علوم و مهندسی کامپیوتر
نشریه علمی پژوهشی انجمن کامپیوتر ایران
مجلد ۶، شماره ۱ (الف)، ۱۳۸۷
صفحات ۱-۱۱
مقاله عادی

بهبود کیفیت سیگنال گفتار نویزی به کمک دینامیک‌های غیرخطی و پویائی جاذب‌ها در شبکه‌های عصبی

لوئیزا دهیادگاری سیدعلی سیدصالحی ایثار نژادقلی

دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده

در این مقاله با استفاده از توانایی شبکه‌های عصبی بازگشتی غیرخطی در بازیابی اطلاعات و خواص جاذب‌های پیوسته سعی در حذف نویز از سیگنال گفتار و بازشناسی آوا از روی سیگنال‌های پاکسازی شده، داریم. شبکه عصبی بازگشتی با گفتار تمیز تعلیم می‌بیند و سپس از آن برای بازشناسی آوا در سیگنال گفتار نویزی که با نویزهای ایستان و غیرایستان نویزی شده است، استفاده می‌شود. در طراحی شبکه عصبی از توابع غیرخطی و اتصالات بازگشتی در لایه پنهان استفاده شده است. برای بررسی کارائی این شبکه، نتایج بدست آمده از آن با نتایج بازشناسی آوا در شبکه جلوسو مقایسه می‌شوند. شبکه علاوه بر طبقه‌بندی صحیح آواها، پاکسازی سیگنال نویزی و نزدیک کردن آن به سیگنال تمیز را با استفاده از خواص جاذب‌ها یاد می‌گیرد. اتصالات بازگشتی توانسته است در نسبت سیگنال به نویز صفر دسی‌بل دقت بازشناسی را برای نویز ایستان ۲۱ و برای نویز غیرایستان ۱۴ درصد بهبود دهد.

کلمات کلیدی: بازشناسی مقاوم گفتار به نویز، شبکه‌های عصبی بازگشتی، دینامیک‌های غیرخطی، جاذب‌های پیوسته، نویز ایستان، نویز غیرایستان.

۱- مقدمه

با رشد روزافزون استفاده از سیستم‌های گفتار در کاربردهای عملی و روزمره نیاز به حفظ راندمان بازشناسی گفتار در محیط‌های واقعی به عنوان امری اجتناب‌ناپذیر مطرح گردیده است[۱]. شرایط ایده‌ال و عاری از نویزی که در کارها و شبیه‌سازی‌های کامپیوتری در نظر گرفته می‌شود در بسیاری از کاربردهای واقعی به صورت جدی نقض می‌شود. بنابراین هنگامی که از سیستم بازشناسی گفتار که در محیط آزمایشگاهی آموزش داده شده است، در محیط واقعی استفاده می‌شود اغلب راندمان سیستم بازشناسی به دلیل عدم انطباق دادگان آموزشی در آزمایشگاه و داده جمع آوری شده در محیط واقعی به مقدار زیادی کاهش می‌یابد[۲].

با توجه به اینکه انسان توانائی بازشناسی گفتاری را که با هر یک از دو نوع نویز ایستان[۱] و غیرایستان[۲] تخریب شده باشد دارد[۳]، [۴]، مبحث مقاوم‌سازی مدل‌های بازشناس گفتار در برابر این نویزها به عنوان یکی از زمینه‌های فعال تحقیقاتی در سال‌های اخیر مطرح بوده است [۵].

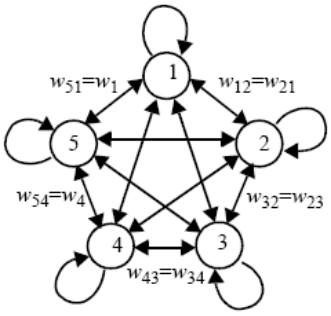
روش‌های مختلفی برای بازیابی گفتار نویزی پیشنهاد شده‌اند[۶]، [۷]. هدف این روش‌ها بدست آوردن تخمین مناسبی از نویز و بهبود شرایط گفتار نویزی می‌باشد[۸]. به عنوان مثال تفریق طیف یک روش ساده بازیابی گفتار است[۹]، [۱۰]. در این روش تخمینی از نویز با استفاده از متوسط‌گیری روی فریم‌هائی از گفتار که در آنها فقط نویز وجود دارد، بدست می‌آید و از سیگنال گفتار کم می‌شود. اما مشکل اصلی این روش این است که در مقابله با نویز غیرایستان مناسب نیست و تخمین مناسبی از نویز را نمی‌تواند بدست آورد [۹].

روش‌های بازیابی گفتار به کمک شبکه‌های عصبی تخمین نرم‌تری[۳] از سیگنال گفتار بدست می‌آورند. توانائی شبکه‌های عصبی مصنوعی در تخمین توابع غیرخطی آنها را برای نویز غیرایستان و نیز توابع غیرخطی که در پارامترهای LHCB سیگنال گفتار [بخش ۴-۱] وجود دارند، مناسب ساخته است.

در تحقیقات مختلف از شبکه‌های عصبی برای بازیابی گفتار استفاده شده است [۷]، [۱۱] ،[۱۲]، [۱۳]. همچنین تحقیقات نشان داده‌اند که شبکه‌ها ابزار بسیار مناسبی برای کاهش بعد دادگان هستند [۱۴]، [۱۵]. شبکه‌های عصبی مصنوعی علیرغم توانائی زیاد نمی‌توانند به راحتی رفتار زمانی سیگنال گفتار را مدل کنند. در نتیجه تنها راه برای رسیدن به این هدف استفاده از اتصالات بازگشتی و تأخیردار در شبکه است [۱۶].

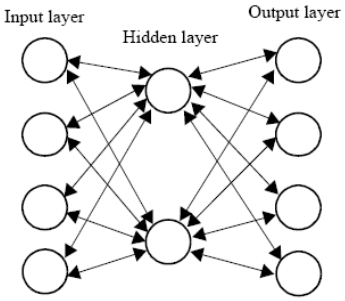
به عنوان مثال شبکه هاپفیلد یک شبکه عصبی بازگشتی[۴] است [۱۷]. گره‌های این شبکه شبیه به نورون‌های واقعی می‌باشند و می‌توانند در دو حالت ممکن آتش[۵] و یا خاموش[۶] قرار بگیرند. شبکه هاپفیلد برای ذخیره تعدادی از الگوها که می‌توانند نمونه‌های نویزی را اصلاح کنند، طراحی شده است (شکل ۱).

این شبکه این کار را با ساختن یک فضای انرژی که در آن جاذب‌ها[۷] (نقاطی که کمینه انرژی را دارند) وجود دارند، انجام می‌دهند. در این فضا هر جاذب نماینده یک الگوی ذخیره شده است. الگوهای نویزی و یا الگوهای ناقص حالت‌هائی از شبکه هستند که به جاذب‌ها منجرمی‌شوند. در یک سیکل شبکه هاپفیلد انرژی الگوهای نویزی آن در فضای انرژی کاهش می‌یابد تا به یک حالت جاذب که نزدیکترین الگوی ذخیره شده به آن می‌باشد، برسد.



شکل ۱- شبکه عصبی هاپفیلد[۱۷]

شبکه‌های هاپفیلد محدودیت‌های خاصی در ظرفیت حافظه دارند. ماشین بولتزمن فرم تکمیل شده شبکه هاپفیلد با واحدهای پنهان است [۱۸]. واحدهای پنهان به ماشین بولتزمن اجازه می‌دهد که نسبت به شبکه هاپفیلد همبستگی‌های[۸] با درجه بالاتر را در داده‌ها پیدا کند. به این ترتیب می‌تواند الگوهای پیچیده‌تری را یاد بگیرد (شکل ۲).



شکل ۲- ساختار ماشین بولتزمن[۱۸]

الگوریتم یادگیری ماشین بولتزمن بر این پایه استوار است که گره می‌تواند خروجی را از روی واحدهای ورودی پیش‌گوئی کند. اصلاح وزن‌ها نیز به‌گونه‌ای است که تفاوت بین توزیع احتمالات مشاهده شده را کاهش می‌دهد. که p_{ij} خروجی مطلوب و p'_{ij} خروجی مشاهده شده است.

$$\Delta w_{ij} = \gamma(p_{ij} - p'_{ij}) \qquad (۱)$$

به طور کلی شبکه‌های عصبی به خاطر خاصیت چند به یک و با تعلیم جلوسو می‌توانند سیگنال با نویز جمع شده را به سیگنال تمیز بنگارد. شبکه‌های جلوسو فضای S(n)+N(n) را به فضای S(n) می‌نگارند. که در آن S(n) سیگنال گفتار جمع شده با نویز، S(v) سیگنال گفتار تمیز و N(n) نشان دهنده نویز می‌باشد.

$$S(n) + N(n) \Rightarrow S(n) \qquad (۲)$$

به شرط آنکه داده کافی از تمام شرایط نویزی داشته باشند، و شبکه نیز ظرفیت یادگیری تمامی شرایط را داشته باشد. این شبکه به نویز تعلیمی وابسته می‌شود ولی در اکثر شرایط همه نویزها در دسترس نیستند و دستیابی به روشی برای حذف نویز و بازیابی سیگنال تمیز که نیاز به برآوردی از ماهیت نویز نداشته باشد، یک ضرورت به نظر می‌رسد. و ما در این تحقیق به ارائه روشی برای رسیدن به این هدف می‌پردازیم. در این راستا از کارائی جاذب‌ها در شبکه‌های عصبی بازگشتی استفاده شده است. استفاده از این جاذب‌ها باعث می‌شود که الگوهای اطراف را به الگوی مورد نظر بنگارد. به عنوان مثال می‌توان به آزمایشات انجام شده در مراجع [۱۹]، [۲۰] اشاره کرد.

