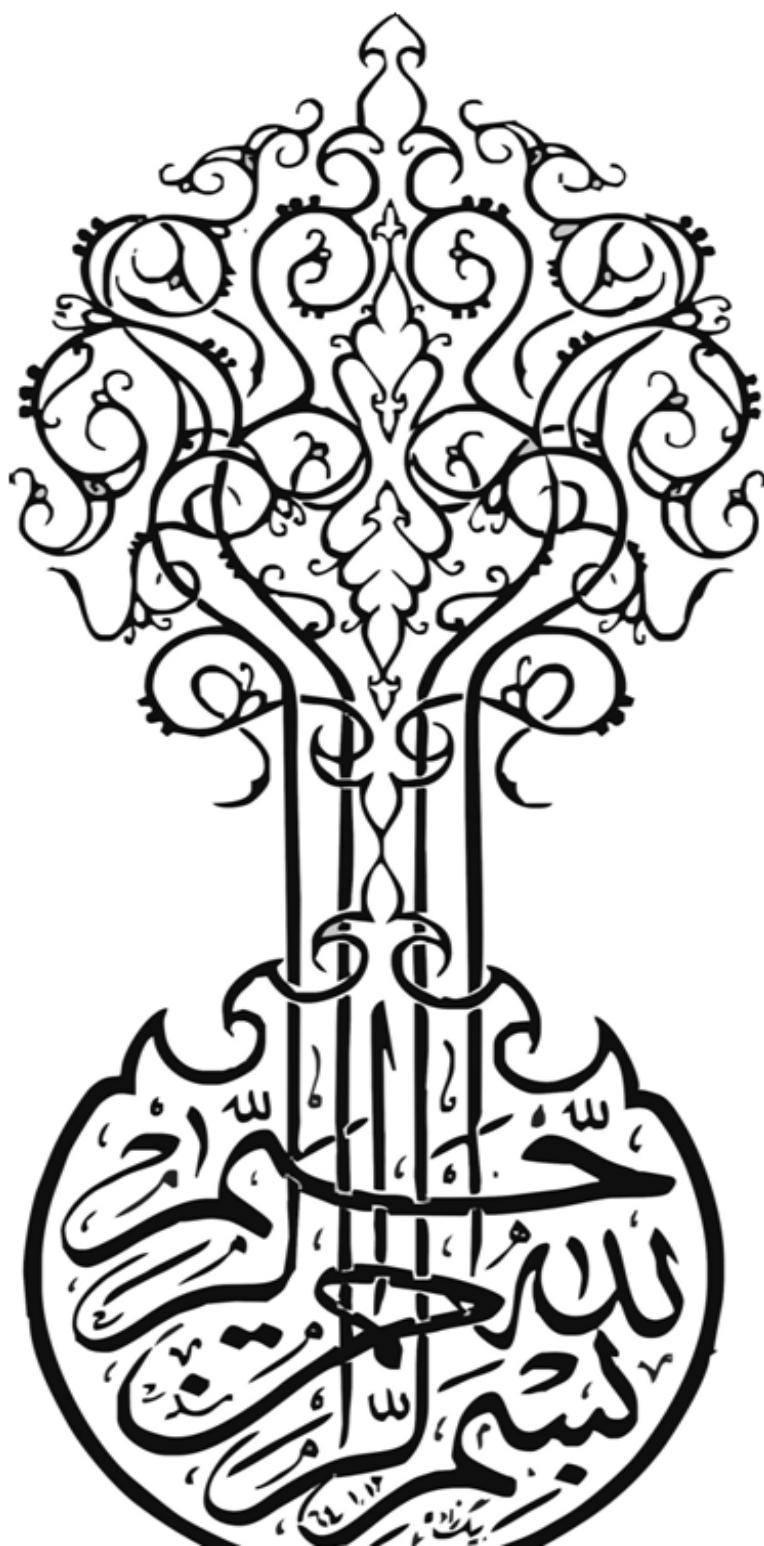






دانشکده : برق و رباتیک

گروه : الکترونیک

بهسازی گفتار با استفاده از Wavenet

دانشجو : سمیرا مغانی

استاد راهنما :

جناب آقای دکتر حسین مروی

استاد مشاور :

دکتر علیرضا احمدی فرد

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

ماه و سال انتشار :

شهریور ۱۳۹۱

شماره : ۷۰۹۷
تاریخ : ۹۱/۰۹/۲۸
ویرایش :

بسمه تعالی



فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند معال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم (اذی) :
ممیرامضانی رشته : برق گرایش : الکترونیک
تحت عنوان : بهسازی گفتار با استفاده از Wavenet
که در تاریخ ۹۱/۰۶/۲۸ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول (یا درجه : خیلی)	اعتبار ۱۹/۲۸	☒	دفاع مجدد	☐	مردود	☐
-----------------------	--------------	-------------------------------------	-----------	--------------------------	-------	--------------------------

۱- عالی (۲۰ - ۱۹) ۲- بسیار خوب (۱۸ - ۱۵/۹۹)

۳- خوب (۱۲/۹۹ - ۱۶) ۴- قابل قبول (۱۵/۹۹ - ۱۴)

۵- نمره کمتر از ۱۴ غیر قابل قبول

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضا
۱- استاد راهنما	علیرضا ارمی	استادیار	
۲- استاد مشاور	عباسرضا احمدی	استادیار	
۳- نماینده شورای تحصیلات تکمیلی	سرتیغ رحیمی	استادیار	
۴- استاد متحن	ایرج محمدعزیزی	استادیار	
۵- استاد متحن	حاجی‌نژاد	استادیار	

رئیس دانشکده :

تقدیم به:

بزرگترین پشتوانه‌های زندگی‌ام

پدر عزیزم

مادر مهربانم

و

همسر وفادارم

که تا هر کجای دنیای علم هم اگر پیش بروم

باز هم رفوزه کلاس اول مدرسه عشق

محبت

صفا و

صمیمیت آنانم.

"من علّمنی حرفاً فقد سیّرنی عبداً؛ هر کس چیزی به من بیاموزد، مرا بنده خویش کرده است."

حضرت علی (ع)

تشکر و قدردانی

در ابتدا، سپاس و ثنای بی حد خود را بر آستان صفات بی‌همتای احدیت که در کمال رافت و در نهایت عطوفت رخصت اتمام این پایان نامه را به بنده عطا فرموده است، نثار خواهم کرد.

از پدر و مادر مهربان و فداکارم تشکر می‌کنم که در تمام طول عمر، مرا مانند کوهی مستحکم در برابر سختی‌های زندگی پرورش دادند و دستان پرمهرشان را می‌بوسم و از خداوند منان می‌خواهم که سایه پر برکتشان را همیشه برای من حفظ کند.

تشکر ویژه‌ای از همسر دلسوز و همیشه همراهم دارم که همواره مانند خورشیدی درخشان گرمابخش و امیدبخش تمامی لحظات زندگیم بوده و حس زیبای خوشبختی را برایم معنا بخشید.

بر خود واجب می‌دانم مراتب سپاس و قدردانی خویش را نثار بزرگوارانی کنم که از هیچ‌گونه بذل راهنمایی‌های عالمانه و دقیق خود، در جهت پر بار شدن محتوای این پژوهش، دریغ نورزیدند. از استاد راهنمای گرامی جناب آقای دکتر حسین مروی، کسی که همیشه مرا از راهنمایی‌های ارزنده و همفکری‌ها و حمایت‌های علمی‌شان بهره‌مند نمودند، بسیار سپاسگزارم. از اساتید محترم جناب آقای دکتر معروضی و جناب آقای دکتر گرایلو که زحمت مطالعه و داوری این پایان‌نامه را برعهده گرفتند، کمال تشکر را دارم.

همچنین قدردان اساتید و معلمان محترمی که مرا از ابتدای تحصیل تا بدین لحظه به سوی پیشرفت و علم‌اندوزی هدایت کردند خواهم بود و امیدوارم در پناه حق همیشه پایدار و مستدام باشند. از تمامی دوستان عزیز و بزرگوارانی که مرا از کمک‌های بی‌دریغشان بی‌نصیب نگذاشتند ممنون و سپاسگزارم.

تعهد نامه

اینجانب سمیرا مغانی دانشجوی دوره کارشناسی ارشد رشته الکترونیک دانشکده برق و ریاتیک دانشگاه صنعتی شاهرود نویسنده پایان نامه

بهسازی گفتار با استفاده از ویونت تحت راهنمایی جناب آقای دکتر حسن مروی متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

بهسازی گفتار یا Speech enhancement بیانگر گروه بزرگی از روش‌هاست که با انجام پردازش‌هایی روی سیگنال‌های نویزی، نهایتاً منجر به بهبود کیفیت و قابلیت فهم گفتار می‌شود.

تاکنون چندین روش در زمینه بهسازی گفتار نویزی ارائه شده است. روش پیشنهادی ما در این پایان‌نامه، شبکه عصبی موجک (ویونت) می‌باشد.

این روش، مدلی مبتنی بر تلفیق شبکه عصبی با تبدیل موجک می‌باشد و به عنوان جایگزین مناسب در شبکه‌های عصبی پس‌خورد جهت تخمین و تقریب توابع غیرخطی اختیاری پیشنهاد می‌شود.

در شبکه استاندارد پس‌خورد تابع فعال سازی نرون لایه پنهان یک تابع سیگموئید است. حال آنکه شبکه‌های عصبی موجک توابع موجک را به عنوان توابع فعال‌سازی نرون‌های لایه پنهان از شبکه پس‌خور مورد استفاده قرار می‌دهند. در روش پیشنهادی دو حالت برای دسترسی به نویز در نظر گرفته شد. حالت الف، حالتی است که دسترسی مستقیم به نویز محیط وجود دارد و حالت ب، حالتی است که از تخمین نویز استفاده کرده‌ایم. روش پیشنهادی برای تخمین نویز، استفاده از تبدیل موجک می‌باشد. موجک استفاده شده در این مرحله دابیشز ۵ (dB5) می‌باشد. روش پیشنهادی برای دادگان فارسی پیاده سازی شده است و برای ارزیابی و عملکرد آن از سه معیار سیگنال به نویز، سیگنال به نویز قطعه‌ای و Log Likelihood Ratio (LLR) استفاده شده است. در نهایت روش پیشنهادی با شبکه عصبی و ترکیب شبکه عصبی با موجک و روش موجک مقایسه شده است. بررسی نتایج و آزمایشات نشان‌دهنده عملکرد قابل قبول روش پیشنهادی است.

کلمات کلیدی:

بهسازی گفتار، نویز، ویونت، تبدیل موجک

- 1) “Improving The Performance Of Speech Enhancement By Using Wavenet “ is accepted for publication in International Journal of Computational Engineering Research (IJCER)
- 2) “Speech Enhancement By Using Wavenet” is submitted to 4th International Conference on Signal Processing Systems - Kuala Lumpur-ICSPS 2012

فهرست مطالب

عنوان	صفحه
۱ مقدمه	۲
۱-۱ مقدمه ای بر بهسازی گفتار	۲
۲-۱ کاربردهای سیستم های بهسازی گفتار	۳
۳-۱ تاریخچه روش های بهسازی گفتار	۳
۴-۱ ساختار پایان نامه	۵
۲ بهسازی گفتار و مروری بر روش های آن	۷
۱-۲ مقدمه	۷
۲-۲ بهسازی گفتار (Speech enhancement)	۷
۳-۲ منابع و مشخصات نویزهای صوتی	۹
۴-۲ طبقه بندی روش های بهسازی گفتار	۱۰
۵-۲ عوامل موثر در طراحی سیستم های بهسازی گفتار	۱۱
۶-۲ ارزیابی سیستم های بهسازی گفتار	۱۲
۱-۶-۲ معیار سیگنال به نویز (SNR)	۱۳
۲-۶-۲ معیار سیگنال به نویز قطعه‌های (SNR _{seg})	۱۴
۳-۶-۲ معیار سیگنال به نویز قطعه‌های وزندار (SNR _{fw-seg})	۱۴
۴-۶-۲ معیار PESQ	۱۵
۵-۶-۲ معیار LLR	۱۶
۷-۲ تشخیص گفتار از نویز	۱۶
۸-۲ روش های تخمین نویز در بهسازی گفتار	۱۷
۹-۲ مروری بر روش های بهسازی گفتار	۱۸
۱-۹-۲ تفریق طیفی	۱۸
۲-۹-۲ زیرفضای سیگنال	۲۱

۲۱	۳-۹-۲ فیلتر وینر
۲۳	۴-۹-۲ روش های آماری
۲۴	۵-۹-۲ فیلتر وفقی
۲۸	۳ مروری بر شبکه عصبی و تبدیل موجک
۲۸	۱-۳ مقدمه
۲۸	۲-۳ شبکه عصبی
۳۰	۱-۲-۳ آموزش شبکه به روش پس انتشار خطا
۳۱	۲-۲-۳ الگوریتم پس انتشار خطا
۳۳	۳-۲-۳ فواید شبکه عصبی
۳۳	۴-۲-۳ محدودیتهای شبکه عصبی
۳۳	۳-۳ بهسازی گفتار با استفاده از شبکه عصبی
۳۴	۱-۳-۳ مثالی از بهسازی گفتار با استفاده از شبکه عصبی
۳۷	۴-۳ تبدیل موجک
۳۹	۱-۴-۳ انواع توابع موجک
۳۹	۱-۱-۴-۳ Morlet تابع موجک
۳۹	۲-۱-۴-۳ تابع موجک (RASP (RAtional functions with Second-order Poles
۴۰	۳-۱-۴-۳ تابع موجک (SLOG (Superposed LOGistic functions
۴۱	۴-۱-۴-۳ توابع موجک (POLYWOG (POLYnomials WindOwed with Gaussian
۴۲	۵-۱-۴-۳ Shannon تابع موجک
۴۳	۵-۳ بهسازی گفتار با استفاده از تبدیل موجک
۴۴	۱-۵-۳ مثالی از کاربرد تبدیل موجک در بهسازی گفتار
۴۷	۴ شبکه ویونت (wavelet neural network)
۴۷	۱-۴ مقدمه
۴۷	۲-۴ معرفی شبکه ویونت
۴۹	۱-۲-۴ شبکه ویونت تک بعدی

۵۰	۲-۲-۴ شبکه ویونت چند بعدی
۵۱	۳-۲-۴ الگوریتم آموزش
۵۴	۵ شبکه ی پیشنهادی و بکارگیری آن در بهسازی گفتار
۵۴	۱-۵ مقدمه
۵۵	۲-۵ شبکه ویونت
۵۵	۱-۲-۵ شبکه پیشنهادی
۵۷	۱-۱-۲-۵ مکانیزم آموزش on-line
۵۹	۲-۱-۲-۵ آنالیز همگرایی
۵۹	۳-۵ منبع نویز
۶۱	۱-۳-۵ تخمین نویز
۶۱	۱-۱-۳-۵ تبدیل موجک گسسته
۶۳	۲-۱-۳-۵ تخمین نویز با استفاده از تبدیل موجک گسسته
۶۴	۳-۱-۳-۵ آستانه گذاری روی ضرایب موجک
۶۴	۲-۳-۵ انتخاب آستانه
۶۷	۶ نتایج شبیه سازی
۶۷	۱-۶ مقدمه
۶۷	۲-۶ معرفی پایگاه داده
۶۷	۳-۶ فرضیات
۶۸	۴-۶ مراحل بهسازی گفتار در روش پیشنهادی
۶۸	۱-۴-۶ نویز
۷۱	۱-۱-۴-۶ انتخاب آستانه
۷۳	۵-۶ نتایج آموزش شبکه ویونت با توابع موجک مختلف
۷۴	۶-۶ نتایج آموزش شبکه عصبی ویولت با تعداد نورونهای متفاوت در لایه ها
۷۷	۷-۶ بررسی نمودار و منحنی های حاصل از پیاده سازی روش پیشنهادی با پارامترهای مطلوب
۸۵	۸-۶ ارزیابی روش پیشنهادی

۸۵.....	۱-۸-۶ معیار ارزیابی سیگنال به نویز با استفاده از روش الف
۸۵	۱-۱-۸-۶ معیار ارزیابی سیگنال به نویز
۸۷	۲-۱-۸-۶ معیار ارزیابی سیگنال به نویز قطعه ای
۸۹.....	۳-۱-۸-۶ معیار ارزیابی LLR
۹۰	۲-۸-۶ معیار ارزیابی سیگنال به نویز با استفاده از روش ب
۹۰	۱-۲-۸-۶ معیار ارزیابی سیگنال به نویز
۹۱.....	۲-۲-۸-۶ معیار ارزیابی سیگنال به نویز قطعه ای
۹۲	۹-۶ مقایسه روش پیشنهادی با روش های دیگر
۹۵	۱-۹-۶ معیار سیگنال به نویز
۱۰۰	۲-۹-۶ معیار سیگنال به نویز قطعه ای
۱۰۴	۳-۹-۶ معیار MSE
۱۰۵	۱۰-۶ نتیجه گیری
۱۰۶.....	۱۱-۶ پیشنهادات
۱۰۷.....	منابع

فهرست شکل ها

- شکل ۱-۲: بلوک دیاگرام حذف نویز..... ۷
- شکل ۲-۲: نویز جمع شونده پس زمینه..... ۹
- شکل ۳-۲: بلوک دیاگرام تفریق طیفی..... ۱۹
- شکل ۴-۲: حذف نویز با فیلتر وینر..... ۲۲
- شکل ۵-۲: نمای کلی روشهای آماری..... ۲۴
- شکل ۶-۲: حذف کنندهی نویز با فیلتر وقتی..... ۲۵
- شکل ۱-۳: طبقه بندی ساختارهای مختلف شبکه عصبی..... ۲۹
- شکل ۲-۳: شبکه عصبی Feed-Forward چند لایه..... ۳۰
- شکل ۳-۳: شبکه عصبی..... ۳۴
- شکل ۴-۳: نمای کلی از شبکه آدالاین..... ۳۵
- شکل ۵-۳: بلوک دیاگرام بهسازی با شبکه آدالاین..... ۳۶
- شکل ۶-۳: سیگنال گفتار تمیز..... ۳۶
- شکل ۷-۳: سیگنال گفتار آلوده به نویز..... ۳۷
- شکل ۸-۳: سیگنال گفتار بهبود یافته..... ۳۷
- شکل ۹-۳: $h(t)$ موجک Morlet کوسینوس-گوسین و تبدیل فوریه اش $H(w)$ ۳۹
- شکل ۱۰-۳: مثال هایی از موجک های RASP..... ۴۰
- شکل ۱۱-۳: تابع سیگموئید لجستیک..... ۴۰
- شکل ۱۲-۳: مثال هایی از موجک مادر SLOG..... ۴۱
- شکل ۱۳-۳: مثال هایی از موجک POLYWOG..... ۴۲
- شکل ۱۴-۳: تابع موجک Shanon..... ۴۳
- شکل ۱۵-۳: بلوک دیاگرام روش شبکه عصبی و موجک..... ۴۴
- شکل ۱۶-۳: سیگنال گفتار آلوده به نویز..... ۴۵
- شکل ۱۷-۳: سیگنال گفتار بهبود یافته..... ۴۵
- شکل ۱-۴: ساختار شبکه ویونت..... ۴۸
- شکل ۲-۴: ساختار یک نرون موجک..... ۴۹
- شکل ۳-۴: شبکه ویونت یک بعدی..... ۵۰
- شکل ۴-۴: شبکه ویونت چند بعدی..... ۵۰
- شکل ۱-۵: بلوک دیاگرام روش پیشنهادی..... ۵۵
- شکل ۲-۵: ساختار شبکه عصبی موجک چهار-لایه..... ۵۷
- شکل ۳-۵: بلوک دیاگرام حذف نویز با استفاده از نویز ضبط شده..... ۶۰

- شکل ۴-۵: بلوک دیاگرام حذف نویز با استفاده از تخمین نویز ۶۰
- شکل ۵-۵: شمای کلی تجزیه سیگنال در تبدیل موجک گسسته ۶۲
- شکل ۶-۵: شمای کلی تجزیه سیگنال در تبدیل بسته موجک گسسته ۶۲
- شکل ۷-۵: شمای کلی بازسازی در تبدیل موجک گسسته ۶۳
- شکل ۸-۵: عملیات فیلتر نمودن سیگنال با استفاده از تبدیل موجک ۶۴
- شکل ۱-۶: مثالی از موجکهای دابیشز ۶۸
- شکل ۲-۶: فیلترهای بالاگذر و پایین گذر و دوگان آنها ۷۰
- شکل ۳-۶: ۹ مرحله تجزیه گفتار نویزی به کمک موجک دابیشز ۷۱
- شکل ۴-۶: خلاصه روند طی شده در تخمین نویز روش پیشنهادی ۷۲
- شکل ۵-۶: مقایسه نویز اصلی و نویز تخمین زده شده ۷۲
- شکل ۶-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز سفید گوسی (b) و گفتار بهبود یافته (c) ۷۸
- شکل ۷-۶: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز سفید گوسی (b) و گفتار بهبود یافته (c) ۷۹
- شکل ۸-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c) ۸۱
- شکل ۹-۶: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c) ۸۲
- شکل ۱۰-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز همه (b) و گفتار بهبود یافته (c) ۸۳
- شکل ۱۱-۶: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c) ۸۴
- شکل ۱۲-۶: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف ۸۶
- شکل ۱۳-۶: مقایسه SNRseg سیگنال خروجی به ازای segSNR ورودی مختلف ۸۸
- شکل ۱۴-۶: مقایسه LLR سیگنال خروجی به ازای LLR ورودی مختلف ۸۹
- شکل ۱۵-۶: شمای کلی شبکه عصبی آدالاین [۲۲] ۹۳
- شکل ۱۶-۶: روند طی شده در بهسازی گفتار با شبکه عصبی آدالاین [۲] ۹۳
- شکل ۱۷-۶: روند طی شده در بهسازی گفتار با شبکه عصبی آدالاین و موجک [۲] ۹۴
- شکل ۱۸-۶: روند طی شده در بهسازی گفتار با روش تبدیل موجک ۹۵
- شکل ۱۹-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی ۹۶
- الف وب بر حسب SNR برای نویز سفید گوسی ۹۶
- شکل ۲۰-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی ۹۷
- الف وب بر حسب SNR برای نویز همه ۹۷
- شکل ۲۱-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی ۹۷

- الف وب بر حسب SNR برای نویز F16 ۹۹
- شکل ۶-۲۲: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی
- الف وب بر حسب SNRseg برای نویز سفید گوسی ۱۰۰
- شکل ۶-۲۳: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی
- الف وب بر حسب SNRseg برای نویز همهمه ۱۰۲
- شکل ۶-۲۴: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی
- الف وب بر حسب SNRseg برای نویز F16 ۱۰۳

فهرست جدول ها

- جدول ۶-۱: SNR سیگنال خروجی به ازای n های مختلف $db(n)$ ۶۹
- جدول ۶-۲: SNR سیگنال خروجی به ازای تعداد مراحل تجزیه (j) ۶۹
- جدول ۶-۳: نتایج MSE شبکه ویونت با توابع موجک مختلف در ۵ مرتبه تکرار الگوریتم آموزش ۷۳
- جدول ۶-۴: اطلاعات شبکه ویونت با ۶ نورون در لایه دوم و ۳ نورون در لایه سوم ۷۴
- جدول ۶-۵: اطلاعات شبکه ویونت با ۸ نورون در لایه دوم و ۴ نورون در لایه سوم ۷۵
- جدول ۶-۶: اطلاعات شبکه ویونت با ۲۴ نورون در لایه دوم و ۱۲ نورون در لایه سوم ۷۶
- جدول ۶-۷: اطلاعات شبکه ویونت با ۴۸ نورون در لایه دوم و ۲۴ نورون در لایه سوم ۷۶
- جدول ۶-۸: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف ۸۷
- جدول ۶-۹: مقایسه SNRseg سیگنال خروجی به ازای SNRseg ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف ۸۸
- جدول ۶-۱۰: مقایسه LLR سیگنال خروجی به ازای LLR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف ۸۹
- جدول ۶-۱۱: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی ب ۹۱
- جدول ۶-۱۲: مقایسه SNRseg سیگنال خروجی به ازای SNRseg ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف ۹۱
- جدول ۶-۱۳: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۹۶
- پیشنهادی الف وب بر حسب SNR برای نویز سفید گوسی ۹۶
- جدول ۶-۱۴: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۹۸
- پیشنهادی الف وب بر حسب SNR برای نویز همهمه ۹۸
- جدول ۶-۱۵: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۹۹
- پیشنهادی الف وب بر حسب SNR برای نویز F16 ۹۹
- جدول ۶-۱۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۱۰۱
- پیشنهادی الف وب بر حسب SNRseg برای نویز سفید گوسی ۱۰۱
- جدول ۶-۱۷: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۱۰۲
- پیشنهادی الف وب بر حسب SNRseg برای نویز همهمه ۱۰۲
- جدول ۶-۱۸: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت ۱۰۳
- پیشنهادی الف وب بر حسب SNRseg برای نویز F16 ۱۰۳

جدول ۶-۱۹: مقایسه MSE روش پیشنهادی با روش های آدالاین، آدالاین و موجک و ویونت پیشنهادی
الف وب بر حسب تعداد تکرارهای مرحله آموزش ۱۰۴

فصل اول

مقدمه

۱-۱ مقدمه ای بر بهسازی گفتار

همان طور که می دانیم مهمترین وسیله ارتباطی افراد از طریق صدا و گفتار می باشد. به همین جهت پردازش گفتار، به جهت کاربردهای فراوان در زندگی بشر، از سالیان قبل مد نظر دانشمندان و محققین بوده است. پردازش گفتار شامل بخشهای متفاوتی مانند: کدینگ، تشخیص گفتار گوینده، سنتز، فشرده سازی، بهسازی گفتار یا speech enhancement می باشد.

یکی از مهمترین بخش های پردازش گفتار، بهسازی گفتار و تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری می باشد.

در قالب یک تعریف کلی، بهسازی گفتار عبارت است از تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری در مواردی که سیگنال گفتار تحت تاثیر نویز، انعکاس و سایر عوامل تخریبی واقع گردیده است. نویز از معضلات دنیای امروزی است. گسترش تکنولوژی و صنعت باعث ایجاد صداهایی با شدت زیاد در اطراف ما شده است. اگر شدت صدا از حد مجاز خود بالاتر برود، به افرادی که در مواجهه با آن هستند آسیب وارد می شود. البته شدت این آسیب با توجه به شدت صدا و مدت زمان مواجهه با آن متغیر است. صدای بلند، چه موسیقی باشد و چه صدایی آزاردهنده، می تواند بر شنوایی ما تأثیرات نامطلوبی داشته باشد.

شدت صدا را با دسی بل می سنجند، به عنوان مثال، شدت صدای ناشی از پرواز یک هواپیمای جت از فاصله ۱۰۰ متری برابر با ۱۳۰ دسی بل است. نیاز به بهسازی گفتار از آنجا ناشی می شود که سیگنال گفتار معمولاً، یا از منبعی تولید می شود که در محیط نویزی قرار دارد، یا تحت تاثیر کانال انتشار و در اثر نویز یا انعکاس دچار تخریب می شود، یا در گیرنده به نویز آلوده می شود و یا ممکن است منشأ نویز ترکیبی از سه حالت فوق باشد [۳۰].

۲-۱ کاربردهای سیستم های بهسازی گفتار

سیستم های بهسازی گفتار کاربردهای وسیع و متفاوتی در حوزه های مختلف دارند که به طور کلی می توان آنها را به صورت زیر طبقه بندی کرد:

- تشخیص خودکار گفتار^۱ و تشخیص هویت گوینده^۲: در سیستم های تشخیص خودکار گفتار و هویت گوینده، عملکرد سیستم در محیط هایی با سیگنال به نویز پایین مناسب نمی باشد. ثابت شده که نرخ تشخیص با به کارگیری الگوریتم های بهسازی گفتار بهبود می یابد [۲۸].

- سیستم های تلفن: که در آنها گفتار اصلی به وسیله نویز زمینه یا نویز موجود در مسیر مخابره و نیز در اثر انعکاس صدای طرفین مکالمه خراب می شود.

- تلفن های عمومی: که در محیط های پر سر و صدا و شلوغ واقع اند.

- سیستم های مخابرات هوا به زمین: که در آنها نویز اتاق خلبان، پیام ارسال شده از سوی خلبان را خراب می کند.

- سمعک: که به عنوان یک تقویت کننده، هم سیگنال گفتار و هم نویز موجود در محیط را تقویت نموده و موجب ناراحتی کاربر می شود.

- حذف نویز در گفتارهای ضبط شده: که در آنها صدای ضبط شده اشخاص آغشته به نویز است و جهت بهبود کیفیت آنها، باید نویز را حذف کرد.

- کدینگ اطلاعات: که اثر نویز بر گفتار به شدت عملکرد آنها را تحت تاثیر قرار می دهد [۴۸].

۳-۱ تاریخچه روش های بهسازی گفتار

با توجه به کاربرد های مهم و فراوان بهسازی گفتار، روش های مختلفی برای این کار پیشنهاد شده اند.

^۱ - Automatic Speech Recognition

^۲ - Speaker Identification

از مهمترین و قدیمی ترین روش های بهسازی گفتار می توان به فیلتر وینر [۱] و مشتقات آن اشاره کرد. در این روش ها، که برای اولین بار در سال ۱۹۴۹ وارد حوزه بهسازی گفتار گردید، سعی می شود که تخمین سیگنال تمیز بر اساس یک معیار بهینه سازی باشد.

در سال ۱۹۵۹، Widrow و Hoff فعالیت هایی در زمینه ی حذف نویز با استفاده از شبکه ی عصبی و توسعه شبکه Adaline انجام دادند [۲].

یکی دیگر از روش های قدیمی و بسیار معروف بهسازی گفتار، تفریق طیفی [۳] و [۴] است که در سال ۱۹۷۵ پایه گذاری شد و در سال ۱۹۷۹ با پردازش در حوزه فرکانس مطرح شد. پس از آن بهبودهای زیادی بر روی این روش صورت گرفت. فیلتر کالمن هم در سال ۱۹۸۷ با مقاله Paliwal & Basu [۵] وارد این حوزه گردید.

روش مهم دیگر بهسازی گفتار که مبتنی بر تئوری جبر خطی می باشد، زیر فضای سیگنال است که در ابتدا در [۶] و [۷] و [۸] و در سال ۱۹۹۱ ارائه گردید و سپس بهبودهای زیادی روی آن صورت پذیرفت. اخیرا روش های مبتنی بر کپسترال، تبدیل موجک و تبدیل فوریه کسینوسی نیز وارد حوزه بهسازی گفتار شده اند. تئوری حذف نویز در حوزه موجک توسط Donoho در سال ۱۹۹۵ مطرح گردید.

با توجه به کاربرد های مهم و فراوانی که برای روش های بهسازی گفتار بیان شد، و اهمیت آن، روش های مختلفی برای این کار پیشنهاد شده اند.