استفاده‌های مشهور از شبکه‌های عصبی بازگشتی و دینامیک‌های جاذب که از نظر بیولوژیکی مدل‌های مختصری از حافظه می‌باشند [۲۱] ، شامل ذخیره حافظه‌های انجمنی [۲۲]، بازسازی تصاویر نویزی [۲۳] و جستجو برای یافتن کوتاهترین مسیر در حل مسئله فروشنده دوره‌گرد [۱۷] است. شبکه‌های عصبی جاذب با ایده فضای انرژی مرتبط می‌باشند [۲۴]. در این شبکه‌ها به دلیل وجود اتصالات بازگشتی خروجی بدست آمده از روی ورودی دوباره به عنوان ورودی به شبکه داده می‌شود و این روند چندین بار تکرار می‌شود تا به این ترتیب حالت شبکه تغییر تدریجی پیدا کند و مکرراً بهنگام شود. این روند را اصطلاحاً دور زدن در شبکه می‌گوئیم. تغییرات تدریجی در حالت شبکه در برخی شرایط به سمت کاهش انرژی پیش می‌رود تا سرانجام به یک حالت تعادل متناظر با یک کمینه محلی انرژی یا قعر بستر جذب برسد.

با توجه به نکات ذکر شده در این مقاله سعی داریم با استفاده از شبکه‌های عصبی بازگشتی با توابع غیرخطی و خواص جاذب‌ها بازشناسی آوا در سیگنال گفتار نویزی را بهبود دهیم. به این منظور از ایده تعلیم سیگنال تمیز به شبکه عصبی بازگشتی استفاده کرده‌ایم. بنابراین در هنگام تعلیم، حالت شبکه به گونه‌ای بهنگام می‌شود که سیگنال‌های صوتی تمیز که به شبکه تعلیم داده شده‌اند، به عنوان قعر بستر جذب‌ها در فضای دینامیک‌های شبکه شکل می‌گیرند. و در هنگام تست شبکه که سیگنال‌های نویزی به عنوان ورودی به شبکه داده می‌شوند، در اثر دور زدن در شبکه، شبکه حالت پویائی خود را طی می‌کند و سیگنال‌های نویزی به سمت قعر این بستر جذب‌ها که سیگنال‌های تمیز هستند هدایت می‌شوند و سپس برای بازشناسی آوا از این سیگنال‌های پاکسازی شده استفاده می‌شود.

به این ترتیب ما از خاصیت پویایی جاذب‌ها، به منظور کاهش بعد غیرخطی فضای سیگنال نویزی به زیر فضای سیگنال تمیز استفاده می‌کنیم. برای این منظور لازم است که زیر فضای گفتار تمیز به عنوان جاذب پیوسته در شبکه عصبی بازگشتی شکل بگیرد. بستر جذب این جاذب پیوسته، کل فضای ورودی خواهد بود که تمام سیگنال‌های نویزی و یا اعوجاج یا تغییر یافته را در بر می‌گیرد.

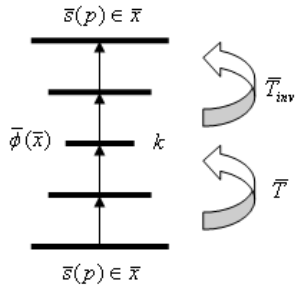
در این مقاله برای ارزیابی مدل‌های طراحی شده، یک مدل مرجع پیشنهاد شده است که یک مدل بازشناس بدون اتصالات بازگشتی است. مدل‌های طراحی شده سعی در حذف نویز از سیگنال گفتار را دارند و از این‌رو انتظار می‌رود که به دقت بازشناسی بهتری نسبت به شبکه مرجع دست پیدا کنند. در ادامه به بررسی

جاذب‌ها می‌پردازیم. سپس ساختار شبکه و بازشناسی مقاوم گفتار را بررسی می‌کنیم و در نهایت آزمایشات انجام شده و نتایج بدست آمده از این آزمایشات را بیان می‌کنیم.

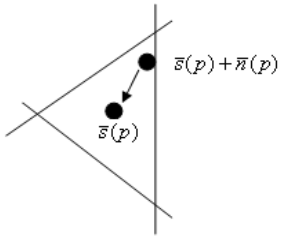
۲- دینامیک‌های جاذب

همان‌گونه که در مقدمه نیز اشاره شد، مدل پیشنهاد شده در این مقاله برای بازشناسی مقاوم گفتار نویزی، یک شبکه عصبی با اتصالات بازگشتی می‌باشد که در طراحی آن از ایده جاذب‌ها برای بازسازی گفتار نویزی در لایه ورودی استفاده شده است. به این دلیل در این قسمت با جاذب‌ها و چگونگی امکان حذف نویز و اعوجاجات توسط آنها در شبکه‌های عصبی بازگشتی غیرخطی بیشتر آشنا می‌شویم.

فرض می‌کنیم یک سیگنال m بعدی گسسته $\bar{s}(p)$ را داریم، و آن را به یک شبکه عصبی خودانجمنی مانند شبکه عصبی شکل ۳ تعلیم می‌دهیم و همچنین فعلاً فرض می‌کنیم که توالی نمونه‌ها مورد نظر نباشد. یعنی عملاً $\bar{s}(p)$ متشکل از p نمونه در فضای ورودی است که تعلیم داده می‌شوند. در ابتدا فرض می‌کنیم که نورون‌های شبکه دارای توابع غیرخطی سخت (پله‌ای) هستند. هر یک از نورون‌ها عملاً در فضای ورودی خود (لایه ماقبل) یک ابرصفحه را بیان می‌کنند. فضای ورودی توسط این ابرصفحات چندی‌سازی[۹] می‌شود و برای هر ناحیه ما بین ابرصفحات، بیان خروجی لایه نورون‌ها بیان واحدی می‌باشد و معنی آن این است که اگر نمونه درون ناحیه در اثر نویز دچار تغییراتی شود، بنحوی که از این ناحیه خارج نشود، این تغییرات در لایه بعدی بروز نخواهند داشت (شکل ۴). این خاصیت از تعمیم یک نقطه به یک ناحیه (درونیابی) حاصل شده است.



شکل ۳- ساختار شبکه عصبی خود انجمنی برای استخراج مؤلفه‌های اساسی



شکل ۴- اگر نمونه درون ناحیه در اثر نویز دچار تغییراتی شود، بنحوی که از این ناحیه خارج نشود، این تغییرات در لایه بعدی بروز نخواهد داشت

در عمل ما سیگنال ورودی را توسط تبدیل غیرخطی با استفاده از یک سری توابع پایه $\phi_j(\bar{x})$ به مؤلفه‌هائی تجزیه کرده‌ایم که در اینجا $\phi_j(\bar{x})$ یک ابرصفحه به ازای هر نورون j ام است، که در خروجی خود ۰ و یا ۱ می‌دهد.

در چنین حالتی مسئله مهم و اساسی این است که موقعیت هر کدام از این ابرصفحات را به گونه‌ای تعیین کنیم که نمونه‌های ورودی ما در وسط هر ناحیه قرار بگیرند تا تحمل حداکثر اعوجاج را داشته باشند. و مسئله دوم اینکه این ابرصفحات به‌گونه‌ای قرار بگیرند که بتوانند تمامی نمونه‌ها را در فضای ورودی از یکدیگر متمایز کنند و این تمایز نمونه‌های ورودی در تمامی لایه‌ها ادامه یابد تا در خروجی به طور متمایزی بتوانند با کمترین خطا مجدداً نمونه‌های ورودی را تولید نمایند.

حال فرض می‌کنیم که نورون‌ها به جای توابع غیرخطی سخت، توابع غیرخطی نرم مثل تابع سیگموئید داشته باشند تا بتوانیم از قانون پس‌انتشار خطا برای تعلیم شبکه استفاده کنیم. در این حالت مرزها نرم و فازی می‌شوند و خروجی هر نورون بسته به موقعیت بیان نمونه $\bar{s}(p)$ در ورودی‌اش یک مقدار پیوسته به خود می‌گیرد و هر نمونه ورودی $\bar{s}(p)$ در هر لایه یک تفسیر با این نورون‌ها خواهد داشت؛ در لایه وسط نیز همین‌گونه است.

ابتدا فرض می‌کنیم که به این شبکه (با توابع نرم) فقط یک نمونه تعلیم داده می‌شود. برای بیان یک نمونه فقط داشتن یک نورون در لایه وسط کافی است و اگر بیشتر از یک نورون در این لایه قرار دهیم مقادیر خروجی همه آنها برای این نمونه پس از تعلیم یکسان خواهد شد [۱۹].

در این حالت عملاً ابرصفحات نرم (فازی) به گونه‌ای قرار می‌گیرند که بهتر بتواند برای این نمونه بیان کاملتری در خروجی نورون‌ها ارائه نمایند و در خروجی شبکه نیز با خطای نزدیک به صفر این نمونه تولید گردد، و در نقطه قعر (جاذب) یک خوشه قرار گیرد که توسط این ابرصفحات نرم ساخته می‌شود. فضای ورودی توسط این ابرصفحات درونیابی می‌شود و هر نقطه دیگر $s'(p) = \bar{s}(p) + \bar{n}(p)$ در این فضا به طور غیرخطی به این مؤلفه واحد تصویر می‌گردد. لیکن این مؤلفه اساسی واحد (یک نورون در لایه وسط) مقدار کمتری را نشان می‌دهد، زیرا نمونه اعوجاج یافته جدید قرار است توسط نمونه اصلی بیان شود و بسته به نزدیک بودن و شبیه بودن خود به آن این بیان قوی‌تر خواهد بود. و این به دلیل درونیابی فضای ورودی توسط تابع کرنل واحد $\phi(\bar{x})$ می‌باشد که بیان‌گر این مؤلفه اساسی است.

$$\bar{s}(p) \Rightarrow \phi_{\max}(\bar{x}) \qquad (۳)$$

$$\bar{s}(p)+\bar{n}(p) \Rightarrow \phi(\bar{x}) < \phi_{\max}(\bar{x}) \Rightarrow \tilde{s}(p)+\tilde{n}(p)$$

به ازای نمونه اصلی که در اینجا با $\bar{s}(p)$ نشان داده شده است، تابع کرنل $\phi(\bar{x})$ ماکزیمم می‌شود، که با $\phi_{\max}(\bar{x})$ نشان داده شده است. حال اگر نمونه‌ها نویزی شوند، $\bar{s}(p)+\bar{n}(p)$ تابع کرنلی $(\phi(\bar{x}))$ بدست می‌آید که از $\phi_{\max}(\bar{x})$ کمتر است. و اگر خروجی بدست آمده از شبکه را دوباره به ورودی بدهیم، ورودی جدید $\tilde{s}(p)+\tilde{n}(p)$ خواهد بود که در آن $\|\tilde{n}(p)\| < \|\bar{n}(p)\|$ می‌باشد. و این یعنی تضعیف نویز و اعوجاج به دلیل آنکه مؤلفه‌های اساسی بیان کننده نویز در شبکه حضور ندارند و در عمل نویز کمی فیلتر سازی غیرخطی می‌گردد. با کمی فیلتر شدن $\bar{n}(p)$ نمونه اعوجاج یافته کمی به نمونه اصلی نزدیک می‌شود و با دادن خروجی شبکه به ورودی می‌توان آن را مجدداً فیلتر نمود. و این کار را آنقدر ادامه داد تا $\bar{n}(p)$ به طور کامل حذف شود؛ آزمایش عملی این موضوع را به خوبی نشان می‌دهد [۲۰].

حال فرض کنیم که به جای یک نمونه تعلیم، p>1 نمونه تعلیم داشته باشیم. در این حالت لازم است که $p, \phi_j(x), (j = 1,2,....,p)$ کرنل متمایز برای هر نمونه تعلیم داشته باشیم که نمونه‌ها را با مرزهای نرم از هم تمایز دهند. در این حالت اگر بتوانیم موقعیت ابرصفحات نرم را به گونه‌ای تعلیم دهیم که این کرنل‌ها را به درستی ایجاد نمایند، آنگاه در این حالت نیز می‌توانیم p کرنل، p جاذب و p

بستر جذب داشته باشیم که توسط آنها نویز افزوده شده به نمونه‌ها فیلتر غیرخطی می‌شود. آزمایش‌های عملی این موضوع را ثابت می‌کند و مسأله مهم در اینجا روشی است که به وسیله آن کرنل‌ها و ابرصفحات نرم در موقعیت‌های بهینه به درستی شکل بگیرند.

حال اگر بخواهیم مسیر پیوسته یک سیگنال گفتار (مسیر بردار بازنمائی) را به شبکه تعلیم دهیم، در عمل مانند آن است که آن را توسط نقاط نزدیک به هم در طول مسیر نمونه‌برداری نماییم. با تعلیم مسیر به شبکه عصبی خودانجمنی در عمل این نقاط به عنوان جاذب‌های نقطه‌ای به شبکه تعلیم داده می‌شوند و شبکه مسیر را به صورت نقاطی گسسته از هم می‌آموزد [۱۹].

رویکرد مورد استفاده در این مقاله، تعلیم مسیر سیگنال مطلوب در فضای ورودی به یک شبکه عصبی بازگشتی به عنوان جاذب پیوسته[۱۰] و فیلتر سازی غیرخطی نویز و غیره از سیگنال مطلوب توسط این شبکه عصبی می‌باشد. عملاً توانائی عملکرد جاذب‌گونه این شبکه عصبی، موجب جذب سیگنال نویزی و یا تغییر یافته، در بستر جذب جاذب سیگنال تمیز به سمت آن می‌گردد. به این وسیله نویز از سیگنال پالایش شده و یا تنوعات ناخواسته از سیگنال حذف می‌گردند.

جاذب‌های پیوسته عملاً مجموعه‌ای از جاذب‌های نقطه‌ای[۱۱] هستند که به دنبال یکدیگر قرار گرفته‌اند و یک یا چند مسیر[۱۲] پیوسته تولید می‌کنند و با ساختارهای مناسب از شبکه‌های عصبی توسط داده‌ها قابل یادگیری هستند و یک ویژگی مفید آنها این است که می‌توانند مسیرهای غیرخطی و پیچیده در فضای بسیار بعدی ورودی را یاد بگیرند و از این طریق امکان پالایش (فیلترسازی) غیرخطی سیگنال‌ها را فراهم کنند. از آنجا که در خیلی از موارد واقعی، تأثیرات نویز و خصوصاً تنوعات[۱۳] بر روی سیگنال ورودی خطی نیست، حذف آنها نیازمند یادگیری مانیفولد[۱۴]های غیرخطی توسط سامانه پالایش کننده (در اینجا شبکه‌های عصبی) است.

۳- بازشناسی مقاوم گفتار با شبکه‌های عصبی بازگشتی

۳-۱- بازشناسی مقاوم گفتار

با مسئله بازشناسی مقاوم گفتار به گونه‌های مختلفی برخورد شده است که در همه اینها هدف، محاسبه خروجی سیستم بازشناس است وقتی که ورودی آن تحت تأثیر عوامل مختلفی مثل نویز تخریب شده است. در برخی تحقیقات [۲۵]، [۲۶]، [۳]، این مسئله با عنوان داده‌های مفقود شده مطرح شده است و ساختارهای مختلفی از شبکه‌های عصبی برای بازشناسی مقاوم گفتار در رویاروئی با داده‌های مفقود شده پیشنهاد شده است. در سال‌های اخیر نیز پروین [۲۷]، [۵]، [۱۶]، [۲۸] از یک ساختار شبکه عصبی بازگشتی برای تخمین مقادیر مفقود شده در بردار ورودی استفاده کرده است. این ساختار که ترکیبی از ساختارهای گینگراس و بنژیو [۲۹] می‌باشد به گونه‌ای طراحی شده است که در آن قسمت‌هایی از ورودی به صورت تصادفی مفقود شده در نظر گرفته می‌شوند و شبکه طبق رابطه زیر سعی در طبقه‌بندی و تکمیل الگوهای ورودی از روی الگوهای قبلی دارد.

$$\tilde{x}(n) = (1-\gamma)x(n) + \gamma y(n-1).vf \qquad (۴)$$

که در این رابطه $\tilde{X}(n)$ داده بازسازی شده در زمان n، X(n) داده ورودی در زمان n-1، γ ضریب بازسازی، vf وزنهای اتصالات بازگشتی از لایه پنهان به ورودی (جهت بازسازی قسمتهای مفقود شده)، f تابع غیرخطی و y(n-1) خروجی لایه پنهان در زمان n-1 می‌باشد.

در این تحقیق با الهام از شبکه پیشنهادی پروین شبکه‌ای طراحی شده است که در آن دادگان تست با هر دو نوع نویز ایستان و غیرایستان نویزی شده است. از نویز سفید به عنوان نویز ایستان استفاده شده است و از صدای عبور ماشین‌ها در زمان‌های مختلف به عنوان نویز غیرایستان استفاده شده است و این دو نویز به صورت نویز جمعی[۱۵] به سیگنال گفتار اضافه شده‌اند. شبکه عصبی بازگشتی با دادگان تمیز تعلیم می‌بینند و برای هر ورودی تعلیم آنقدر ادامه پیدا می‌کند تا به بهترین خطای ممکن برای طبقه‌بندی و تخمین ورودی برسد. ساختار شبکه عصبی بازگشتی و شیوه تعلیم آن در ادامه بررسی می‌شود.

۳-۲- ساختار شبکه عصبی بازگشتی

از آنجا که در این تحقیق، تعلیم شبکه با دادگان تمیز و تست آن با دادگان نویزی انجام می‌گیرد، در طراحی شبکه سعی کرده‌ایم، بگونه‌ای شبکه را تعلیم دهیم که شبکه علاوه بر طبقه‌بندی صحیح دادگان، بازسازی ورودی از روی دادگان نویزی را نیز یاد بگیرد.

در تحقیقات قبلی با الهام از شبکه خانم پروین یک شبکه عصبی با دو اتصال بازگشتی طراحی کردیم، که برای فیلترسازی غیرخطی سیگنال گفتار نویزی از آن استفاده شد و نتایج قابل قبولی به دست آمد. اتصالات بازگشتی در ساختار شبکه اتصالات کاملی هستند که یکی از آنها، از لایه پنهان قبلی به لایه پنهان بعدی و دیگری از لایه پنهان به لایه ورودی با یک واحد تأخیر متصل می‌باشند. اتصال بازگشتی اول توجه به زمینه را افزایش می‌دهد و به اتصال بلندمدت گذشته به آینده و به بازشناسی بهتر آوا کمک می‌کند و اتصال بازگشتی دوم با توجه به رابطه ۵ برای حذف نویز مؤثرتر است و نقش بیشتری را در این زمینه و در تخمین کوتاه مدت در حالت نویزدار بر عهده دارد [۲۰].

$$\tilde{x}(n) = (1-\gamma)x(n) + \gamma y(n-1).vf \qquad (۵)$$

که در آن $\tilde{x}(n)$ ورودی بازیابی شده توسط شبکه، x(n) ورودی در لحظه n، vf وزن‌های اتصالات بازگشتی و γ ضریب بازسازی شبکه است که پس از آزمایشات مقدار آن ۰/۷ در نظر گرفته شده است. در ابتدای تعلیم شبکه مقادیر y(n-1) صفر در نظر گرفته می‌شود و vf مقادیر تصادفی کوچک به خود می‌گیرد. و پس از هر دوره تعلیم با توجه به رابطه ۵، $\tilde{x}(n)$ محاسبه می‌شود.