تا امروز مقالات و روش های بسیاری در زمینه بهسازی گفتار ارائه شده است، اما تخمین سیگنال تمیز از روی سیگنال نویزی هنوز یک مساله حل نشده در پردازش گفتار است و چالش های متعددی را در خود نگه داشته است. یکی از دلایل ساده برای اثبات این ادعا این است که علی رغم قدمت بیشتر بهسازی گفتار نسبت به برخی کاربردهای دیگر پردازش گفتار مثل بازشناسی گفتار، هنوز روش قطعی و غالبی در این زمینه وجود ندارد. از دلایل اصلی مشکل بودن حل این مساله، نامشخص بودن جواب دقیق برخی از سوالات کلیدی و ابهامات در این حوزه است. از جمله این سوالات عبارتند از: آیا می توان یک رابطه مشخص را برای معیارهای

کیفیت و قابلیت فهم، به صورتی که انسان درک می کند، به دست آورد؟ پاسخ به این سوال، به نوبه خود منجر به یافتن معیار بهینه سازی معنی داری می شود که با ادراک انسان متناسب است. با داشتن روابط مشخص برای این دو، می توان یکی از آنها را بدون داشتن اثر منفی بر دیگری بهبود بخشید و می توان به این سوال نیز پاسخ داد که چه روش هایی برای بهسازی گفتار در کاربرد کمک شنیداری مناسب است؟ کدام یک برای کاربرد بازشناسی گفتار و کدام یک برای کدینگ مناسب می باشد؟ چه پردازش هایی در سطوح مختلف سیستم شنوایی انسان انجام می گیرد و چگونه می توان آنها را مدل کرد؟ پاسخ این سوال نیز منجر به روشن شدن شباهت دیگری مانند نحوه ترکیب تکنیک های پردازش سیگنال با روش های ادراکی انسان می شود [۴۶].

۴-۱ ساختار پایان نامه

این پایان نامه شامل شش فصل می باشد. تعاریف اولیه و کلیاتی درباره بهسازی گفتار در فصل جاری مطرح شد. در فصل دوم مختصراً به بررسی بهسازی گفتار، عوامل موثر در طراحی سیستم های بهسازی گفتار و چگونگی ارزیابی سیستم های بهسازی و تاریخچه ای از روشهای بیان شده در این زمینه خواهیم پرداخت. فصل سوم نیز مروری بر شبکه عصبی و تبدیل موجک و کاربرد آنها در بهسازی گفتار می باشد. در فصل چهارم به معرفی ویونت (شبکه عصبی موجک) پرداخته ایم. فصل پنجم شامل بیان شبکه ی پیشنهادی و نحوه بکارگیری آن در بهسازی گفتار می باشد. در فصل ششم که فصل آخر نیز می باشد، نتایج شبیه سازی روش پیشنهادی و پیشنهادات برای کارهای آینده بیان شده است. مقایسه روش پیشنهادی با سایر روشهای معروف بهسازی گفتار در این فصل آمده است و نتیجه گیری نهایی در این فصل بیان خواهد شد. هدف از انجام این پایان نامه و جهت گیری های اصلی آن، انجام بهسازی گفتار آلوده به نویز جمع شونده به کمک شبکه ویونت و ارزیابی عملکرد شبکه با معیارهای سیگنال به نویز و سیگنال به نویز قطعه ای و LLR می باشد.

بهسازی گفتار و مروری بر روش های آن

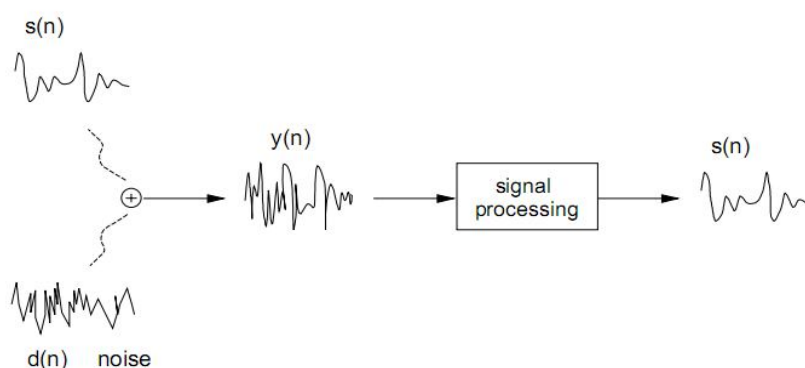
۲ بهسازی گفتار و مروری بر روش های آن

۱-۲ مقدمه

همان طور که قبلا بیان شد، بهسازی گفتار تلاش برای بهبود کیفیت عملکرد سیستم های ارتباط گفتاری در مواردی که سیگنال گفتار تحت تاثیر نویز، انعکاس و سایر عوامل تخریبی واقع گردیده است، می باشد. در این فصل به بررسی مفاهیم مرتبط با بهسازی گفتار پرداخته می شود و انواع طبقه بندی های موجود برای بهسازی گفتار بیان خواهد شد. همچنین مطالبی نیز راجع به روش های ارزیابی سیستم های بهسازی گفتار آورده شده است. منابع و مشخصات نویزهای صوتی نیز به عنوان یکی از عوامل مهم در شناخت ضرورت بهسازی گفتار، بررسی شده است.

۲-۲ بهسازی گفتار (Speech enhancement)

بهسازی گفتار به روش هایی منتهی می شود که با انجام پردازش هایی روی سیگنال های نویزی، نهایتا منجر به بهبود کیفیت و قابلیت فهم گفتار می شود [۹].



شکل ۱-۲: بلوک دیاگرام حذف نویز

کیفیت و قابلیت فهم دو معیار مهم ارزیابی یک سیگنال صوتی برای انسان می باشند. کیفیت یک معیار ذهنی ارزیابی سیگنال است که وضوح مطالب بیان شده و سطح نویز زمینه موجود در آن را اندازه می گیرد. می توان

این معیار را با میزان خوشایند بودن سیگنال برای شنونده یا میزان تلاش شنونده برای درک مطلب بیان شده در سیگنال بیان کرد. در مقابل قابلیت فهم یک معیار عینی است که درصد کلماتی (که می توانند بی معنی باشند) که به درستی توسط شنونده تشخیص داده شده را بیان می کند. این دو معیار از هم مستقل هستند، یعنی یک سیگنال صوتی ممکن است کیفیت بالایی داشته باشد، در حالیکه قابلیت فهم آن پایین باشد یا بالعکس. به عنوان مثال صدای تولید شده توسط ماشین دارای کیفیت بسیار پایین است، اما قابلیت فهم بالایی دارد [۱۰].

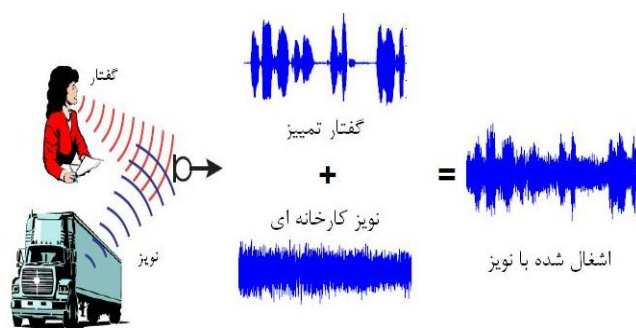
بهبود کیفیت و افزایش قابلیت فهم سیگنال صوتی به صورت توأم، از اهداف بهسازی گفتار است. اما تقریباً تمامی الگوریتم های بهسازی تک کاناله فقط کیفیت را بهبود می بخشند، در حالیکه قابلیت فهم را نه تنها افزایش نمی دهند، بلکه به علت اعوجاجی که در سیگنال ایجاد می کنند، آن را کاهش نیز می دهند. این مشکل به علت تاثیر گذاری روش بهسازی بر بخش های نویزگونه انرژی پایین سیگنال گفتار مانند صامت ها به وجود می آید. زیرا این روش ها معمولاً صامت ها را نویز تشخیص داده و حذف می کنند. در حالیکه صامت ها نقش موثرتری نسبت به مصوت ها در فهم گفتار دارند [۱۱]. از این رو بیشتر شنونده ها اطلاعات بیشتری را از سیگنال نویزی در مقایسه با سیگنال بهسازی شده استخراج می کنند. یکی از نکات کلیدی در طراحی یک سیستم بهسازی گفتار این است که روش بهسازی قادر به حذف نویز زمینه (افزایش کیفیت سیگنال) باشد، بدون اینکه سیگنال اصلی را دچار اعوجاج کند. این بدین معنی است که این سیستم ها در سیگنال بهسازی شده دو نوع اثر از خود به جای می گذارند، یکی حذف نویز زمینه و دیگری اعوجاج سیگنال.

در صورت دسترسی به معیاری دقیق برای فرموله کردن کیفیت و قابلیت فهم، می توان کیفیت را بدون ایجاد اعوجاج در سیگنال، افزایش داد.

۲-۳ منابع و مشخصات نویزهای صوتی

نویز یا صداهای مزاحم تقریباً در همه موارد همراه سیگنال صوت می‌باشند. زمانی که شما از خیابان عبور می‌کنید (صدای خودروها و ساخت و ساز و ...)، وقتی داخل خودرو نشسته اید و در حرکت هستید (صدای موتور ماشین، صدای باد و ...) زمانی که به یک فروشگاه یا رستوران می‌روید (صدای همه‌همه دیگران)، وقتی که به دفتر محل کار خود می‌رسید (صدای کولر، زنگ تلفن و ...) و به طور خلاصه در همه حالات، شما با انواعی از نویزها، که هر کدام دارای ویژگی‌های خاصی هستند، سروکار دارید. به طور کلی نویزها را می‌توان به دو دسته ایستان و غیر ایستان تقسیم نمود. نویز ایستان، نویزی است که مشخصات آماری آن با زمان تغییر نمی‌کند. مانند صدای کولر. در مقابل نویز غیر ایستان، نویزی است که مشخصه‌های زمانی و طیفی آن متغیر باشد. مانند نویز فرودگاه که ترکیبی از نویز همه‌همه و سایر صداهاست.

اثر نویز بر گفتار می‌تواند جمع شونده (مثل نویز)، کانوالوشونده (مثل نویز کانال تلفن یا نویز پژواک) و یا ضرب شونده (مانند نویز محوسازی) باشد. جنس نویز می‌تواند شبیه به گفتار باشد و با گفتار همبستگی داشته باشد، یا فقط یک نویز زمینه (مانند صدای کولر و پنکه) بوده و غیر همبسته با گفتار باشد. شکل زیر نمونه‌ای از تولید سیگنال نویزی را نشان می‌دهد.



شکل ۲-۲: نویز جمع شونده پس زمینه

با توجه به تنوع موجود در شکل و مشخصات نویزها، واضح است که قبل از هرکاری نیاز به شناسایی دقیق نویزها از نظر مشخصات زمانی و طیفی و درک تفاوت بین آنها هستیم. از ویژگی‌های مهمی که انواع نویز را از

هم متمایز می‌کند شکل طیف و توزیع انرژی آنها می‌باشد [۳۰].

۲-۴ طبقه بندی روش های بهسازی گفتار

در طول دهه های اخیر و با پیشرفت سخت افزاری و نرم افزاری روش های مختلف پردازش سیگنال گفتار، طیف وسیعی از رویکردها و روش ها برای بهسازی گفتار ارائه گردیده است. با توجه به تعدد روش های بهسازی گفتار، آنها را به روش های متفاوتی طبقه بندی می‌کنند.

یکی از روش های تقسیم بندی تکنیک های بهسازی گفتار بر پایه تعداد کانال (میکروفون) ورودی می‌باشد. بر این اساس دو خانواده تک کاناله و چند کاناله برای گروه بندی روش های مختلف موجود می‌باشد.

در روش تک کاناله تنها یک میکروفون ورودی در دسترس بوده و کلیه اطلاعات لازم باید از همین سیگنال استخراج شود. اساس کار این خانواده از روش ها، مبتنی بر ایستادن بودن نویز می‌باشد. فرضی که در مواجهه با نویزهای غیرایستادن زیر سؤال رفته و عملکرد روش را دچار مشکل می‌نماید.

در روش های چندکاناله، دو یا چند گیرنده (میکروفن) در ورودی سیستم بهسازی گفتار مورد استفاده واقع می‌شود. افزایش تعداد میکروفون ها و یا کانال های ورودی، قدرت روش را در پاکسازی سیگنال نویزی بالا می‌برد و کار بهسازی را ساده تر می‌کند، اما در مقابل پر هزینه بوده و پیچیدگی پیاده سازی سیستم را نیز می‌افزاید.

از روش های تک کاناله می‌توان به تفریق طیفی، فیلتر وینر، زیرفضای سیگنال، فیلتر کالمن، روشهای مبتنی بر تبدیل ویولت، روشهای مبتنی بر کپسترال و روش های آماری اشاره کرد.

از روش های چند کاناله نیز می‌توان فیلتر فوقی (بیشتر به صورت دوکاناله)، جداسازی سیگنال (قابل استفاده هم به صورت تک کاناله و هم به صورت چند کاناله) و روش های تشکیل پرتو را نام برد.

روش های بهسازی گفتار را می‌توان براساس حوزه فعالیت نیز دسته بندی کرد که بر این اساس آنها را به دو دسته کلی حوزه زمان (مانند فیلتر کالمن و فیلتر فوقی) و حوزه فرکانس یا طیف (مانند تفریق طیفی و فیلتر

وینر) تقسیم می‌کنند.

بیشتر روش‌های بهسازی گفتار در حوزه‌ی فرکانس و روی طیف عمل می‌کنند. زیرا در حوزه طیف قابلیت جداسازی نویز و سیگنال بالاست و می‌توان روش‌های بهینه یا ابتکاری را به خوبی پیاده‌سازی کرد. همچنین، مستقل بودن مولفه‌های طیف امکان کار با هر کدام از آنها را به صورت مستقل فراهم می‌کند که این به نوبه خود منجر به ساده‌تر شدن روابط ریاضی و کاهش محاسبات می‌گردد [۱۶].

یکی دیگر از راه‌های دسته‌بندی کردن روشهای بهسازی گفتار تفکیک این روش‌ها براساس پارامتری یا غیر پارامتری بودن آنهاست. روش‌هایی مانند تفریق طیفی و زیر فضای سیگنال به دلیل کار بر روی خود سیگنال، روش‌های غیر پارامتری محسوب می‌شوند و روش‌های مبتنی بر مدل مخفی مارکوف (HMM) و مدل مخلوط گوسی (GMM) به دلیل استفاده از مدل‌های آماری پارامتری برای سیگنال و نویز در ردیف روش‌های پارامتری قرار می‌گیرند.

روش‌های بهسازی گفتار را می‌توان به راهکارهای آماری و غیر آماری، وفقی و غیروفقی نیز دسته‌بندی شوند. با وجود طبقه‌بندی‌های فوق، به دلیل تشابه زیاد روش‌های بهسازی گفتاری در ایده‌ها و پایه‌های بنیادی، تقسیم‌بندی آنها در قالب چند روش کاملاً مجزا، کار ساده‌ای نیست. به گونه‌ای که می‌توان گفت برخی از روش‌ها، حالت کلی یا تعمیم یافته برخی روش‌های دیگر بوده و یا این که برخی از روش‌ها صرف نظر از استفاده از تبدیل‌هایی متفاوت در نمایش سیگنال (مانند فوریه و موجک) دارای اساس مشترکی هستند.

۲-۵ عوامل موثر در طراحی سیستم‌های بهسازی گفتار

علی‌رغم وجود روش‌های مختلف در بهسازی گفتار، باید توجه داشت که طراحی یک سیستم بهسازی گفتار به عوامل مختلفی وابسته است. از جمله این عوامل می‌توان به موارد زیر اشاره کرد:

کاربرد مورد نظر: دربرخی کاربردها افزایش کیفیت مساله اصلی است مانند و سایل کمک شنوایی، ولی در برخی از آنها قابلیت فهم و درک (مانند کاربرد‌های نظامی و باز شناسی گفتار) مهم‌تر است.

نوع نویز و مشخصات آن: اینکه نویز ایستان یا غیر ایستان باشد، از جنس گفتار (مانند نویز همه‌مه)، جنس نویز زمینه (صدای ماشین، پنکه و باد) و یا از نوع پژواک اتاق و نویز کانال باشد، یا اینکه نویز دارای پهنای باند بالا یا کم باشد، در طراحی سیستم بهسازی گفتار موثر است.

رابطه نویز و گفتار تمیز: اثر نویز برگفتار می‌تواند جمع شونده (مثل نویز سفید)، کانالوشونده (مثل نویز کانال تلفن یا نویز پژواک) و یا ضرب شونده (نویز محوسازی) باشد.

نویز و گفتار می‌توانند همبسته یا غیر همبسته باشند.

تعداد میکروفون‌ها: هر چه تعداد میکروفون‌ها بیشتر باشد، بهسازی گفتار راحت‌تر است [۱۲].

۲-۶ ارزیابی سیستم‌های بهسازی گفتار

مشابه سایر سیستم‌ها، برای ارزیابی کارایی روش‌های بهسازی گفتار و مقایسه تکنیک‌های مختلف با هم، نیازمند داشتن ابزارهایی هستیم که بتواند عملکرد یک سیستم بهسازی گفتار را به صورت کمی نشان دهند.

ارزیابی یک سیستم بهسازی گفتار می‌تواند وابسته به کاربردی که از این سیستم مورد نظر است، باشد. مثلاً اگر هدف از بهسازی گفتار، بهبود کارایی یک سیستم بازشناسی گفتار است، دقت یا صحت سیستم بازشناسی می‌تواند بهترین معیار ارزیابی برای روش‌های بهسازی گفتار باشد.

در یک نگاه کلی می‌توان تکنیک‌های ارزیابی کیفیت گفتار را به دو دسته ذهنی و عینی تقسیم بندی کرد.

معیارهای ذهنی، براساس نظر شنونده‌های انسانی بوده که به صورت شهودی رتبه‌ای را به سیگنال گفتار در یک بازه‌ی امتیازی از قبل تعیین شده اختصاص می‌دهند. از معایب این معیارها این است که، نتایج یک آزمایش ذهنی تحت تاثیر عوامل زیادی از جمله محتوای آزمایش و علایق انجام دهنده آزمایش است. به علاوه این روش‌ها نیاز به دسترسی به تعداد افراد زیادی دارد که فرآیند ارزیابی را وقت گیر و پرهزینه می‌سازد. در روش‌های عینی از تکنیک‌های محاسباتی جهت مقایسه سیگنال اصلی و سیگنال پردازش شده استفاده می‌کنند. اکثر روش‌های عینی اندازه‌گیری کمی کیفیت سیگنال را از روی مدل‌های سیستم شنوایی

انسان انجام می‌دهند. امروزه به این روش‌ها در ارزیابی (علی‌رغم ارجحیت ارزیابی ذهنی) توجه بیشتری می‌شود [۱۳]. در زیر به چند نمونه معیار ارزیابی اشاره شده است.

۱-۶-۲ معیار سیگنال به نویز (SNR)

محاسبه سیگنال به نویز از جمله قدیمی‌ترین و معمول‌ترین روش‌های ارزیابی عینی کیفیت سیگنال است که علاوه بر SNR کلی، انواع مختلفی از آن مانند SNR قطعه‌ای و SNR قطعه‌ای وزن دار معرفی شده است.

با فرض جمع شدن نویز $d(n)$ با سیگنال گفتار تمیز $s(n)$ ، می‌توان سیگنال نویزی y را به صورت

$$y(n) = s(n) + d(n) \text{ در نظر گرفت. اگر } \hat{S}(n) \text{ را سیگنال بهسازی شده در نظر بگیریم، آنگاه رابطه}$$

$$e(n) = s(n) - \hat{S}(n) \text{، معادل خطای بین سیگنال اصلی و سیگنال تخمین زده شده برای نمونه } n \text{ ام است. با}$$

استفاده از این فرضیات، SNR کلی برابر با نسبت انرژی سیگنال اصلی به انرژی خطا تعریف می‌شود که رابطه

آن بر حسب دسی بل به صورت رابطه (۱-۲) است که در آن N تعداد کل نمونه‌های سیگنال است [۱۸].

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N S^2(n)}{\sum_{n=1}^N [s(n) - \hat{S}(n)]^2} \quad (1-2)$$

بدیهی است که در این رابطه، به سیگنال تمیز نیاز داریم. بنابراین این رابطه بیشتر در شبیه‌سازی‌ها و برای زمانی که هر دو سیگنال تمیز و نویزی در دسترس هستند، کاربرد دارد. مهم‌ترین مزیت این روش سادگی آن از نظر ریاضی است اما به علت عدم همبستگی با معیارهای ذهنی و یکسان فرض کردن کل سیگنال، در بسیاری از موارد تخمین مناسبی از کیفیت ارائه نمی‌دهد.

می‌دانیم که در عمل، انرژی سیگنال گفتار با زمان تغییر می‌کند و بعضی از مولفه‌های سیگنال نقش بیشتری در تامین کیفیت ایفا می‌کنند. یکی از حالت‌هایی که می‌تواند به طور نادرست منجر به SNR بالا شود، وقتی است که سیگنال گفتار دارای قسمت‌های گفتاری متمرکزی باشد، در حالی که نویز بر روی قسمت‌های گفتار با انرژی کم اثر ادراکی شدیدی داشته باشد. از این رو اگر فرض کنیم که اعوجاج ناشی از نویز نوسانات انرژی کمی دارد، اندازه‌گیری می‌تواند قاب به قاب تغییر کند. این مسائل مبنای ارائه روش

SNR قطعه ای است.

۲-۶-۲ معیار سیگنال به نویز قطعه‌ای (SNR_{seg})

در این معیار SNR در قاب های کوچک تر محاسبه شده و میانگین این مقادیر به عنوان ارزیابی نهایی ارائه می گردد.

رابطه این معیار به صورت رابطه (۲-۲) است که در آن M تعداد قاب ها و L طول هر قاب است.

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{f=1}^M \log_{10} \left[\frac{\sum_{n=L.F}^{L.F+L-1} s^2(n)}{\sum_{n=L.F}^{L.F+L-1} [s(n) - \hat{s}(n)]^2} \right] \quad (2-2)$$

در برخی از موارد که قاب های سکوت نیز در سیگنال وجود دارند ممکن است این مقدار منفی شود. برای رفع این مشکل می توان بازه های سکوت را تشخیص داد و آنها را در محاسبه لحاظ نکرد. به علاوه، با در نظر گرفتن یک آستانه پایین و جایگزین کردن مقادیر کوچکتر از آن با مقدار آستانه، می توان از منفی شدن جلوگیری کرد.

همچنین گوش انسان بین قاب هایی با مقادیر SNR بالاتر از ۳۵dB تمایز چندانی نمی گذارد، بنابراین یک آستانه بالا برای متعادل کردن SNR هایی که بسیار بالا هستند، به کار می رود. این دو آستانه که معمولاً ۰dB و ۳۵dB است، باعث می شوند تا جواب نهایی در بازه قابل قبولی باقی بمانند [۳۰].

۳-۶-۲ معیار سیگنال به نویز قطعه‌ای وزندار (SNR_{fw-seg})

توسعه رابطه SNR به حوزه فرکانس با نام SNR قطعه ای وزندار، به صورت زیر است که در آن K تعداد کل باندهای فرکانسی و W_k هر کدام از آنهاست. این معیار در رابطه (۳-۲) آمده است.

M تعداد قاب ها، F(f, k) و F̂(f, k) دامنه فیلتر بانک برای سیگنال تمیز و تخمینی در فرکانس kام از قاب f است.

$$SR_{fw-seg} = \frac{10}{M} \sum_{f=1}^M \frac{\sum_{k=1}^K W_k \log_{10} \left[\frac{F^2(f, k)}{[F(f, k) - \hat{F}(f, k)]^2} \right]}{\sum_{k=1}^K W_k} \quad (3-2)$$

۴-۶-۲ معیار (PESQ (perceptual Evaluation Speech Quality

برخی از معیارهای ارزیابی مورد استفاده در بهسازی گفتار از حوزه های دیگر پردازش گفتار، از جمله سنتز و کدینگ گفتار گرفته شده اند که روش PESQ یکی از آنهاست. در سال ۲۰۰۰ مسابقه ای توسط ITU برای معرفی یک معیار عینی جدید ترتیب داده شد.

هدف اصلی از تعریف این معیار عینی جدید قابلیت اطمینان بالا در شبکه های مختلف و شرایط متفاوت بود و در این میان PESQ معرفی شد. به طور خلاصه، PESQ یک معیار عینی برای تخمین یا پیش بینی امتیاز یک آزمایش واقعی (معیار ذهنی MOS) است. در این آزمایش واقعی، کیفیت سیگنال از طریق گوش دادن افراد مختلف به سیگنال صوتی و امتیاز دادن به آن توسط افراد، سنجیده می شود. ضریب همبستگی بین این روش و روش های ذهنی ۹۳.۵٪ است. علی رغم این همبستگی بالا بر این نکته که این معیار نباید به جای روش های ذهنی به کار گرفته شود، تاکید شده است. (امتیاز این روش به وسیله پایگاه داده بزرگی از معیار های ذهنی کالیبره شده است).

در این روش سیگنال بهسازی شده با سیگنال اصلی مقایسه می شود و نمره ای بین ۰.۵ تا ۴.۵ داده می شود. هر چقدر که این معیار بیشتر باشد به معنای کیفیت بالاتر سیگنال صوتی است.

در اولین گام برای محاسبه امتیاز PESQ، ابتدا تاخیر بین دو سیگنال با مکانیزم هایی که در [ITU2000] تشریح شده است، حذف می شود. سپس دو سیگنال با استفاده از یک مدل ادراکی با هم مقایسه می شوند.

در این مقایسه سعی شده نمایشی که سیستم شنوایی انسان از سیگنال ایجاد می کند، مورد استفاده قرار گیرد که برای رسیدن به این هدف چند گام مانند تطبیق زمانی، تطبیق سطوح سیگنال ها با سطوح شنوایی،

تبدیل زمان-فرکانس و سنجش شدت سیگنال در نظر گرفته شده است.

به صورت خلاصه می توان گفت که الگوریتم PESQ به این ترتیب عمل می کند. ابتدا هر دو سیگنال (اصلی و اعوجاج یافته) به یک سطح استاندارد شنوایی نگاشته شده و با فیلتری میان گذر با فرکانس های قطع ۳۰۰ و ۳۰۰۰ هرتز، فیلتر می شوند. در مرحله بعد دو سیگنال از لحاظ زمانی همزمان شده و به این ترتیب تاخیرهای زمانی (که در کار کدینگ به خاطر عبور از شبکه معمول تر است) حذف می شود.

پس از این مرحله، سیگنال ها توسط یک تبدیل شنوایی پردازش شده و میزان بلندی آنها به دست می آید. تفاوت بلندی دو سیگنال محاسبه شده و در طول زمان و فرکانس میانگین گیری می شوند تا تخمینی از معیار MOS به دست آید. از لحاظ پیچیدگی محاسباتی PESQ پیچیده ترین معیار عینی است [۱۳].

۵-۶-۲ معیار LLR

پس از بررسی معیارهای بالا به بررسی LLR(Log Likelihood Ratio) می پردازیم که به صورت رابطه ی زیر تعریف می شود:

$$LLR = \log\left(\frac{a_e^T R_x a_e}{a_x^T R_x a_x}\right) \quad (۴-۲)$$

که در آن a_e و a_x به ترتیب ضرایب LPC سیگنال تمیز و سیگنال بهسازی شده و R_x ماتریس خود همبستگی سیگنال تمیز است. معیار LLR برای همه پنجره های یک جمله حساب می شود. در واقع این مقدار برای هر قاب از سیگنال محاسبه شده و در نهایت از مقادیر به دست آمده برای قابها میانگین گرفته می-شود. مقدار این معیار به بازه [۰ و ۲] محدود می شود [۱۹]. کمتر بودن مقدار LLR به معنی بهتر بودن کیفیت سیگنال خروجی است.

۷-۲ تشخیص گفتار از نویز

در سیستم پردازش گفتار، معمولاً سیگنال دریافتی، از دو قسمت گفتار و غیر گفتار که معمولاً سکوت و یا یک

نویز زمینه است، تشکیل می‌گردد.

تشخیص گفتار از سکوت در سیستم های بهسازی گفتار نیز نقش بسیار مهمی دارد، زیرا در این نوع سیستم ها باید نمونه ای از نویز خالص را که در نواحی سکوت یافت می شود، بدست آوریم.

روش تشخیص گفتار از سکوت غالباً مبتنی بر انرژی سیگنال عمل می کنند. همچنین ترکیب این روش با معیارهای دیگر مانند نرخ عبور از صفر و یا آنتروپی و روش مبتنی بر ضرایب کپسترال نیز در این زمینه استفاده می شود [۴۴].

۸-۲ روش های تخمین نویز در بهسازی گفتار

یکی از نیازهای کلیه روش های بهسازی گفتار تک کاناله، تخمین نویز یا مشخصات آن است که کارایی روش بهسازی گفتار را به شدت تحت تاثیر قرار می دهد. در روش های بیان شده تا به حال فرض شد که تخمینی از نویز در دسترس است. در این قسمت روش های تخمین نویز بررسی خواهند شد.

ساده ترین روش تخمین نویز VAD است. در این روش با استفاده از انرژی، نرخ عبور از صفر، متناوب بودن و ... برای هر قاب، در مورد گفتار یا غیر گفتار بودن آن تصمیم گیری می شود. تصمیم گیری اغلب براساس مقایسه این مقادیر با مقادیر آستانه است. روش های مختلفی برای VAD تا به حال پیشنهاد شده است. برخی از روش ها مبتنی بر انرژی و نرخ عبور از صفر هستند و برخی دیگر بر اساس ضرایب کپسترال، معیار فاصله LPC ایتاکورا و متناوب بودن گفتار عمل می کنند. در این روش مشخصات نویز در بخش های سکوت و غیر گفتاری سیگنال به روز می شود. این روش در SNR های بالا و برای نویزهای ایستادن مناسب است. به علاوه برخی از آنها نیاز به برخی پارامترها مانند مقادیر آستانه و برخی تنظیمات دیگر دارند که ممکن است از محیطی به محیط دیگر تغییر کند. سیستم های بهسازی گفتار برای داشتن عملکرد مطلوب در کاربردهای واقعی باید از توانایی تخمین مشخصات نویز در شرایط غیر ایستادن و دنبال کردن این مشخصات برای همه زمان ها حتی، در شرایطی که گفتار وجود داشته باشد، برخوردار باشند. از این رو روش های تخمین نویز به

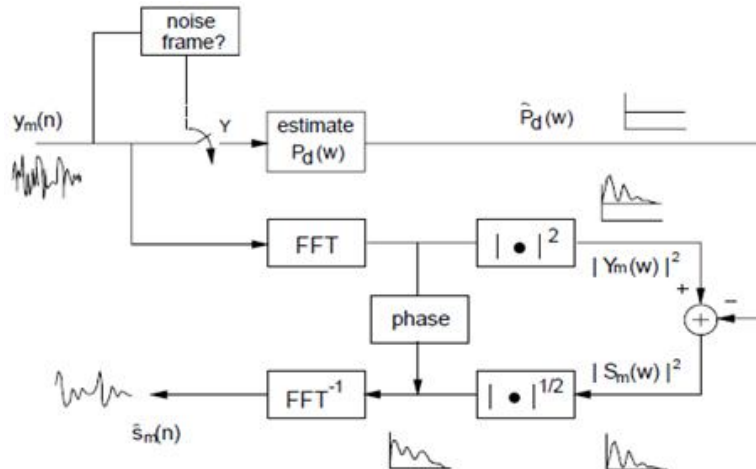
عنوان یکی از بخش های مهم در سیستم های بهسازی گفتار مطرح شده اند.