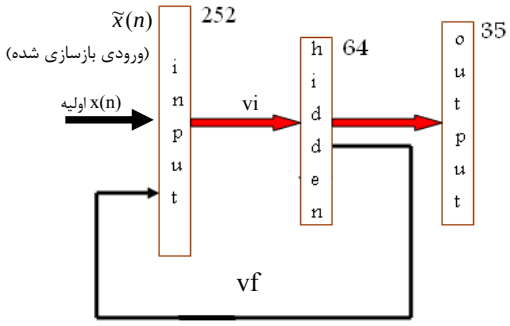
در این تحقیق در ادامه کارهای قبلی به دنبال بهبود شبکه عصبی بازگشتی هستیم به‌گونه‌ای که بتوان در آن از خاصیت جاذب‌ها برای بازسازی سیگنال نویزی استفاده کرد. همان‌گونه که در رابطه ۵ دیده می‌شود برای تخمین ورودی بازسازی شده در لحظه n یعنی $\tilde{x}(n)$ از ۰/۳ ورودی به اضافه ۰/۷ خروجی اتصالات بازگشتی لایه پنهان در لحظه n-1 استفاده شده است. یعنی در هر دوره تعلیم[۱۶] ورودی از روی ورودی‌های زمان قبل تخمین زده و بازسازی می‌شود.

بررسی شبکه نشان می‌دهد، چنانچه بتوانیم بجای تخمین سیگنال نویزی از روی ورودی‌های زمان قبل، سیگنال را از روی خودش در دوره تعلیم قبل بازسازی کنیم باید به نتایج معتبرتری دست پیدا کنیم. زیرا هر سیگنال برای بازسازی خود، دارای اطلاعات مفیدتری نسبت به سیگنال‌های زمان قبل می‌باشد. به این منظور بجای تخمین هر ورودی از ورودی‌های قبلی شبکه را بگونه‌ای تعلیم دهیم که اتصال بازگشتی از لایه پنهان به لایه ورودی، بجای ورودی لحظه بعد ورودی همان لحظه را به صورت خودانجمنی و با استفاده از خواص جاذب‌ها بازسازی کند و در

واقع تأخیر یک فریم ورودی را نداشته باشد (شکل۵). بنابراین رابطه ۵ طبق رابطه ۶ تغییر پیدا می‌کند.

$$\tilde{x}(n) = (1-\gamma)x(n) + \gamma\, y(n).vf \qquad (۶)$$

در این رابطه، x(n) بردار حاوی پارامترهای بازنمایی استخراج شده از ۱۴ فریم متوالی گفتاری است (۱۴ ×۱۸ پارامتر). به کمک بستر جذب و جاذب‌های یادگیری شده در شبکه بازگشتی، که قبلاً براساس الگوهای تعلیمی یعنی گفتار تمیز تعلیم یافته اند، این ورودی x(n) داده شده در هر گام پیشروی بر روی مسیر گفتار، به صورت غیر خطی از نویزها و آلاینده های خارج از مسیر جاذب‌ها پالایش می‌گردد و به سمت جاذب مربوط به خود که گفتار تمیز نظیر آن x(n) است، جذب می‌گردد. به این وسیله فضای با بعد بالاتر S(n)+N(n) به زیر فضای با بعد کمتر جاذب پیوسته گفتار تمیز نظیر آن به طور غیرخطی تصویر می شود. باید توجه نمود که در این شیوه کاهش بعد غیر خطی توسط جاذب‌ها، n یعنی گام زمانی پیشروی روی مسیر گفتار تا پایان دور زدن در اتصال بازگشتی و پالایش غیر خطی نویز، مقدار ثابتی است و از این دیدگاه پیشروی در محور زمان بر روی مسیر گفتار ورودی نداریم. ولی برای عملکرد جاذب‌ها در اتصالات بازگشتی طبیعتاً نیاز به تأخیر زمانی است تا x(n) ورودی با دور زدن مکرر در شبکه از مسیر اتصالات بازگشتی به جاذب نظیر خود برود و از این دیدگاه فرصت زمانی کافی به شبکه بازگشتی می بایستی داده شود.



شکل ۵- ساختار شبکه عصبی بازگشتی، در این ساختار اتصال بازگشتی لایه پنهان به ورودی برای بازسازی ورودی در همان لحظه n است

تنها تفاوت رابطه ۶ با رابطه ۵ در این است که در این حالت y(n) خروجی لایه پنهان برای ورودی لحظه n می‌باشد. با این شیوه تعلیم شبکه عصبی بهتر می‌تواند سیگنال‌های تمیز را در قعر بستر جذب‌ها قرار دهد. در این رابطه $\tilde{x}(n)$ الگوی ورودی در حال بازیابی توسط شبکه بازگشتی است که x(n) ورودی اولیه شبکه را دریافت می کند و از طریق اتصالات بازگشتی، $\tilde{x}(n)$ را به طور مکرر محاسبه و آن را بازسازی می کند.

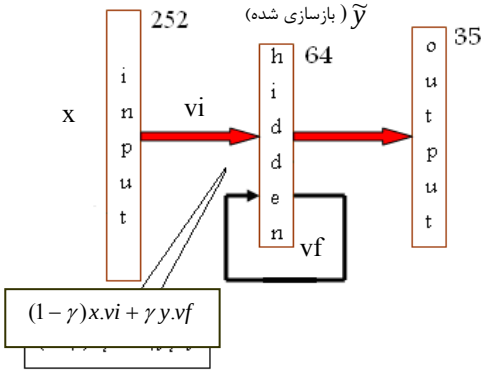
اما با توجه به اینکه تابع استفاده شده برای تخمین x(n) در خروجی لایه بازگشتی، یک تابع خطی است اتصال بازگشتی را با توجه به رابطه‌های زیر به لایه پنهان می‌توان منتقل کرد.

$$\tilde{y}(n) = f(\tilde{x}(n).vi) \qquad (۷)$$

که در این رابطه f تابع غیرخطی سیگموئید دوقطبی و vi وزنهای اتصال مستقیم لایه ورودی به پنهان می‌باشند. با توجه به مقدار $\tilde{x}(n)$ از رابطه ۶، رابطه ۷ به ترتیب زیر تغییر می‌کند.

$$\tilde{y}(n) = f[((1-\gamma)x(n) + \gamma\, y(n).vf).vi]$$
$$\tilde{y}(n) = f[(1-\gamma)x(n).vi + \gamma\, y(n).vf.vi] \qquad (۸)$$

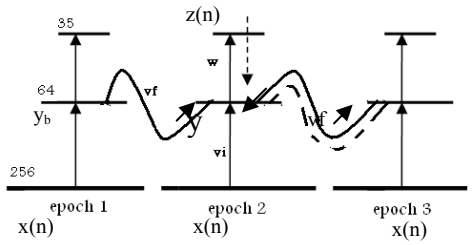
اگر مسؤولیت حاصلضرب دو وزن vf.vi را به وزن vf منتقل کنیم، می‌توانیم اتصال بازگشتی را طبق شکل زیر به لایه پنهان منتقل کنیم.



شکل ۶- ساختار شبکه عصبی بازگشتی با انتقال اتصال بازگشتی به لایه پنهان، در این حالت ورودی‌های لایه پنهان با نسبتهای γ و $1-\gamma$ ترکیب می‌شوند

ساختار شبکه عصبی که در آزمایشات استفاده شده است، در شکل ۶ نشان داده شده است. این شبکه شامل لایه‌های ورودی، پنهان و خروجی است. اتصال بازگشتی نیز در لایه پنهان وجود دارد و مقادیر لایه پنهان هر دوره را به دوره تعلیم قبلی متصل می‌کند. در لایه ورودی ۲۵۲ گره وجود دارد که به تعداد پارامترهای ۱۴ فریم سیگنال گفتار می‌باشد [بخش۴-۱]. در لایه پنهان پس از انجام آزمایشات مختلف ۶۴ نورون در نظر گرفته شد که قابلیت کافی برای ایجاد قدرت تفکیک لازم در فضای ورودی را داشته باشد. و در لایه خروجی ۳۵ نورون به تعداد آواهای زبان فارسی در نظر گرفته شد.

شبکه عصبی بازگشتی یک شبکه غیرخطی با تابع فعالیت سیگوئید دوقطبی است و در آن اتصال بازگشتی در لایه پنهان به فیلتر سازی نویز ورودی با استفاده از پویائی جاذب‌هائی که توسط آن ساخته می‌شوند، کمک می‌کند. بنابراین در طی تعلیم شبکه و برای پس‌انتشار خطا باید از دو خطای مربوط به طبقه‌بندی آواها و بازسازی ورودی لایه پنهان یعنی x.vi استفاده کرد. شکل ۷ وزنهای اتصالات شبکه و نحوه پس‌انتشار خطا در شبکه را نشان می‌دهد. خطوط ممتد نشان دهنده اتصالات کامل شبکه می‌باشد که وزنهای این اتصالات توسط الگوریتم پس‌انتشار خطا باید اصلاح شوند. خطوط نقطه‌چین نیز نشان دهنده نحوه پس‌انتشار خطا در شبکه می‌باشند.



شکل ۷- وزن‌های شبکه عصبی بازگشتی و نحوه پس‌انتشار خطا. خطوط تیره وزن‌های شبکه و خطوط نقطه چین نحوه پس‌انتشار خطا را نشان می‌دهد

شبکه در لایه خروجی طبق رابطه (۲) طبقه بندی صحیح آواها را تعلیم می‌بیند و در هر بار تعلیم فریم هفتم ورودی را در خروجی تشخیص می‌دهد، که در این رابطه $d(n,i)$ خروجی مطلوب در لحظه n برای نورون iام خروجی و $z(n,i)$ خروجی بدست آمده در طی تعلیم شبکه در لحظه n برای خروجی iام و E_{1n} مجموع مربعات خطای مربوط به بازشناسی آواها در خروجی در لحظه n و l تعداد گره‌های خروجی شبکه است.

$$E_{1n} = \sum_{i=1}^{l} (d(n,i) - z(n,i))^2 \qquad (۹)$$

از اتصال بازگشتی برای بازسازی ورودی از روی داده نویزی استفاده شده است. از آنجا که در هنگام تعلیم، ورودی‌های شبکه دادگان تمیز می‌باشند، وزن‌های اتصال بازگشتی به گونه‌ای همگرا می‌شوند که بتوانند سیگنال نویزی ورودی را ابتداً توسط پویائی جاذب‌های ناشی از این اتصالات بازگشتی حتی‌الامکان بازسازی کرده و سپس عمل بازشناسی روی نتیجه حاصله انجام پذیرد. در هر دوره تعلیم خروجی لایه پنهان طبق رابطه (۸) بدست می‌آید و سپس پس‌انتشار خطا طبق رابطه (۱۰) به گونه‌ای انجام می‌شود که حاصلضرب خروجی لایه پنهان در وزن‌های بازگشتی با حداقل خطا به حاصلضرب ورودی در وزن‌های لایه ورودی نزدیک شود.

$$E_2 = \sum_{i=1}^{m} [\sum_{j=1}^{m} x(n,j).vi_{ji} - \sum_{j=1}^{m} y_b(n,j).vf_{ji}]^2 \qquad (۱۰)$$

در اینجا n شماره الگوی تعلیمی ورودی و y_b خروجی لایه y مربوط به دوره تعلیم قبلی است و m تعداد گره‌های لایه پنهان است. نحوه محاسبه سیگنال خطای دلتا، پس انتشار شده به لایه y به صورت زیر است.

$$\delta_y = (\delta_z W^T + (x.vi - \tilde{y}.vf)vf^T) y(1-y) \qquad (۱۱)$$

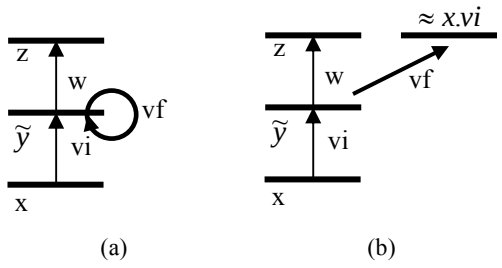
$$\tilde{y} = f((1-\gamma)x.vi + \gamma \tilde{y}_b.vf) \qquad (۱۲)$$

در روابط فوق، δ_z بردار سیگنال خطای دلتا در خروجی شبکه مربوط به خطای طبقه بندی آواها، W ماتریس وزنهای لایه خروجی شبکه ، δ_y بردار سیگنال خطای دلتا پس انتشار شده در لایه پنهان، $\tilde{y}$ خروجی لایه پنهان در این دوره تعلیم، f تابع غیرخطی سیگموئید دو قطبی ، y(1-y) مشتق تابع f، $\tilde{y}_b$ خزوجی لایه پنهان در دوره تعلیم قبلی و γ ضریب بازسازی است.

باید توجه داشت که در این ساختار شبکه عصبی و این نحوه تعلیم، ما عملاً به نحوی از یادگیری چند تکلیفی[۱۷] استفاده می کنیم. یعنی وزنهای نورون‌های لایه پنهان به نحوی تعلیم می‌بینند که بتوانند هم در خروجی شبکه، طبقات مطلوب آوایی را تولید نمایند و هم در خروجی اتصالات بازگشتی، مقادیر x.vi را به عنوان مقادیر مطلوب حتی الامکان ایجاد نمایند.

a) ساختار شبکه در حین استفاده، b) ساختار شبکه در حین یادگیری چند تکلیفی، وزن‌های vi باید طوری یادگیری شوند که هم بتوانند طبقات مطلوب را در خروجی z تولید کنند و هم در خروجی اتصالات vf، مقادیر x.vi را به عنوان مقادیر مطلوب حتی الامکان بسازند.

در ابتدای تعلیم مقدار خروجی لایه پنهان صفر در نظر گرفته می‌شود و به این ترتیب در ابتدای تعلیم خروجی لایه پنهان تماماً از روی ورودی ساخته می‌شود، و به تدریج که تخمین ورودی توسط اتصال بازگشتی به ورودی اصلی نزدیکتر می‌شود، تأثیر آن در بدست آوردن خروجی لایه پنهان نیز بیشتر می‌شود. در رابطه (۸) با الهام از تحقیقات پروین [۵]، اتصالات رسیده از ورودی و از خروجی اتصالات بازگشتی با نسبت γ با هم ترکیب می‌شوند. و این از آن جهت است که حضور نسبی از الگوی ورودی موجب گردد تا در بازیابی الگوی بدون نویز متناظر با آن کمک نماید و الگوی ورودی را در مسیر حرکت به سوی جاذب در بستر جذب متناظر با همین الگو حفظ نماید. اتصالات بازگشتی نیز بستر جذب‌ها را شکل می‌دهند و به بیان دیگر با تکرار دور زدن در این لایه فیلتر سازی نویز از سیگنال به دفعات و مکرراً انجام می‌پذیرد.



شکل ۸- ساختار شبکه و یادگیری چند تکلیفی

در هنگام تست نیز برای بدست آوردن خروجی از همین رابطه استفاده می‌شود، ولی برای آنکه جاذب‌ها بتوانند نقش خود را در حذف نویز حفظ نمایند و فرصت کافی به شبکه داده شود تا در بستر جذب جاذب‌ها به سمت جاذب‌های مربوطه حرکت کند، نیاز است که به شبکه فرصت داده شود تا خروجی لایه پنهان از مسیر اتصال بازگشتی چند بار بازسازی شود. که در هر دور با ورودی رسیده از لایه ورودی نیز جمع می‌شود. به این ترتیب می‌توانیم طبقه‌بندی صحیح‌تری در خروجی داشته باشیم.

لذا با توجه به این عملکرد شبکه، در مرحله تعلیم نیز پس‌انتشار خطا از دو منشا مختلف انجام می‌پذیرد. خطای طبقه‌بندی آواها (رابطه ۹) از لایه خروجی به سمت لایه پنهان پس‌انتشار می‌شود و وزن‌های لایه پنهان در این جهت اصلاح می‌شوند. در مرحله تعلیم خروجی مطلوب لایه بازگشتی عملاً همان خروجی رسیده از لایه ورودی است (رابطه۱۰)، و در هنگام تعلیم سیگنال گفتار تمیز، اتصال بازگشتی باید یاد بگیرد که تخمین مناسبی از سیگنال رسیده به لایه پنهان را از طریق ورودی بسازد. وزن لایه ورودی به گونه‌ای اصلاح می‌شود که بتواند همزمان هم طبقه‌بندی صحیح انجام پذیرد و هم به عنوان ورودی مناسبی برای لایه بازگشتی باشد. در واقع هر دو خطا در تصحیح این وزن دخالت دارند. در ادامه به بررسی آزمایشات انجام شده بر روی این شبکه عصبی می‌پردازیم.

۴- مجموعه آزمایشات و دادگان

۴-۱- مجموعه دادگان

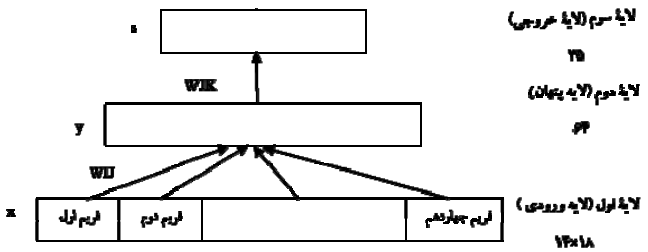
۲۰ جمله از دادگان[۱۸] فارس‌دات [۳۰] که توسط یک گوینده بیان شده است در آزمایشات اولیه این تحقیق استفاده شده‌اند، که ۱۰ جمله به عنوان دادگان تعلیم و۱۰ جمله به عنوان دادگان تست در نظر گرفته شده‌اند. این جملات شامل همه آواهای فارسی می‌باشند و می‌توانند نتایج واقعی داشته باشند. در آزمایشات نهائی نیز دادگان را به ۸۰۰ جمله از ۱۰ نفر افزایش دادیم.

به منظور طبقه‌بندی واحدهای گفتار باید دادگان را در طی چند مرحله آماده‌سازی کنیم. اولین مرحله استخراج الگو از روی سیگنال صوتی می‌باشد. در

این مرحله ما از روش لگاریتم انرژی بانک فیلترهای مجذور هنینگ[19] برای استخراج الگو استفاده کردیم، زیرا این روش در تحقیقات قبلی [۳۱]، [۳۲]، نتیجه مطلوبی را برای بازشناسی آواهای فارسی داشته است. مرحله بعد هنجارسازی پارامترهای استخراج شده است. در اینجا از روش هنجار سازی[20] طولی (شبیه روش هنجارسازی به میانگین و واریانس) استفاده کردیم، که در آن پارامترها نسبت به تغییراتشان در طی کل دادگان هنجارسازی می‌شوند. مرحله آخر برچسب دهی به فریم‌های گفتار می باشد که در دادگان فارس‌دات [۳۰] همه فریم‌ها برچسب دهی شده‌اند.

۴-۲- شبکه مرجع

از آنجا که می‌خواهیم کارائی اتصالات بازگشتی را در بهبود بازشناسی سیگنال گفتار نویزی بررسی کنیم لازم است یک شبکه جلوسو بدون اتصال بازگشتی را به عنوان مقایسه تعریف کنیم. این شبکه را شبکه مرجع می‌گوئیم. کار مدل مرجع به طبقه‌بندی ساده و تشخیص زنجیره بردارهای بازنمائی داده شده در ورودی آن محدود می‌شود. ساختار مدل طبق شکل ۹ شبکه عصبی با تأخیر زمانی می‌باشد که در لایه ورودی ۱۴ فریم متوالی را دریافت می‌کند. این مدل فریم هفتم ورودی را به عنوان یکی از ۳۵ آوای فارسی در خروجی تشخیص می‌دهد. این شبکه یک لایه پنهان با ۶۴ نورون دارد و تابع استفاده شده برای همه نورون‌های آن، تابع غیرخطی سیگموئید دوقطبی می‌باشد. از الگوریتم پس‌انتشار خطا برای تعلیم شبکه استفاده شده است و پس از تعلیم شبکه دقت بازشناسی مدل را برای همه گروه‌های آوای فارسی و روی دادگان تست بدست آوردیم که در جدول آمده است. انتظار داریم که از مدل‌های طراحی شده دقت بازشناسی بهتری نسبت به شبکه مرجع بگیریم.