۹-۲ مروری بر روش های بهسازی گفتار

روش های زیادی برای مساله بهسازی گفتار ارائه شده اند که می توان آنها را در دو گروه اصلی روش های آماری و غیرآماری تقسیم کرد. علی رغم وجود ماهیت تصادفی و آماری سیگنال، منظور از روش های غیرآماری، روش هایی مانند تفریق طیفی و زیر فضای سیگنال است که در آنها هیچ فرض خاصی در مورد شکل توزیع سیگنال یا طیف آن و اطلاعات پیشین مرتبط انجام نمی شود. در مقابل در روش های آماری فرض می شود که سیگنال یا طیف آن از توزیع خاصی پیروی می کنند. برای حذف نویز از سیگنال گفتار، می توان طیف فرکانسی سیگنال نویزی را بهگونه ای اصلاح نمود که پس از بازیابی سیگنال در حوزه زمان، صداهای پس زمینه گفتار حذف گردد و یا کاسته شود. به این منظور ابتدا باید سیگنال گفتار با تبدیلی نظیر تبدیل فوریه به حوزه فرکانس انتقال یابد و سپس اصلاحات لازم بر روی طیف فرکانسی به شیوه ای صورت گیرد تا نویز کاهش یابد. در این قسمت به مرور چندی از روش های بهسازی گفتار بیان شده خواهیم پرداخت.

۱-۹-۲ تفریق طیفی

تفریق طیفی از قدیمی ترین و رایج ترین الگوریتم های بهسازی گفتار است که به خاطر سادگی و پایین بودن محاسبات آن هنوز هم مورد توجه محققین قرار دارد. ایده این روش بسیار واضح و ساده است [۴۳].



شکل ۲-۳: بلوک دیاگرام تفریق طیفی

اگر فرض شود که سیگنال گفتار تمیز $s(n)$ در اثر جمع شدن با نویز مستقل $d(n)$ به سیگنال نویزی $y(n)$ تبدیل شود، آنگاه با اعمال تبدیل فوریه زمان کوتاه، رابطه اول، به صورت رابطه دوم خواهد شد.

$$y(n) = s(n) + d(n) \quad (۵-۲)$$

$$Y(\omega_k) = S(\omega_k) + D(\omega_k) \Rightarrow Y_k e^{j\theta_{y(k)}} = S_k e^{j\theta_{s(k)}} + D_k e^{j\theta_{d(k)}}$$

که در آن $\omega_k = 2k\pi/N$ $k=0,1,2,\dots,N$ و N طول هر قاب سیگنال گفتار است.

در این رابطه Y_k ، S_k و D_k به ترتیب دامنه طیف سیگنال نویزی، تمیز و نویز هستند و $\theta_{y(k)}$ ، $\theta_{s(k)}$ و $\theta_{d(k)}$ به ترتیب فاز سیگنال های نویزی، تمیز و نویز هستند. در این روش برای تخمین دامنه طیف سیگنال تمیز از روی سیگنال نویزی می توان به سادگی دامنه طیف نویز را از دامنه طیف سیگنال نویزی پیدا کرد.

یعنی:

$$\hat{S}_K = Y_K - D_K \quad (۶-۲)$$

نهایتاً در تخمین نهایی می توان از فاز سیگنال نویزی به عنوان فاز سیگنال تمیز استفاده کرد.

$$\hat{S}(\omega_k) = \hat{S}_K e^{j\theta_{s(k)}} = [Y_K - D_K] e^{j\theta_{y(k)}} \quad (۷-۲)$$

همانگونه که از رابطه ی بالا مشخص است برای پیاده سازی به تخمینی از دامنه ی طیف نویز نیاز داریم. روشهای مختلفی برای این کار وجود دارند که در تفریق طیفی معمولاً از VAD یا آشکار کننده ی گفتار از

غیرگفتار استفاده می شود. VAD از دو قسمت استخراج ویژگی و تصمیم گیری تشکیل شده است. در بخش استخراج ویژگی از سیگنال گفتار ویژگی هایی مانند انرژی، نرخ عبور از صفر و ضرایب کپسترال استخراج می-شود و سپس واحد تصمیم گیری با استفاده از یک روش دسته بندی یکی از دو حالت گفتار یا سکوت را به سیگنال نسبت میدهد. در این روش، فرض اساسی ایستادن بودن موضعی نویز در فاصله میان دو سکوت ضروری است. زیرا فرض می شود مشخصات نویز در فاصله دو سکوت ثابت بوده و تخمین طیف نویز در این فاصله به روز نمی شود.

با وجود سادگی روش تفریق طیفی عمل تفریق باید با دقت انجام شود تا باعث ایجاد اعوجاج در سیگنال تخمینی نشود. در شرایطی به علت تخمین بیش از حد نویز ممکن است مقداری از اطلاعات سیگنال گفتار حذف شود. از طرف دیگر با تخمین کم نویز سیگنال بهسازی شده دارای نویز خواهد بود. یکی از مشکلاتی که معمولاً در عمل در تفریق به وجود می آید منفی شدن برخی از مولفه های طیف تخمینی است که این معمولاً به علت تخمین نامناسب طیف نویز ایجاد می شود. در این حالت نویزی بنام نویز موزیکال ایجاد می شود [۴] و [۳].

در بسیاری از سیستم های ساده به جای این مقادیر منفی از صفر یا مقادیر کوچک مثبتی استفاده می شود. در نتیجه در سیگنال بهسازی شده برخی مولفه های طیف به تناوب خاموش و روشن شده و به صورت نویز موزیکال شنیده می شود. معمول ترین راه حل رفع این مشکل استفاده از یکسوسازهای نیم موج است.

یکی دیگر از کاستی های این روش که در مقابل نویز موسیقی از اهمیت کمتری برخوردار است، استفاده از فاز سیگنال نویزی به عنوان فاز سیگنال بهسازی شده است. این جایگزینی در برخی از موارد می تواند باعث تغییرات زیاد و مداوم در کیفیت سیگنال شود. علت این جایگزینی کم اثر بودن فاز سیگنال نویزی بر روی کیفیت سیگنال گفتار و عدم حساسیت زیاد سیستم شنوایی انسان به آن است [۲۰]. اما فقط در صورت بالا بودن SNR سیگنال گفتار (بالتر از ۸dB) می توان فاز سیگنال نویزی را بدون از دست دادن کیفیت به جای فاز سیگنال تمیز استفاده کرد [۲۱].

به طور خلاصه فرضیات این روش، نویز ایستان یا ایستان زمان کوتاه، مستقل و جمع شونده با گفتار می‌باشد. از مزایای آن محاسبات پایین و سادگی می‌باشد. از معایب آن ایجاد نویز زمینه (نویز موسیقی) در سیگنال بهسازی شده می‌باشد.

۲-۹-۲ زیرفضای سیگنال

اغلب روش های بهسازی براساس تئوری پردازش سیگنال و تخمین آماری هستند. اما روش های زیرفضای سیگنال مبتنی بر تئوری جبرخطی هستند. در واقع این روش بر این اصل استوار است که می‌توان سیگنال نویزی را به زیر فضاهایی تجزیه کرد که در آنها بخش های سیگنال تمیز و نویز متمایز می‌شوند و با صفر کردن بخش های نویزی در این زیر فضا می‌توان تخمینی از سیگنال تمیز بدست آورد.

تجزیه سیگنال نویزی به دو زیر فضای سیگنال تمیز و نویز با استفاده از روش های شناخته شده تجزیه ماتریس های متعامد در جبرخطی مثل تجزیه مقادیر تکین (SVD) یا تجزیه براساس بردارهای ویژه و مقادیر ویژه امکان پذیر است. روش تجزیه بر اساس مقادیر ویژه و بردارهای ویژه به اسم تبدیل کارهونن - لاف، تبدیل هتلینگ و تحلیل اجزای اصلی نیز معروف است.

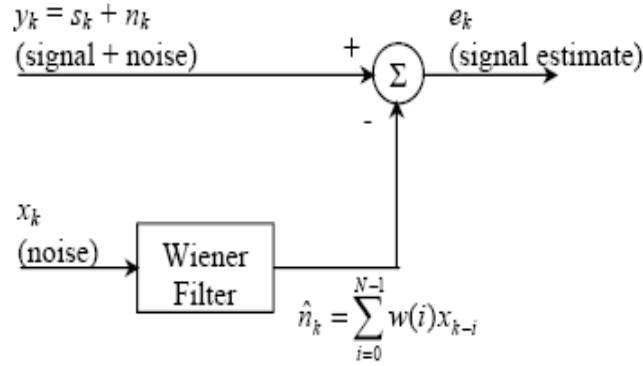
به طور خلاصه در فرضیات این روش، نویز و گفتار مستقل بوده و اطلاعات آماری درجه دوم آنها در دسترس می‌باشد. از مزایای آن محاسبات پایین و سادگی می‌باشد. از معایب آن عدم کارایی بالا است.

۳-۹-۲ فیلتر وینر

فیلتر وینر یکی از روش های شناخته شده در زمینه فیلتر کردن آماری سیگنال است و در زمینه های مختلفی مانند پردازش سیگنال دیجیتال، پیش بینی خطی و ... کاربرد دارد.

این روش که ابتدا توسط نوربرت وینر (۱۹۴۹) فرموله شده است، به حل مساله زیر می‌پردازد.

نقطه شروع در روابط فوقی، تعریف مشخص فیلتر بهینه می‌باشد. فیلتر Wiener رایج ترین نوع مورد استفاده می‌باشد.



شکل ۲-۴: حذف نویز با فیلتر وینر

اگر نمونه سیگنال y ، y_k ، شامل سیگنال اصلی s_k و نویز n_k می باشد، که نویز n_k خود با x_k همبسته است. فیلتر wiener به کمک یک فیلتر دیجیتال (در این مثال فیلتر FIR مرتبه N) نویز را تخمین می زند. $w(i)$ ، i امین ضریب فیلتر می باشد، از آنجاییکه با سیگنال گسسته کار می کنیم می توانیم روابط را به صورت ماتریسی بیان کنیم:

$$X_k = \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ \vdots \\ x_{k-(N-1)} \end{bmatrix} \quad W = \begin{bmatrix} w_{(0)} \\ w_{(1)} \\ \vdots \\ \vdots \\ w_{(N-1)} \end{bmatrix} \quad (۸-۲)$$

$$e_k = y_k - w^T X_k = y_k - x_k^T w \quad (۹-۲)$$

مربع خطای لحظه ای سیگنال به صورت زیر است:

$$e_k^2 = y_k^2 - 2w^T(y_k X_k) + w^T X_k X_k^T w \quad (۱۰-۲)$$

میانگین مربع خطا همان امید ریاضی عبارت بالا می باشد:

$$\varepsilon = \text{MSE} = E\{e_k^2\} - 2w^T E\{y_k X_k\} + w^T E\{X_k X_k^T\} w \quad (۱۱-۲)$$

اگر $E\{X_k X_k^T\}$ را با R_{XX} و $E\{y_k X_k\}$ را با R_{yx} جایگزین کنیم، رابطه بالا به صورت زیر ساده می شود:

$$\text{MSE} = E\{y_k^2\} - 2w^T R_{yx} + w^T R_{XX} w \quad (۱۲-۲)$$

هدف می نیمم کردن تابع هدف MSE فیلتر وینر می باشد.

برای می‌نیم کردن از MSE بر حسب w مشتق می‌گیریم که این رابطه در معادله (۱۳-۲) آمده است:

$$\begin{aligned}\nabla &= E \left\{ \frac{\partial e_k^2}{\partial w} \right\} = E \left\{ 2e_k \frac{\partial e_k}{\partial w} \right\} = 2E \left\{ (y_k - X_k^T w) \frac{\partial (y_k - X_k^T w)}{\partial w} \right\} \\ &= -2E \{ (y_k - X_k^T w) X_k \} = -2E \{ X_k y_k \} + 2E \{ X_k X_k^T \} w \\ &= -2R_{yx} + 2R_{xx}w\end{aligned}\quad (13-2)$$

وقتی ضرایب فیلتر مقدار حداقل را برمی‌گزینند که مشتق آن برابر صفر شود.

$$-2R_{yx} + 2R_{xx}w_{opt} = 0 \Rightarrow w_{opt} = R_{xx}^{-1} R_{yx} \quad (14-2)$$

R ماتریس همبستگی سیگنال ورودی و R_{yx} ماتریس همبستگی بین ورودی x و سیگنال y می‌باشد.

۲-۹-۴ روش‌های آماری

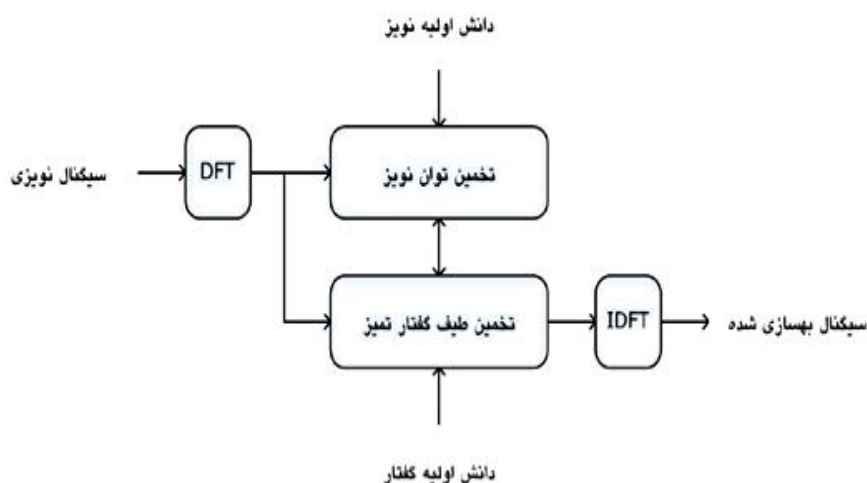
روش‌های آماری به تکنیک‌هایی گفته می‌شود که در آنها یک مدل آماری برای سیگنال گفتار تمیز و نویز در نظر گرفته می‌شود و پردازش براساس این مدل‌ها صورت می‌گیرد.

این روش به دو دسته روش‌های آماری محض و روش‌های مبتنی بر مدل تقسیم می‌شود. روش‌های آماری محض، در آن یک مدل آماری خاص (مثلاً توزیع گاوسی) برای ضرایب فوریه یا پارامترهای دیگر هر قاب سیگنال گفتار و نویز فرض می‌شود و سپس روابط بهسازی گفتار، متناسب با این مدل فرض شده به دست می‌آید. این روش‌ها منجر به روابط و فیلترهایی می‌شوند که در سطح قاب عمل کرده و اطلاعات سیگنال گفتار و نویز را در همین سطح مدل سازی می‌کند.

از این رو در برخی منابع به این روش‌ها، روش‌های مبتنی بر مدل کوتاه مدت یا روش‌های آماری بدون آموزش گفته می‌شود. دسته دیگر از روش‌های آماری روش‌های مبتنی بر مدل یا روش‌های long term یا روش آموزش یافته هستند.

در این روش‌ها به کمک داده‌های آموزشی مدل یا مدل‌هایی ایجاد می‌شود که خصوصیات مورد نظر را (که معمولاً تابع توزیع احتمال و پارامترهای آن است) در خود دارد. در این روش با استفاده از داده‌های

آموزشی، خصوصیات بلند مدت و مشترک بین قاب های مختلف فرا گرفته می شود. همانطور که از شکل ۵-۲ نیز مشخص است، در روش های آماری، علاوه بر توزیع های آماری که بخش جدانشدنی الگوریتم هستند، نیاز به یک معیار اعوجاج یا تخمین گر نیز وجود دارد که بر اساس کمینه کردن اعوجاج یا بهینه کردن رابطه مربوط به تخمین گر، روابط تخمین سیگنال تمیز به دست می آید.



شکل ۵-۲: نمای کلی روشهای آماری

از معروفترین معیارهای مورد استفاده در روش های آماری، می توان به تخمین گرهای کمترین میانگین مربعات خطا (MMSE)، بیشترین شباهت (ML) و بیشترین احتمال پسین (MAP) اشاره کرد. به طور خلاصه فرضیات در این روش استقلال نویز و گفتار، وجود دانش قبلی از سیگنال تمیز و نویز مثلاً نوع توزیع آنها می- باشد. از مزایای این روش کارایی بالا نسبت به سایر روش ها، ایجاد نویز موسیقی کم، اعوجاج کمتر در سیگنال می باشد. از معایب آن محاسبات و پیچیدگی بالای این روش است [۴۶].

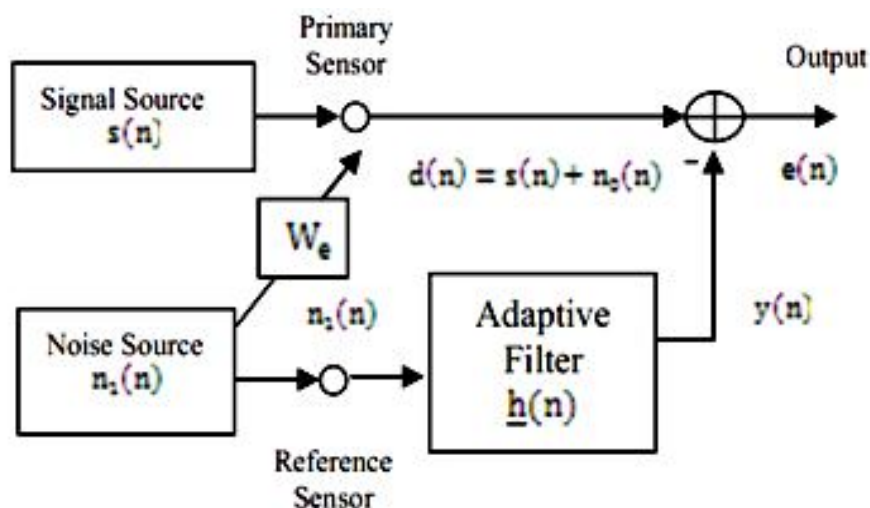
۵-۹-۲ فیلتر فوقی

در فیلتر فوقی حذف کننده نویز، هدف بدست آوردن یک تخمینی از سیگنال نویز و کم کردن آن از سیگنال خراب شده است. بلوک دیاگرام سیستم حذف نویز فیلتر فوقی را در شکل ۶-۲ می بینیم. برای این

هدف باید $e(n)$ و $s(n)$ را یکسان کنیم. به طور کلی سیستم حذف کننده نویز از دو ورودی:

۱- سیگنال مطلوب $(s(n))$ ۲- سیگنال نویز مرجع $(n_1(n))$ تشکیل شده است.

سیگنال اصلی $d(n)$ شامل سیگنال مطلوب و یک نسخه فیلتر شده یا تاخیر یافته از سیگنال نویز مرجع است.



شکل ۲-۶: حذف کننده‌ی نویز با فیلتر فوقی

$$d(n) = s(n) + n_0(n) \quad (۱۵-۲)$$

$$n_0 = n_1(n) * w_e$$

$n_1(n)$ سیگنال نویز مرجع و $d(n)$ سیگنال نویزی می‌باشد. w_e ضرایب فیلتر مسیر نویز می‌باشد. این فیلتر در

واقع بیانگر محیط انتشار نویز تا قبل از سوار شدن بر سیگنال تمیز است.

فرض بر اینست که سیگنال مطلوب $s(n)$ با سیگنال نویز مرجع $n_1(n)$ و نسخه فیلتر شده آن $n_0(n)$ ناهمبسته باشد.

$$E[s(n).n_0(n-m)] = 0 \quad (۱۶-۲)$$

$$E[s(n).n_1(n-m)] = 0$$

فیلتر فوقی با پردازش بر روی سیگنال نویز مرجع به دنبال تخمینی از w_e است.

$$e(n) = s(n) + n_1(n).w_e - n_1(n).h(n) \quad (۱۷-۲)$$

$$e(n) = s(n) + n_1(n).[w_e - h(n)]$$

کارایی سیستم های فوقی حذف نویز وابسته با فرضیات زیر است :

- سیگنال و نویز در خروجی سنسور اصلی باید ناهمبسته باشد .
- سیگنال نویز مرجع باید با سیگنال نویزی که سیگنال مطلوب را خراب کرده همبستگی بالایی داشته باشند.
- عناصر سیگنال مطلوب موجود در سنسور اصلی ، نباید در خروجی سنسور نویز مرجع آشکار شوند .
- طول فیلتر برای حذف بخش های مهم نویز کافی باشد .
- سیستم باید به گونه ای طراحی شود که نویز ناشی از کوانتیزه کردن به حداقل برسد .
- در بسیاری از کاربردهای لغو نویز مشخصات سیگنال تغییرات بسیار سریعی دارند. بر این اساس الگوریتم هایی که همگرایی سریعتری نسبت به بقیه دارند مورد نیاز است.

مروری بر شبکه عصبی و تبدیل موجک و کاربرد آنها در بهسازی گفتار

۳ مروری بر شبکه عصبی و تبدیل موجک

۱-۳ مقدمه

شبکه عصبی مصنوعی از شبکه عصبی بیولوژیکی الهام گرفته است بدین معنی که یک شبکه عصبی از چندین عنصر بنام نرون تشکیل شده است و هریک از این نرونها با هم در ارتباطند. هدف شبکه عصبی مصنوعی تقلید از نرونهای سیستم عصبی انسان است که این کار با اتصال تعداد زیادی از پردازشگرها موسوم به نرون های مصنوعی یا گره ها انجام می شود [۴۵].

معمولترین ساختار های شبکه های عصبی به صورت زیر می باشند:

ساختار های لایه ای پیشخور (Feed Forward):

گروهی از نرون ها که نسبت به سیگنال های ورودی و خروجی در موقعیت مشابهی قرار دارند و به آن لایه (Layer) می گویند.

ساختارهای بازگشتی (Recurrent or Feedback):

ساختاری است که در آن خروجی یک نرون، ورودی نرون دیگر بوده به نحوی که چند نرون در یک حلقه بسته جای می گیرند.

ساختار های مشبک (Lattice):

ساختاری که در آن ارتباط نرون ها با مفهوم لایه یا حلقه همراه نیست.

از جمله کاربردهای شبکه عصبی ، کلاس بندی ، استخراج ویژگی و... می باشد.

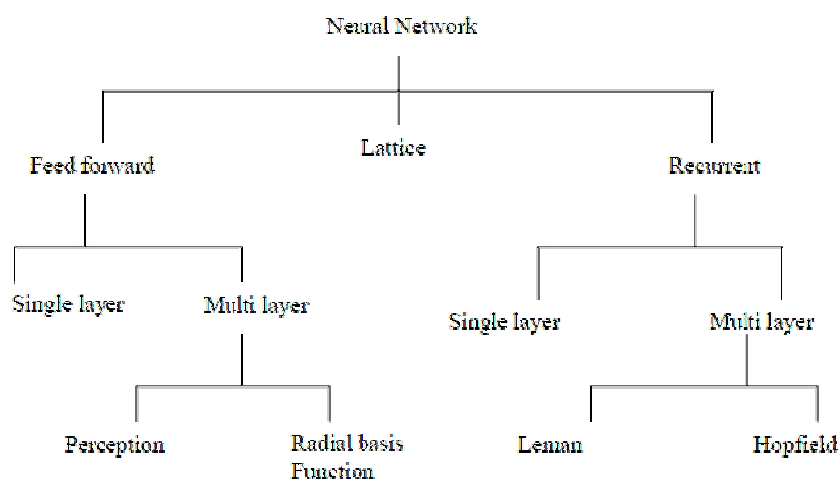
یکی از کاربردهای شبکه عصبی در زمینه بهسازی گفتار است.

۲-۳ شبکه عصبی

شبکه های عصبی نوعی مدلسازی ساده شده از سیستمهای عصبی واقعی هستند که کاربرد فراوانی در حل

مسائل مختلف در علوم دارند. حوزه کاربرد این شبکه ها آنچنان گسترده است که از کاربردهای طبقه بندی گرفته تا کاربردهایی نظیر درون یابی، تخمین، آشکارسازی و ... را شامل می شود. شاید مهمترین مزیت این شبکه ها، توانایی وافر آن ها در کنار سهولت استفاده از آن ها باشد.

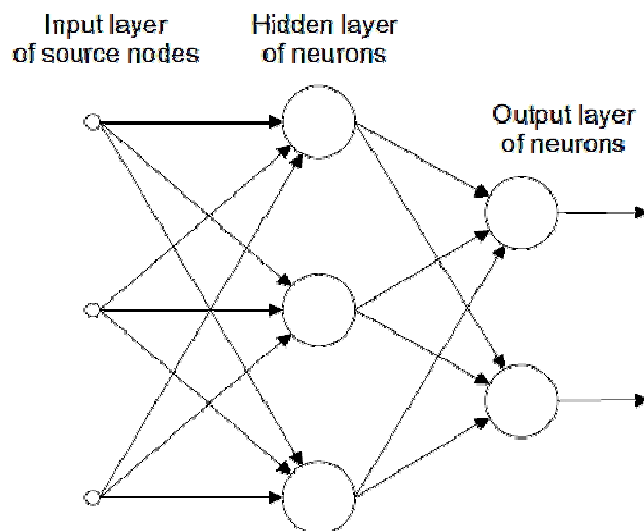
در شکل زیر ساختارهای مختلف شبکه عصبی را مشاهده می کنید.



شکل ۳-۱: طبقه بندی ساختارهای مختلف شبکه عصبی

یکی از ساده ترین و در عین حال کارآمدترین چیدمان های پیشنهادی برای استفاده در مدل سازی عصبی واقعی، مدل پرسپترون چندلایه (Multi layer perceptron) یا به اختصار MLP می باشد که از یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل یافته است. در این ساختار، تمام نورون های یک لایه به تمام نورون های لایه بعد متصلند. این چیدمان اصطلاحاً یک شبکه با اتصالات کامل را تشکیل می دهد.

شکل ۳-۲ شمای یک شبکه پرسپترون سه لایه را نشان می دهد. به سادگی می توان استنباط نمود که تعداد نرون های هر لایه، مستقل از تعداد نرون های دیگر لایه ها می باشد.



شکل ۳-۲: شبکه عصبی Feed-Forward چند لایه

۳-۲-۱ آموزش شبکه به روش پس انتشار خطا

به طور کلی شبکه های عصبی مصنوعی از لحاظ یادگیری بر دو دسته اند: شبکه های وزن ثابت و شبکه های با وزن متغیر (شبکه های یادگیرنده). خود شبکه های یادگیرنده نیز به دو دسته باسرپرست (Supervised) و بدون سرپرست (Unsupervised) تقسیم می شوند. در شبکه های باسرپرست، در فاز آموزش از نمونه هایی استفاده می گردد که خروجی ایده آل متناظر با آن ها از پیش دانسته است. به عبارت دیگر در این گونه شبکه ها، نمونه های داده ورودی، برچسب دارند. در شبکه های بی سرپرست، بر اساس یک معیار (مثلاً فاصله) و بر اساس نوعی رقابت، خروجی مورد نظر در کلاس جداگانه قرار می گیرد.

با توجه به این که شبکه عصبی، مدل ساده شده اعصاب بدن است، درست به مانند آن ها قابلیت یادگیری دارد. به عبارت دیگر، شبکه با استفاده از اطلاعاتی که از ورودی و توسط سرپرست خود دریافت می کند، قادر به فراگیری روند موجود در الگوهاست. لذا به طور مشابه با انسان، روند یادگیری در شبکه عصبی نیز از مدل های انسانی الهام گرفته است بدین صورت که مثال های بسیاری را به دفعات بایستی به شبکه ارائه نمود تا بتواند با تغییر وزنهای شبکه، خروجی موردنظر را دنبال کند. ارائه نمونه داده های ورودی به شبکه عصبی به دو روش

امکان پذیر است:

روش ارائه یک جا (Batch mode): در این روش، تمام نمونه ها به شبکه ارائه می گردند و در آخر، خطای شبکه نسبت به کل نمونه ها محاسبه گشته و وزن ها بر اساس آن خطا تغییر می کنند. در مرحله بعد، مجدداً تمام داده ها یکبار دیگر به شبکه ارائه شده و روند فوق نظیر به نظیر انجام می پذیرد تا نهایتاً خطا به سطح قابل قبولی برسد. مسلماً این روش پیچیده و زمان بر بوده و نیاز به حافظه زیادی دارد. همچنین امکان گیر کردن الگوریتم در مینیمم های محلی وجود دارد.

روش ارائه الگو (Pattern mode) : در این روش، در هربار نمونه ها به صورت تک تک به شبکه داده شده و خطای متناظر با همان داده بلافاصله محاسبه شده و بر اساس آن، وزنهاى شبکه تغییر می کنند. سپس نمونه بعدی به شبکه ارائه شده و روند بالا مشابهاً انجام می پذیرد. چون در این روش، در هر گام، اصلاح وزن ها بر اساس هر نمونه انجام می پذیرد، الگوریتم همگرایی خوبی داشته و با توجه به ماهیت تصادفی موجود در ارائه تکی داده ها، خطر مینیمم های محلی منتفی است.

به منظور آموزش شبکه و اصلاح وزن ها تا رسیدن به یک خطای معنادار، روش های بسیار زیادی وجود دارد. یکی از مشهورترین این روش ها، الگوریتم پس انتشار خطا (Error back propagation algorithm) است که در ادامه توضیح داده می شود [۴۵].

۳-۲-۲ الگوریتم پس انتشار خطا

این الگوریتم که در سال ۱۹۸۶ توسط روم لهارت و مک کلیلاند پیشنهاد گردید، در شبکه های عصبی پیش سو (Feed forward) مورد استفاده قرار می گیرد. پیش سو بودن به این معناست که نرون های مصنوعی در لایه های متوالی قرار گرفته اند و خروجی (سیگنال) خود را رو به جلو می فرستند. واژه پس انتشار نیز به معنای این است که خطاها به سمت عقب در شبکه تغذیه می شوند تا وزنها را اصلاح کنند و پس از آن، مجدداً ورودی مسیر پیش سوی خود تا خروجی را تکرار کند. روش پس انتشار خطا از روش های باسرپرست است به این

مفهوم که نمونه های ورودی برچسب خورده اند و خروجی مورد انتظار هر یک از آنها از پیش دانسته است. لذا خروجی شبکه با این خروجی های اید ه آل مقایسه شده و خطای شبکه محاسبه می گردد. در این الگوریتم ابتدا فرض بر این است که وزن های شبکه به طور تصادفی انتخاب شده اند. در هر گام خروجی شبکه محاسبه شده و بر حسب میزان اختلاف آن با خروجی مطلوب، وزنها تصحیح می گردند تا در نهایت این خطا، می نیم شود. در الگوریتم پس انتشار خطا، تابع تحریک هر عصب به صورت جمع وزندار ورودی های مربوط به آن عصب در نظر گرفته می شود. بدین ترتیب با فرض این که w وزن های متناظر بین لایه ورودی و لایه بعد باشد می توان نوشت:

$$A_j(\bar{x}, \bar{w}) = \sum_{i=0}^n x_i w_{ij} \quad (1-3)$$

به وضوح می توان دید که خروجی تابع تحریک عصب فقط به ورودی و وزنها متناظر بستگی دارد. حال بایستی به بررسی ارتباط خطا با ورودی ها، وزن ها و خروجی ها بپردازیم. برای این کار روش های متفاوتی وجود دارد که برخی از مهمترین آنها عبارتند از:

- روش گرادیان شیب (Gradient descent)

- روش نیوتن

- روش اندازه حرکت (Momentum)

- روش آنتروپی متقابل (Cross entropy)

- روش Marquardt-Levenberg

در این پایان نامه، از روش گرادیان شیب به دلیل این که ساده ترین و درعین حال پرکاربردترین روش است، استفاده کرده ایم. که در ادامه روابط آن بیان خواهد شد. در روش گرادیان شیب، ابتدا یک تابع هزینه درجه دوم تعریف می گردد. هدف نهایی الگوریتم پس انتشار خطا، مینیم کردن این تابع هزینه است. بر اساس روش گرادیان شیب، با توجه به درجه ۲ و مثبت بودن تابع هزینه، فرض می شود این تابع رفتاری سهموی دارد. لذا برای رسیدن به مینیم کلی آن بایستی در خلاف جهت شیب تابع حرکت کنیم. بنابراین با فرض این که وزن

ها در ابتدای کار به صورت تصادفی انتخاب شده باشند، باید شیب تابع خطا را نسبت به وزنهای محاسبه نموده و در جهت خلاف آن، وزنهای را تغییر دهیم و این روند را تا آنجا ادامه دهیم که به مینیمم کلی یا یک خطای قابل قبول برسیم.