شکل ۹- ساختار شبکه مرجع

۴-۳- آزمایشات

در مراحل تحقیق به منظور تحلیل سریعتر شبکه عصبی بازگشتی در ابتدا ۲۰ جمله از دادگان فارس‌دات برای تعلیم و تست شبکه استفاده شدند و پس از آن دادگان را به ۸۰۰ جمله از دادگان فارس‌دات افزایش دادیم. تعلیم شبکه با دادگان تمیز و عاری از نویز انجام می‌شود و برای تست شبکه از دادگان نویزی با نسبت‌های سیگنال به نویز ۰، ۵، ۱۰، ۱۵ و ۲۰ دسی‌بل استفاده می‌شود. نویز استفاده شده در آزمایشات بر دو نوع نویز ایستان و نویز غیرایستان می‌باشد که نویز ایستان از نوع نویز سفید و نویز غیرایستان صدای خودروها در خیابان انتخاب شده است.

در هنگام تست مقاوم بودن کیفیت بازشناسی آواها توسط شبکه نسبت به نویزهای ایستان و غیرایستان مورد بررسی قرار می‌گیرد و عملکرد شبکه بازگشتی که مسیر گفتار تمیز را به عنوان جاذب پیوسته تعلیم دیده است، نسبت به شبکه مرجع ارزیابی می‌شود. نتایج تعلیم و تست شبکه‌ها با هر دو دسته از دادگان در ادامه آمده است.

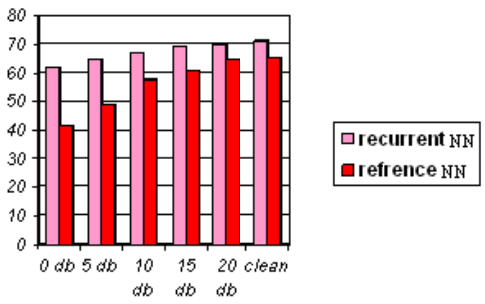
۵- نتایج

۵-۱- تعلیم جاذب‌های پیوسته با گفتار تمیز از یک گوینده

در ابتدا هر دو شبکه بازگشتی و مرجع با ۱۰ جمله از دادگان فارس‌دات تعلیم داده می‌شوند. سپس ۱۰ جمله دیگر از دادگان که توسط همان گوینده بیان شده‌اند که با دو نوع نویز ایستان و غیرایستان و با نسبت‌های مختلف سیگنال به نویز، نویزی شده‌اند برای بازشناسی به شبکه داده می‌شوند. نتایج بدست آمده از دو شبکه در شکل‌های ۱۰ و ۱۱ و جداول ۱ و ۲ آمده است. این نتایج نشان دهنده عملکرد قوی‌تر شبکه بازگشتی نسبت به شبکه مرجع برای هر دو نوع نویز ایستان و غیرایستان و نیز در بازشناسی گفتار تمیز بوده است. به عنوان مثال در SNR=0، ۲۰/۴۲٪ افزایش صحت بازشناسی دیده می‌شود.

جدول ۱- نتایج بازشناسی در حضور نویز ایستان

SNR	شبکه بازگشتی	شبکه مرجع
0 db	۶۲.۰۰۷۰	۴۱.۵۸۹۳
5 db	۶۴.۵۵۹۲	۴۹.۱۲۹۹
10 db	۶۶.۹۳۷۴	۵۷.۷۷۲۶
15 db	۶۹.۰۸۳۵	۶۰.۷۸۸۹
20 db	۷۰.۰۶۹۶	۶۴.۶۱۷۲
سیگنال تمیز	۷۱.۰۵۵۷	۶۵.۱۹۷۲

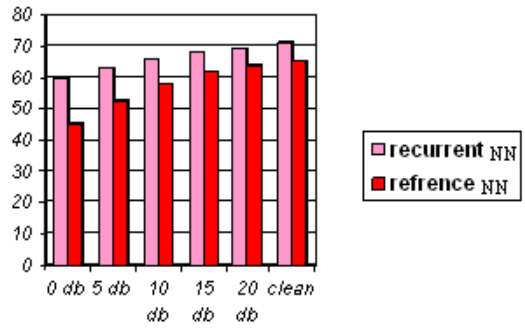


شکل ۱۰- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز ایستان

در آزمایشات بعدی این دو شبکه را که با گفتار یک نفر تعلیم دیده‌اند، با گفتار ۴۰ نفر از افراد با جنسیت‌ها، لهجه‌ها و بیان‌های مختلف، یعنی با ۴۰۰ جمله از دادگان تست کردیم. در این آزمایش کارائی شبکه مرجع به دلیل افزایش تنوعات خصوصاً تنوعات گوینده به شدت افت پیدا می‌کند، اما شبکه بازگشتی افت کمتری داشته است و با توجه به شکل‌های ۱۲ و ۱۳ و جدول‌های ۳ و ۴ توانسته است در مقابل هر دو نوع نویز ایستان و غیرایستان کارائی قابل قبولی داشته باشد و نسبت به مرجع صحت بازشناسی سیگنال نویزی ایستان را ۳۰٪، سیگنال نویزی غیرایستان را ۲۱٪ و سیگنال تمیز را ۲۷٪ بهبود دهد.

جدول ۲- نتایج بازشناسی در حضور نویز غیرایستان

SNR	شبکه بازگشتی	شبکه مرجع
0 db	۵۹.۹۷۶۸	۴۵.۳۵۹۶
5 db	۶۳.۱۶۷۱	۵۲.۶۶۸۲
10 db	۶۵.۸۳۵۳	۵۷.۸۸۸۶
15 db	۶۸.۲۱۳۵	۶۲.۰۰۷۰
20 db	۶۹.۲۵۷۵	۶۳.۸۰۵۱
سیگنال تمیز	۷۱.۰۵۵۷	۶۵.۱۹۷۲

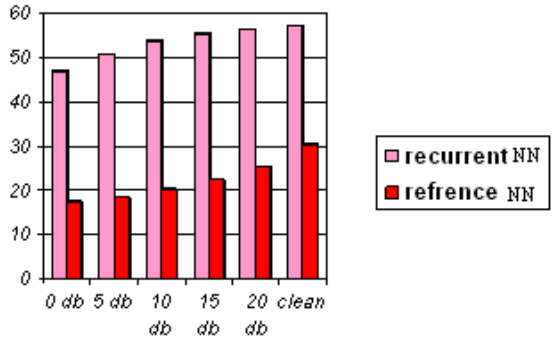


شکل ۱۱- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز غیر ایستان

این نتایج نشان می‌دهند که جاذب‌های شکل گرفته در شبکه بازگشتی برای صحبت یک نفر در هنگام تعلیم حذف تأثیرات تنوعات گوینده و نویز را همزمان انجام داده‌اند و شبکه برای حذف تنوعات گویندگان با وجود نویز غیرایستان نیز کارائی خوبی داشته است. علت این مطلب عملکرد مطلوب جاذب پیوسته یادگیری شده در شبکه بازگشتی است که ابتدا گفتارهای افراد متنوع را بصورت غیرخطی در فضای ورودی به گفتار نظیر از گوینده‌ای که به آن تعلیم یافته است، می‌برد و سپس آنرا بازشناسی می‌کند. این توانائی پالایش و فیلترسازی غیرخطی تأثیرات تنوعات (گوینده و غیره) از سیگنال ورودی، یک ویژگی برجسته این شبکه بازگشتی می‌باشد و آنرا به عنوان یک ابزار قوی برای پردازش غیرخطی سیگنال‌ها مطرح می‌کند.

جدول ۳- نتایج بازشناسی در حضور نویز ایستان (تعلیم با یک نفر تست با ۴۰ نفر)

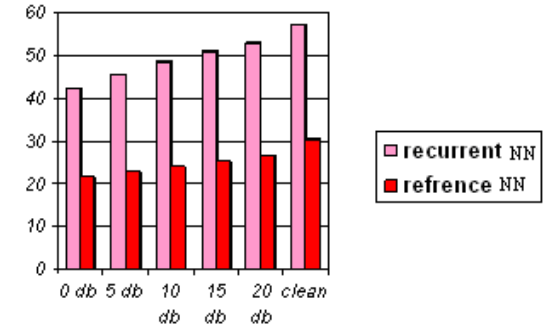
SNR	شبکه بازگشتی	شبکه مرجع
0 db	۴۶.۹۴۰۲	۱۷.۴۱
5 db	۵۰.۸۲۷۳	۱۸.۵۵۴۵
10 db	۵۳.۷۰۴۱	۲۰.۲۲۸۹
15 db	۵۵.۳۶۲۰	۲۲.۳۶۲۱
20 db	۵۶.۳۷۸۹	۲۵.۲۹۶۹
سیگنال تمیز	۵۷.۰۶۴۵	۳۰.۶۳



شکل ۱۲- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز ایستان (تعلیم با یک نفر تست با ۴۰ نفر)

جدول ۴- نتایج بازشناسی در حضور نویز غیرایستان (تعلیم با یک نفر تست با ۴۰ نفر)

SNR	شبکه بازگشتی	شبکه مرجع
0 db	۴۲.۴۰۲۲	۲۱.۷۳۴۴
5 db	۴۵.۵۳۰۷	۲۲.۸۵۴۰
10 db	۴۸.۵۱۰۲	۲۴.۰۶۶۳
15 db	۵۰.۹۸۷۹	۲۵.۲۸۸۶
20 db	۵۲.۸۵۶۱	۲۶.۴۰۴۹
سیگنال تمیز	۵۷.۰۶۴۵	۳۰.۶۳



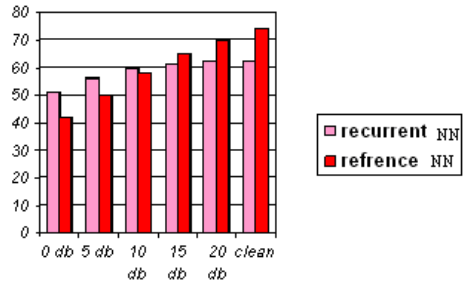
شکل ۱۳- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز غیرایستان (تعلیم با یک نفر تست با ۴۰ نفر)

۵-۲- تعلیم جاذب‌های پیوسته با گفتار تمیز از ۴۰ گوینده

پس از بررسی کامل شبکه و بدست آوردن مقادیر مناسب برای تعداد نورون‌های لایه پنهان، ضریب یادگیری و تعداد دوره‌های تعلیم بهینه، شبکه‌های عصبی مرجع و بازگشتی با دادگان زیاد تعلیم داده شدند. در این مرحله دادگان تعلیم ۴۰۰ جمله از ۴۰ گوینده مختلف و دادگان تست ۴۰۰ جمله دیگر از همین گویندگان بوده است که با نویزهای ایستان و غیرایستان و با نسبت‌های سیگنال به نویز مختلف نویزی شده‌اند. همچنین تعداد ی لایه پنهان به ۲۵۶ نورون افزایش یافت. نتایج بدست آمده در ادامه آمده است.

جدول ۵- نتایج بازشناسی در حضور نویز ایستان (تعلیم و تست با ۴۰ نفر)

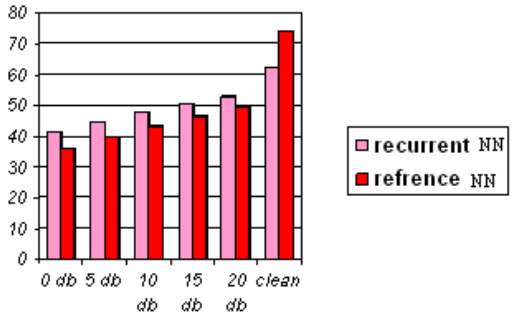
SNR	شبکه بازگشتی	شبکه مرجع
0 db	۵۱.۰۷	۴۲.۰۵۲
5 db	۵۶.۲۱۳۳	۴۹.۹۵۲۸
10 db	۵۹.۵۷۰۴	۵۸.۱۱۴۶
15 db	۶۱.۳۵۰۸	۶۵.۰۹۷۱
20 db	۶۲.۱۹۲۲	۶۹.۹۶۸۰
سیگنال تمیز	۶۲.۲۴	۷۴.۰۷



شکل ۱۴- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز ایستان (تعلیم و تست با ۴۰ نفر)

جدول ۶- نتایج بازشناسی در حضور نویز غیرایستان (تعلیم و تست با ۴۰ نفر)

SNR	شبکه بازگشتی	شبکه مرجع
0 db	۴۱.۵۵۹۲	۳۶.۳۵۵۴
5 db	۴۴.۷۴۵۷	۳۹.۹۵۲۶
10 db	۴۷.۸۷۷۶	۴۳.۳۷۴۴
15 db	۵۰.۵۶۲۳	۴۶.۵۷۹۱
20 db	۵۲.۹۸۲۰	۴۹.۶۹۴۴
سیگنال تمیز	۶۲.۲۴	۷۴.۰۷



شکل ۱۵- مقایسه عمکرد شبکه بازگشتی و مرجع در حضور نویز غیرایستان (تعلیم و تست با ۴۰ نفر)

نتایج نشان می‌دهد شبکه بازگشتی می‌تواند سیگنال نویزی شده با نسبت سیگنال به نویز صفر دسی‌بل را نسبت به شبکه مرجع ۶٪ بهتر بازشناسی کند، و این بدان معنی است که این شبکه توانسته است هدف ما که بهبود بازشناسی سیگنال گفتار نویزی است را برآورده کند. اما در مورد سیگنال تمیز شبکه کارائی شبکه مرجع را نداشته است و در حدود ۱۲٪ ضعیفتر عمل کرده است. شاید بتوان علت این بازشناسی ضعیف را تعلیم ناقص شبکه دانست که در این صورت باید به دنبال الگوریتم‌ها و روش‌هائی برای بهتر کردن همگرائی شبکه باشیم. همچنین حساسیت بیشتر شبکه بازگشتی نسبت به تنوع گوینده می تواند دلیل این امر باشد، که در این صورت می‌توان با استفاده از یک شبکه برای هنجارسازی دادگان نسبت به تنوع گوینده درصد صحت بازشناسی را بالا برد.

عملکرد ضعیف شبکه بازگشتی را می‌توان اینگونه بیان کرد که در تعلیم دادگان زیاد به شبکه بازگشتی، در فضای دینامیک‌های شبکه، تعداد جاذب‌ها (قعرها) زیاد می‌شود، یعنی در واقع با افزایش تعداد گوینده‌ها جاذب‌ها مغشوش شده‌اند و بازشناسی ضعیفتر شده است. این مسئله این فرضیه را نیز قوی‌تر می‌کند که می‌گوید انسان در بازشناسی صحبت‌های تنوع‌دار ابتدا آن را به صحبت متناظر از خود گوینده تصویر می‌کند و کاهش چند به یک انجام می‌دهد و سپس عمل بازشناسی آن را انجام می‌دهد.

۶- نتیجه گیری

در این تحقیق با استفاده از شبکه‌های عصبی بازگشتی غیرخطی و توانائی آنها در بازیابی اطلاعات و نیز با استفاده از جاذب‌های پیوسته و خواص آنها سعی در حذف نویز از سیگنال گفتار و بازشناسی آوا از روی سیگنال‌های نویزی پاکسازی شده، داشتیم. ساختار پیشنهادی توانست درصد صحت بازشناسی در نسبت سیگنال به نویز صفر دسی‌بل را نسبت به شبکه مرجع ۲۰٪ بهبود دهد. آزمایشات نشان می‌دهند که در تعلیم دادگان مختلف به شبکه‌های بازگشتی، این دادگان به عنوان قعرهای شبکه و نقاط جاذب در شبکه شکل می‌گیرند و شبکه هر الگوی ورودی را به سمت یکی از این قعرها حرکت می‌دهد. با دور زدن در شبکه، امکان حرکت به سمت قعر و تشخیص الگوی جاذب در خروجی فراهم می‌شود.

در آزمایشات شبکه با دادگان تمیز تعلیم می‌بیند و نحوه به قعر رفتن دادگان را نیز یاد می‌گیرد. به این ترتیب که برای هر ورودی در هنگام تعلیم، شبکه آنقدر دور می‌زند تا به بهترین خطای ممکن دست پیدا کند. با این نحوه تعلیم در واقع شبکه ورودی را از روی همان ورودی در لحظه قبل اصلاح می‌کند و این اصلاح را آنقدر ادامه می‌دهد تا به یک حالت بهینه یا به عبارتی به جاذب آن ورودی برود. بنابراین هنگامی که ورودی نویزی به شبکه داده می‌شود، شبکه با چند بار دور زدن ورودی را که در اثر نویز از قعر خود خارج شده است به سمت جاذب هدایت می‌کند.

در انتها دادگان تست را زیاد کردیم تا عملکرد مدل را در حذف تنوعات ورودی، عملاً مورد ارزیابی قرار دهیم. نتایج بدست آمده نشان می‌دهند که شبکه مرجع در این آزمایشات عملکرد بسیار ضعیفی داشته است و درصد صحت بازشناسی آن در مورد دادگان نویزی با نویز صفر دسی‌بل تنها ۱۷٪ بوده است. و این درصد در مورد دادگان تمیز به ۳۰/۶٪ رسیده است. در حالی که در شبکه بازگشتی در صد صحت بازشناسی سیگنال نویزی ۳۰٪ و سیگنال تمیز ۲۷٪ نسبت به شبکه مرجع بهبود داشته است. این نشان می‌دهد که جاذب‌های شکل گرفته در شبکه بازگشتی در هنگام بازشناسی حذف تنوعات و نویز را همزمان انجام داده‌اند و شبکه برای حذف تنوعات گویندگان نیز کارائی مطلوبی داشته است.

پس از آن علاوه بر دادگان تست، نفرات گویندگان در دادگان تعلیم را نیز زیاد کردیم. چون در گفتار یک نفر در دادگان فارس‌دات همه تنوعات آوائی وجود ندارد

و ممکن است با صحبت یک نفر شبکه‌ها خوب تعلیم ندیده باشند و نتوانند همه تنوعات آوائی در صحبت افراد زیاد را بازشناسی کنند.

نتایج بدست آمده نشان دادند که تعلیم دادگان زیاد به شبکه مرجع قدرت بازشناسی آن را بسیار بالا برده است و این شبکه نسبت به حالتی که تنها با جملات یک نفر تعلیم می‌دید در حدود ۳۵٪ بهبود داشته است. شبکه بازگشتی نیز می‌تواند سیگنال نویزی شده با نسبت سیگنال به نویز صفر دسی‌بل را نسبت به شبکه مرجع ۶٪ بهتر بازشناسی کند، و این نشان می‌دهد که این شبکه توانسته است هدف ما که بهبود بازشناسی سیگنال گفتار نویزی است را تا حدودی برآورده کند. اما در مورد سیگنال تمیز شبکه کارائی شبکه مرجع را نداشته است و در حدود ۱۲٪ ضعیفتر عمل کرده است. شاید بتوان علت این بازشناسی ضعیف را تعلیم ناقص شبکه دانست که در این صورت باید به دنبال الگوریتم‌ها و روش‌هائی برای بهتر کردن همگرائی شبکه باشیم.