۳-۲-۳ فواید شبکه عصبی

استفاده از شبکه عصبی فواید زیر را دارد:

- غیر خطی بودن
- یادگیری و تطبیقی
- پردازش توزیع شده بصورت موازی
- آنالیز و طراحی یکنواخت
- سیستم چند متغیری

۳-۲-۴ محدودیتهای شبکه عصبی

روش نظام مندی وجود ندارد تا ساختار شبکه عصبی را برای کاربردهای مختلف توصیه کند؛ بطور مثال، چه تعداد لایه بایستی در شبکه استفاده شود، چه تعداد نورون در لایه ها بکار رود. که این معمولا با روش سعی و خطا بدست می آید. همچنین زمان آموزش طولانی برای شبکه عصبی وجود دارد تا بخوبی آموزش ببیند و هرچه سیستم پیچیده تر، زمان طولانی تر خواهد شد. یکی دیگر از محدودیت های اینست که این شبکه عصبی مشخصات یک جعبه سیاه را دارد و درک اطلاعات ذخیره شده در شبکه مشکل است [۳۰].

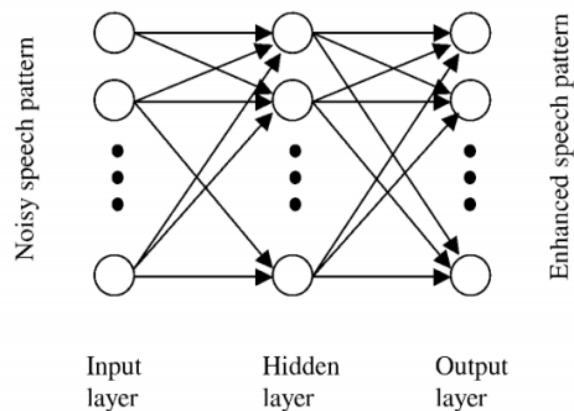
۳-۳ بهسازی گفتار با استفاده از شبکه عصبی:

شبکه عصبی متداول استفاده شده ، همانطور که در شکل ۳-۳ دیده می شود، شبکه ای ۳ لایه است. لایه ابتدایی لایه ورودی می باشد. در انتها لایه خروجی و بین این دو لایه، لایه میانی قرار می گیرد. در شکل

۳-۳ نمای کلی آن در آمده است.

هر لایه شامل چندین نرون می باشد. ورودی های هر نرون مجموع وزنداری از نرونهای لایه قبل است که از توابع اکتیو ساز عبور می کند. بین تمام نودهای لایه ورودی و تمام نودهای لایه میانی و همچنین نودهای لایه میانی با لایه خروجی ارتباط برقرار است و به این حالت fully connected گویند.

در روش بهسازی گفتار با شبکه ی عصبی، معمولاً سیگنال گفتار نویزی به عنوان ورودی استفاده می شود و سیگنال بهبود یافته از لایه خروجی بدست می آید. اگر x_k را سیگنال تمیز در نظر بگیریم و y_k را ورودی نویزی پنجره بندی شده، شبکه عصبی با تنظیم وزندهای بین لایه های خود، امید ریاضی شرطی $E[x_k | y_k]$ را تقریب می زند [۲۷]. فرض بر پایدار ماندن نویز و سیگنال نویزیست.



شکل ۳-۳: شبکه عصبی [۲۴]

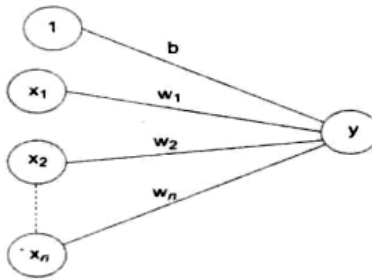
از مزایای این روش محاسبات قوی، انعطاف پذیری، درجه آزادی بالا جهت اصلاح، تقریب توابع غیر خطی می باشد. از معایب این روش مثال زیاد برای یادگیری شبکه و نیاز به داده آموزشی دقیق خواهد بود.

۳-۳-۱ مثالی از بهسازی گفتار با استفاده از شبکه عصبی

در این قسمت ابتدا به معرفی شبکه عصبی به نام آدالین پرداخته شده است.

این شبکه یک خروجی دارد که ورودی را از واحدهای مختلف دریافت می کند. شکل ۳-۴ نمایی کلی از شبکه

آدالاین را نشان می‌دهد.



شکل ۳-۴: نمای کلی از شبکه آدالاین [۲]

از ویژگی‌های این شبکه عمل کردن بصورت فیلتر وفقی و معماری ساده همچنین سرعت بالای پردازش آن را می‌توان برشمرد. این شبکه در زمینه بهسازی گفتار قابل استفاده می‌باشد. در پیاده سازی شبکه، اولین مرحله ایجاد گفتار آلوده نویزی است. سیگنال نویز به عنوان ورودی و سیگنال آلوده ی نویزی به عنوان تارگت به شبکه اعمال می‌شود. خروجی شبکه تخمینی از سیگنال نویزی خواهد بود. در این شبکه هدف کم کردن اختلاف سیگنال خروجی و سیگنال مطلوبست. این شبکه با الگوریتم LMS (least mean squares) وزنه‌های خود را به روز می‌کند. سیگنال تمیز تخمین زده شده، از اختلاف تارگت و خروجی شبکه حاصل می‌شود. بلوک دیاگرام روند طی شده در شکل ۳-۵ نشان داده شده است. مراحل الگوریتم حذف نویز در زیر آمده است:

- تنظیم نرخ یادگیری برابر α

- تنظیم وزنها بصورت $(\alpha - \text{randn}(1,p) * 1.2)$ که p برابر با ابعاد سیگنال ورودی می‌باشد. منظور از randn مقادیر رندوم بین صفر و یک می‌باشد.

- تنظیم سیگنال آلوده‌ی نویزی به عنوان تارگت (T)

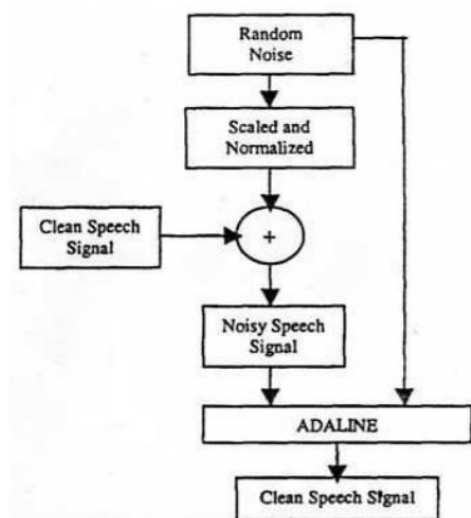
- تنظیم سیگنال نویز به عنوان ورودی (x)

- به روز کردن خروجی (y) و خطا (E) در هر مرحله مطابق با رابطه (۳-۲):

$$y(i) = w(i) \times X(i) + b(i) \quad , \quad E(i) = T(i) - y(i) \quad (۳-۲)$$

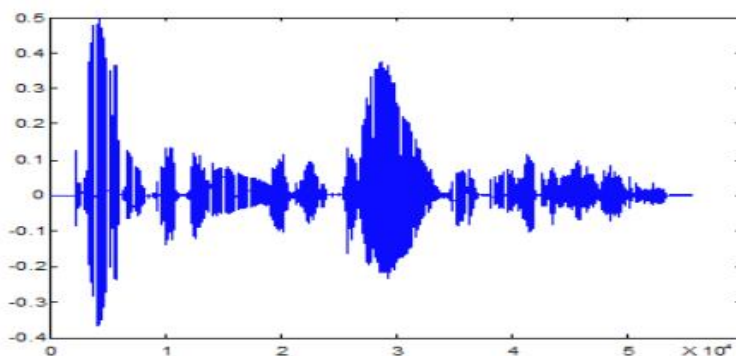
- تنظیم وزنها و بایاس (b) مطابق رابطه (۳-۳):

$$w(i+1) = w(i) + \alpha(E(i) \times I(i)) \quad , \quad b(i+1) = b(i) + \alpha(E(i)) \quad (3-3)$$

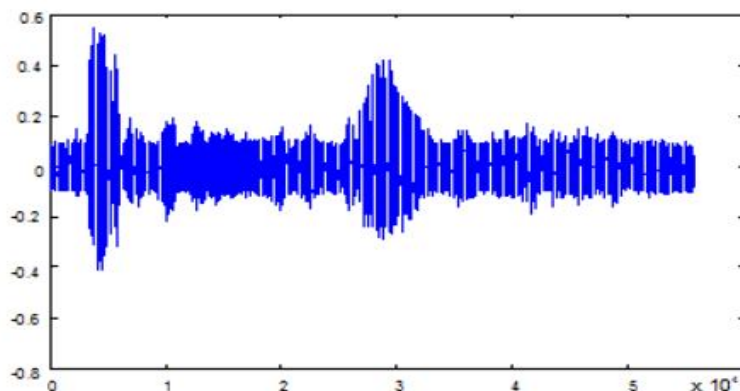


شکل ۳-۵: بلوک دیاگرام بهسازی با شبکه آدالاین [۲۲]

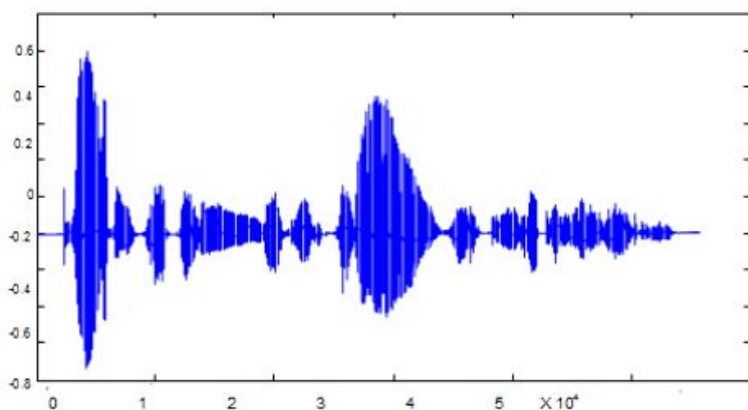
شکل ۳-۶، سیگنال گفتار تمیز و بدون نویز را نشان می‌دهد. این سیگنال با نویز رندومی جمع شده و گفتار نویزی را در شکل ۳-۷ بوجود آورده است. سیگنال حاصل به عنوان ورودی به شبکه اعمال شده است. نرخ یادگیری (α) برابر با ۰.۵ در نظر گرفته شده است. سیگنال گفتار تمیز، آلوده به نویز و گفتار بهبود یافته را در اشکال زیر مشاهده می‌کنید [۲].



شکل ۳-۶: سیگنال گفتار تمیز



شکل ۳-۷: سیگنال گفتار آلوده به نویز



شکل ۳-۸: سیگنال گفتار بهبود یافته

۳-۴ تبدیل موجک

مفهوم موجک در اوایل ۱۹۸۰ برای اولین بار مطرح گردید [۴۲]. از موجک ها به عنوان ابزار مناسبی در ریاضیات ، فیزیک ، پردازش سیگنال، محاسبات عددی ، فشرده سازی اطلاعات و... بهره گیری می شود. تبدیل موجک یک تبدیل انتگرالی است که از توابع موجک پایه استفاده می کند [۴۵].

تبدیل موجک، عملیاتی می باشد که با استفاده از ترکیب نسخه های اصلاح شده از توابع هسته، تابع را می توان نمایش داد. توابع هسته توابع مادر و نسخه های اصلاح شده آن نیز توابع فرزند نامیده می شوند. یک تابع را زمانی تابع موجک مادر می نامیم که بتواند در شرط پذیرفتگی صدق کند. یک تابع مادر می بایست نوسانی بوده

و همچنین دامنه آن نیز به سرعت به سمت صفر میل کند. اگر این دو شرط را همراه با شرط صفر بودن مولفه جریان مستقیم یا مولفه فرکانس صفر را برای یک تابع در نظر بگیریم، گفته می شود شرط پذیرفتگی مورلت در تابع مربوطه برقرار می باشد. به عبارت دیگر برای یک تابع $h \in L^2(R)$ شرط پذیرفتگی برقرار می باشد [۳۱] اگر:

$$c_h = \int_{-\infty}^{\infty} \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty \quad (۴-۳)$$

که $H(\omega)$ تبدیل فوریه تابع $h(t)$ می باشد. ثابت C_h ، ثابت پذیرفتگی تابع $h(t)$ و لازم است که برای وجود عکس تبدیل موجک شرط فوق برقرار باشد. هر تابعی که در شرط پذیرفتگی فوق صدق کند، می تواند یک تابع موجک مادر باشد. برای تابع داده شده $h(t)$ ، شرط $c_h < \infty$ برقرار است تنها اگر $H(0) = 0$ باشد. موجکها ذاتا همانند باند عبور یک فیلتر در حوزه فوریه و یا به طور مشابه $\int_{-\infty}^{\infty} h(t) dt = 0$ که نشانگر مقدار متوسط صفر در حوزه زمان می باشد، عمل می کنند. تبدیل موجک از یک تابع $f \in L^2(R)$ با توجه به شرط پذیرفتگی تابع موجک مادر $h(t)$ همانند زیر تعریف می شود [۳۶]:

$$W_f(a, b) = \int_{-\infty}^{\infty} f(t) h_{a,b}^*(t) dt \quad (۵-۳)$$

که * نشانگر مزدوج مختلط می باشد. اگر چه اکثر توابع موجک، توابع حقیقی می باشند، توابع موجک فرزند از یک تابع موجک مادر $h(t)$ با تغییر در پارامترهای انتقال و مقیاس آن تولید می شود:

$$h_{a,b}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right) \quad (۶-۳)$$

که $a > 0$ پارامتر مقیاس و b پارامتر انتقال می باشد. ثابت $a^{-1/2}$ برای نرمال سازی انرژی تابع موجک می باشد که از این طریق انرژی تابع موجک فرزند با انرژی تابع موجک مادر برابر می شود.

اخیرا توابع موجک جدیدی که به طور کلی جالبتر از توابع موجک متعامد می باشند، پدید آمده اند که می توان از آنها به عنوان توابع مادر استفاده نمود. این نوع توابع موجک توانایی های این توابع را برای کاربردهایی نظیر

فرآیندهای سیگنال تطبیقی و بویژه کاربردهای شبکه عصبی بطور چشمگیری افزایش می‌دهد. در ادامه به معرفی چند نمونه از توابع موجک خواهیم پرداخت [۳۷].