آنگونه که در آزمایشات مشاهده شد، نیازی نیست که همان عبارت گفتار نویزی شده عیناً به شبکه تعلیم داده شده باشد، بلکه کافی است که زیر فضای گفتار تمیز توسط تعدادی جملات کافی دیگر از گوینده مورد نظر، تعلیم داده شده باشد. به عبارت دیگر، جاذب پیوسته می‌بایستی بیانگر زیرفضای گفتار گوینده مورد نظر باشد.

به بیان دیگر می توان گفت که این جاذب پیوسته توسط کنار هم چیده شدن مؤلفه‌های پایه گفتار گوینده مورد تعلیم، شکل گرفته است و این مؤلفه‌ها هستند که هر یک به عنوان یک جاذب نقطه ای، در بازشناسی مقاوم الگوهای ورودی ایفای نقش می‌کنند. ما فکر می کنیم که یک علت توانایی بالای مغز انسان در بازشناسی مقاوم الگوها، احتمالاً بر مبنای روشی مشابه این شیوه کاهش بعد غیرخطی الگوهای ورودی است.

مراجع

[1] S. Furui, "Robust Methods in Automatic Speech Recognition and Understanding," *Proc. EUROSPEECH*, Vol. 3, pp.1993-1998, 2003.

[2] R. P. Lippmann, "Speech Recognition by Machines and Humans," *Speech Communication*, Vol. 22, No. 1, pp. 1-15, 1997.

[3] R. P. Lippmann, and B. A. Carlson, "Using Missing Feature Theory to Actively Select Features for Robust Speech Recognition with Interruptions, Filtering and Noise," *Proc. EUROSPEECH*, Vol. 1, pp. 37-40, 1997.

[4] G. A. Miller, and J. C. R. Licklider, "The intelligibility of interrupted speech", *Journal of the Acoustic Society of America*, Vol. 22, pp. 167-173, 1950.

[5] Sh. Parveen, *Connectionist Approaches to the Deployment of Prior Knowledge for Improving Robustness in Automatic Speech Recognition*, Ph. D. thesis, Department of Computer Science, The University of Sheffield, UK, 2003.

[6] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of IEEE*, Vol. 80, No. 10, 1pp. 1526-1555, 1992.

[7] M. Jones, and S. sridharan, "Improving the effectiveness of existing noise reducing techniques using neural networks," *Proceeding of signal processing*, Vol. 1, pp. 387-388, 1996.

[8] M. Gaafar, K. Saleh, and M. Niranjan, "Speech enhancement in a Bayesian framework," *proc. ICASSP*, Vol. 1, pp. 389-392, 1998.

[9] M. Berouti, R. Schwarts, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *proc. ICASSP*, Vol. 4, pp. 208-211, 1997.

[10] P. Lockwood, "Noise reduction for speech enhancement in cars: nonlinear spectral subtraction/kalman filtering," *Proc. EUROSPEECH*, Vol. 1, pp. 83-86, 1991.

[11] I. Nejadgholi, and S. A. Seyyedsalehi, "Nonlinear normalization of input patterns to speaker variability in speech recognition neural networks", *Neural Computing and Applications*, Vol. 18, pp. 45-55, 2009.

[12] S. Tamura, and M. Nakamura, "Improvements to the noise reduction neural network," *proc. ICASSP*, Vol. 2, pp. 825-828, 1990.

[13] S. Moon, and J. Hwang, "Coordinated training of noise removing networks," *proc. ICASSP*, Vol. 1, 573-576, 1993.

[14] G. E. Hinton, and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, Vol. 313, No. 5786, pp. 504 – 507, 2006.

[15] M. Scholz, F. Kaplan, C. Guy, J. Kopka, and J. Selbig, "Non-linear PCA: a missing data approach," *Bioinformatics*, Vol. 21, No. 20, pp. 3887-3895, 2005.

[16] S. Parveen, and P. D. Green, "Speech enhancement with missing data techniques using R. N. N," *NIPS*, 2003.

[17] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *In Proceedings of the National Academy of Sciences*, Vol. 79, No. 8, pp. 2554-2558, 1982.

[18] D. H. Ackley, G. E. Hinton, and T. Sejnowski, "A learning algorithm for Botlzman Machines," *Cognitive Science*, Vol. 9, pp. 147-169, 1985.

[۱۹] س. ع. سیدصالحی، ا. نژادقلی، و ف. توحیدخواه، "افزایش کارائی بازشناخت الگوی شبکه‌های عصبی جلوسو از طریق توسعه روشهائی برای دوسویه کردن عملکرد آنها،" دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، مهر ۱۳۸۳.

[۲۰] ل. دهیادگاری، *بهبود کیفیت سیگنال گفتار آغشته به نویز و اعوجاج توسط شبکه‌های عصبی*، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، ۱۳۸۴.

[21] M. D. Skoronski, *Biologically inspired noise robust speech recognition for both man and machine*, Ph. D. thesis, University of Florida, USA, 2004.

[22] D. J. Amit, *Modeling brain function*, Cambridge University Press, 1989.

[23] C. Koch, J. Marroquin, and A. Yuille, "Analog "neuronal" networks in early vision," *Proceedings of the*

National Academy of Sciences,", Vol. 83, pp. 4263–4267, 1986.

[24] L. K. Saul, and M. I. Jordan, "Attractor Dynamics in Feed Forward Neural Networks," *Neural Computation*, Vol. 12, No. 6, pp. 1313-1335, 2000.

[25] M. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," *Speech Communication*, Vol. 34, No. 3, pp. 267-285, 2001.

[26] T. Eltoft, and Q. Kristiansen, "ICA and Nonlinear Time series Prediction for Recovering Missing Data Segments in Multivariant Signals," *Proc. ICA*, pp. 716-721, 2001.

[27] S. Parveen, and P. D. Green, "Speech recognition with missing data using Recurrent Neural Nets," *NIPS*, 2001.

[28] S. Parveen, and P. Green, "Speech enhancement with missing data techniques using recurrent neural networks," *ICASSP*, Vol. 1, pp. 17-21, 2004.

[29] Y. Bengio, and F. Gingras, "Recurrent Neural Networks for Missing or Asynchronous Data," in *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge, MA, 1996.

[30] M. Bijankhan, J. Sheikhzadegan, M. R. Roohani, Y. Samareh, C. Lucas, and M. Tebyani, "FARSDAT-The speech database of farsi spoken language," *Proc. SST*, pp. 826-831, 1994.

[۳۱] م. رحیمی‌نژاد، س. ع. سیدصالحی، "مقایسه و ارزیابی کارائی انواع روش‌های استخراج پارامترهای بازنمائی و هنجارسازی در بازشناسی مستقل از گوینده گفتار،" نشریه علمی پژوهشی امیرکبیر، سال چهاردهم، شماره آ-۵۵، تابستان ۱۳۸۲.

[۳۲] م. رحیمی نژاد، بهبود کارآیی روش‌های استخراج پارامترهای بازنمایی در سیستم‌های بازشناسی گفتار، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، ۱۳۸۱.

سیدعلی سیدصالحی مدرک کارشناسی خود را در مهندسی برق از دانشگاه صنعتی شریف در سال ۱۳۶۱، کارشناسی ارشد را در مهندسی برق از دانشگاه صنعتی امیرکبیر در سال ۱۳۶۷ و دکتری خود را در مهندسی برق - بیوالکتریک از دانشگاه تربیت مدرس در سال ۱۳۷۴ دریافت نموده است. وی در حال حاضر استادیار دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر می‌باشد. زمینه‌های پژوهشی مورد علاقه ایشان پردازش و بازشناسی گفتار، شبکه‌های عصبی مصنوعی و زیستی، مدل سازی عملکرد مغز و پردازش خطی و غیرخطی سیگنال می‌باشد.
آدرس پست‌الکترونیکی ایشان عبارت است از:
.....

ایثار نژادقلی مدرک کارشناسی خود را در مهندسی برق - کنترل از دانشگاه صنعتی شریف در سال ۱۳۸۰ و مدرک کارشناسی ارشد را در مهندسی پزشکی-بیوالکتریک از دانشگاه صنعتی امیرکبیر در سال ۱۳۸۲ دریافت کرده است. وی اکنون مشغول به تحصیل در دوره دکتری مهندسی پزشکی - بیوالکتریک در دانشگاه صنعتی امیرکبیر می‌باشد. زمینه‌های مورد علاقه او پردازش سیگنال با بهره‌گیری از روش‌های هوش مصنوعی، مدل سازی عملکرد مغز و فن‌آوری شبکه‌های عصبی مصنوعی می‌باشد.
آدرس پست‌الکترونیکی ایشان عبارت است از:
.....

---

[1] Stationary Noise
[2] Nonstationary Noise
[3] Smooth
[4] Recurrent Neural Network (RNN)
[5] Firing
[6] Nofiring
[7] Attractors
[8] Correlations
[9] Quantize
[10] Continuous Attractor
[11] Point Attractor
[12] Trajectory
[13] Variability
[۱۴] در توپولوژی، مانیفولد به یک فضای هندسی اطلاق می‌شود که هر نقطه روی آن دارای یک همسایگی هوموموفیک در فضای n بعدی R^n می‌باشد.
[15] Additive Noise
[16] Epoch
[17] Multi-Task Learning
[18] Data Base
[19] Logarithm of Hanning Critical Band Filter Bank (LHCB)
[20] Normalization

لوئیزا دهیادگاری کارشناسی خود را در رشته مهندسی پزشکی- بالینی از دانشگاه اصفهان در سال ۱۳۸۲ و کارشناسی ارشد را در رشته مهندسی پزشکی-بیوالکتریک از دانشگاه صنعتی امیرکبیر در سال ۱۳۸۴ دریافت نموده است. او در حال حاضر مدرس دانشگاه‌های آزاد و تکنولوژی شهرستان سیرجان می‌باشد. زمینه‌های پژوهشی مورد علاقه او پردازش سیگنال‌های حیاتی، پردازش گفتار، شبکه‌های عصبی مصنوعی زیستی و شبکه‌های عصبی کوانتومی می‌باشد.
آدرس پست‌الکترونیکی ایشان عبارت است از:
.....