۱-۴-۳ انواع توابع موجک

۱-۱-۴-۳ تابع موجک Morlet

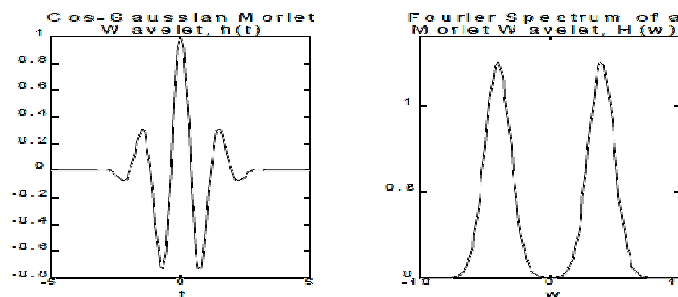
تابع پایه Morlet تابعی است که از ضرب پایه تبدیل فوریه با یک پنجره گوسی بدست آمده است:

$$h(t) = \exp(jw_0 t) \exp(-0.5t^2) \quad (۷-۳)$$

که قسمت حقیقی آن کسینوس تابع گوسی و قسمت موهومی آن سینوس تابع گوسی می باشد. تبدیل فوریه قسمت کسینوسی موجک آن برابر با شیف‌ت تابع گوسی به اندازه w_0 و $-w_0$ می باشد.

$$H(w) = \sqrt{\frac{\pi}{2}} (\exp[-0.5(w - w_0)^2] + \exp[-0.5(w + w_0)^2]) \quad (۸-۳)$$

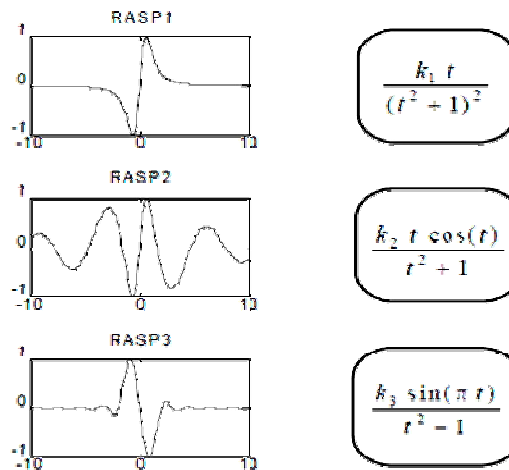
که تابع فوق یک تابع مثبت و زوج می باشد. شکل ۹-۳ یک تابع موجک Morlet و تبدیل فوریه آنرا برای $w_0 = 4$ نشان می دهد.



شکل ۹-۳: $h(t)$ موجک Morlet کوسینوس-گوسین و تبدیل فوریه اش $H(w)$

۲-۱-۴-۳ تابع موجک RASP (Rational functions with Second-order Poles)

توابع کسری با قطب های مرتبه دوم برآمده از تئوری قضیه مانده در متغیرهای مختلط می باشد. شکل ۱۰-۳ سه مثال از موجک مادر RASP را نشان می دهد.



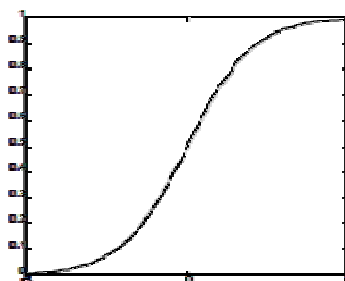
شکل ۳-۱۰: مثال هایی از موجک های RASP

۳-۴-۱-۳ تابع موجک SLOG (Superposed LOGistic functions)

توابع لجستیک مجموع، نتیجه از مجموع محدود ترمهای وزن داده شده و تاخیر داده شده از توابع لجستیک می باشد. یک تابع لجستیک یک نوع تابع یکنواخت افزایشی همانند تابع مجانبی سیگموید (Sigmoid) می باشد که معمولا در نرون خروجی یک شبکه عصبی به عنوان نمایش یک تابع استانه از آن استفاده می شود. مشخصه تابع لجستیک نمایش یک تابع تک قطبی نامتقارن به فرم زیر می باشد:

$$f_{logistic}(t) = \frac{1}{1 + e^{-t}} \quad (۳-۹)$$

شکل ۳-۱۱ یک تابع لجستیک را نمایش می دهد.



شکل ۳-۱۱: تابع سیگموید لجستیک

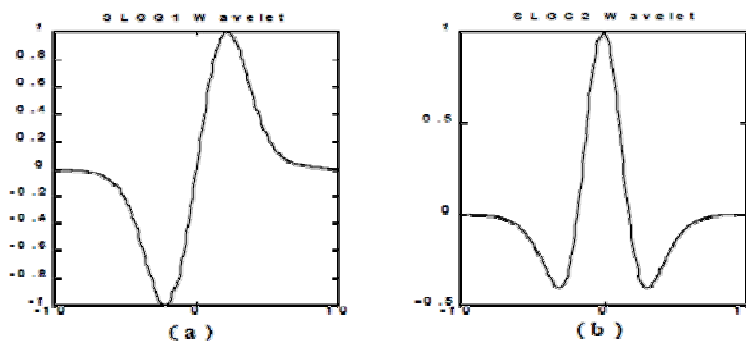
شکل ۳-۱۲ (الف) موجکهای مادر SLOG را نمایش می‌دهد که حاصل از جمع توابع سیگموئید زیر می‌باشد:

$$h_{slog1}(t) = \frac{1}{1+e^{-t+1}} - \frac{1}{1+e^{-t+3}} - \frac{1}{1+e^{-t-3}} + \frac{1}{1+e^{-t-1}} \quad (۱۰-۳)$$

این توابع موجک مادر یک تابع موجک با حداقل نوسان را نمایش می‌دهد. شکل ۳-۱۲ (ب) یک نوع دیگر از تابع موجک SLOG را نشان می‌دهد که یک نیم نوسان بیشتری را به نمایش می‌گذارد. در ساخت این تابع نیز یک مجموع محدودی از توابع لجستیکی تاخیر داده شده و وزن داده شده استفاده شده است:

$$h_{slog2}(t) = \frac{3}{1+e^{-t-1}} - \frac{3}{1+e^{-t+1}} - \frac{1}{1+e^{-t-3}} + \frac{1}{1+e^{-t+3}} \quad (۱۱-۳)$$

بطور مشابه همه کلاسهای توابع موجک SLOG از طریق همین روش با تعداد نوسانات بیشتری قابل تولید می‌باشد.



شکل ۳-۱۲: مثالی از موجک مادر SLOG

۴-۱-۴-۳ توابع موجک (POLYWOG (POLYnomials WindOwed with Gaussian)

تعداد زیادی از توابع موجک مادر با استفاده از توابع چند جمله ای و انواعی از توابع گوسی با نام توابع موجک POLYWOG ساخته می‌شوند. برای مثال تمام مشتقات یک تابع گوسی، از نوع POLYWOG می‌باشد و بنابراین می‌توان آنها را نیز به عنوان موجک مادر انتخاب نمود. نمونه ای از این توابع در زیر داده شده است:

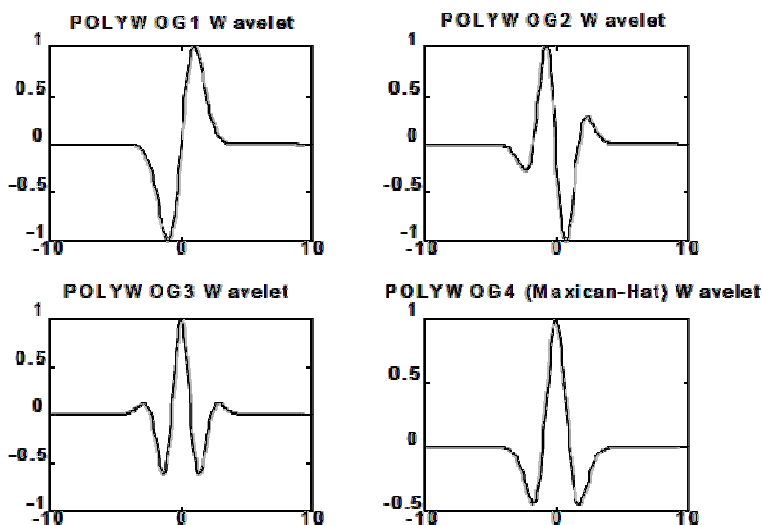
$$\text{POLYWOG1: } h_{\text{POLYWOG1}} = k_1 t \exp\left(-\frac{t^2}{2}\right), \quad k_1 = \sqrt{e} \quad (۱۲-۳)$$

$$\text{POLYWOG2: } h_{\text{POLYWOG2}} = k_2(t^3 - 3t) \exp\left(-\frac{t^2}{2}\right), \quad k_2 = 0.724 \quad (۱۳-۳)$$

$$\text{POLYWOG3: } h_{\text{POLYWOG3}} = k_3(t^4 - 6t^2 + 3) \exp\left(-\frac{t^2}{2}\right), \quad k_3 = \frac{1}{3} \quad (۱۴-۳)$$

$$\text{POLYWOG4: } h_{\text{POLYWOG4}} = (1 - t^2) \exp\left(-\frac{t^2}{2}\right) \quad (۱۵-۳)$$

شکل ۳-۱۳، چهار نمونه از مشتق تابع گوسی POLYWOG را نشان می‌دهد. انواع متفاوت توابع موجک POLYWOG ممکن است توسط دیگر عملگرهای محاسباتی نیز بدست آیند.



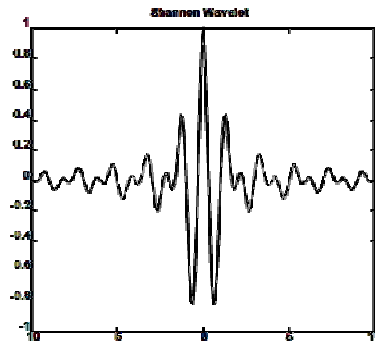
شکل ۳-۱۳: مثال هایی از موجک POLYWOG

۳-۴-۱-۵ تابع موجک Shannon

تابع موجک Shannon یک موجک مادر می‌باشد که یک پایه متعامد را در فضای $L^2(R)$ تولید می‌کند. این موجک به صورت گسسته در فرکانس می‌باشد و از این رو در حوزه زمان گسترش یافته‌اند. تابع موجک Shannon به صورت زیر تعریف می‌شود:

$$h_{\text{shannon}}(t) = \frac{\sin 2\pi t - \sin \pi t}{\pi t} \quad (۱۶-۳)$$

تابع موجک Shannon در شکل ۳-۱۴ نشان داده شده است.



شکل ۳-۱۴: تابع موجک Shanon

۳-۵ بهسازی گفتار با استفاده از تبدیل موجک :

سیستمهای بهبود گفتار مبتنی بر موجک موفقیت خود را در عمل نشان داده اند. با استفاده از این سیستم ها سیگنال صوتی در هریک از دو حوزه زمان و فرکانس با رزولوشن دلخواه تجزیه می گردد. سپس با استفاده از آستانه گذاری یا روشهای فیلتر کردن، اجزا سیگنال بهبودیافته، بازسازی می گردد [۳۸].

سیگنال گفتار که طبیعت پایین گذر دارد، انرژی اش در تعداد کمی از ابعاد موجک در مولفه های فرکانس پایین متمرکز شده است. ضرایب تبدیل موجک در محلهایی که انرژی سیگنال زیاد است؛ خود را بیشتر نشان می دهد.

نویز که بر روی سیگنال گفتار سوار می شود، انرژی کمتری نسبت به سیگنال گفتار خواهد داشت. بنابراین با آستانه گیری و صفر قرار دادن ضرایب با دامنه ی کوچکتر از حد آستانه، به نحو مطلوبی می توان نویز را حذف نمود، در حالیکه اطلاعات سیگنال اصلی حفظ شود [۳۹].

فرض کنید y دنباله مشاهدات نویزی و سیگنال x (با طول محدود) تخریب شده با نویز سفید گوسی n با متوسط صفر است:

$$y = x + n \quad (۳-۱۷)$$

در اینجا هدف بازسازی سیگنال x از مشاهدات نویزی y است. اگر w ماتریس تبدیل موجک گسسته باشد،

فرمول در حوزه موجک بصورت زیر نوشته می شود :

$$Y = X + N \quad (18-3)$$

که:

$$Y = w_y, \quad X = w_x, \quad N = w_n \quad (19-3)$$

با آستانه گذاری بر روی ضرایب Y ، تخمینی از سیگنال تمیز (X) در حوزه موجک، بدست آورده ایم. سیگنال تمیز x با استفاده از تبدیل موجک عکس، بصورت رابطه (۲۰-۳) تخمین زده می شود [۱۷].

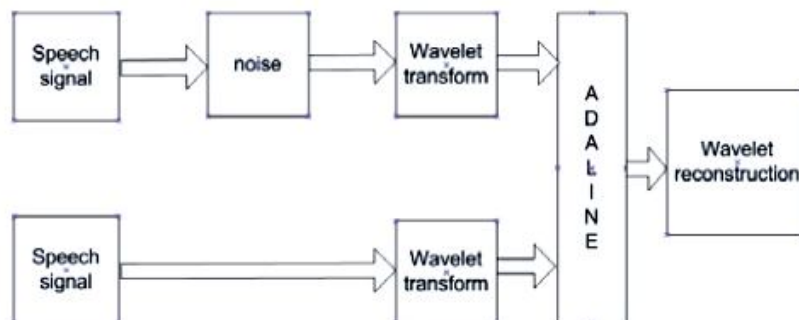
$$x = w^{-1}(X) \quad (20-3)$$

مقدار مناسب حد آستانه به روشهای مختلفی تعیین می شود.

۳-۵-۱ مثالی از کاربرد تبدیل موجک در بهسازی گفتار

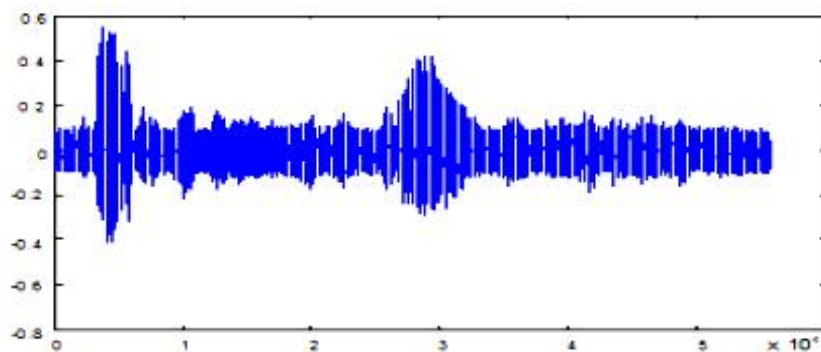
در قسمت بعد، از تبدیل موجک گسسته استفاده کرده و از ورودی شبکه آدالاین قبل از ورود به شبکه، تبدیل موجک گسسته گرفته می شود. به این ترتیب از ویژگی تبدیل موجک استفاده کرده و تعداد نمونه های ورودی را کاهش می دهیم و به انجام سریعتر روند بهسازی کمک خواهیم کرد.

در شکل ۳-۱۵ روند بیان شده آمده است. نتیجه حاصل از آزمایش این دو روش (روش آدالاین و ترکیب آن با شبکه عصبی) بر روی گفتار انگلیسی با فرکانس نمونه برداری ۸KHz، نشان داده است که در شبکه آدالاین کیفیت سیگنال تست شده تا ۱۰db و با روش شبکه آدالاین و تبدیل موجک تا ۸db بهبود یافته است [۲].

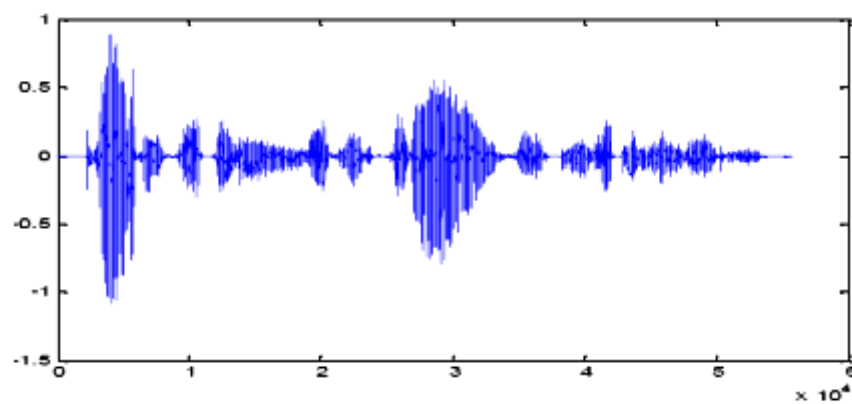


شکل ۳-۱۵: بلوک دیاگرام روش شبکه عصبی و موجک [۲]

در نهایت از سیگنال خروجی عکس تبدیل موجک گسسته گرفته می‌شود. نتایج حاصل از تکرار آزمایش با شبکه عصبی و تبدیل موجک بصورت زیر است [۲]:



شکل ۳-۱۶: سیگنال گفتار آلوده به نویز



شکل ۳-۱۷: سیگنال گفتار بهبود یافته

ویونت (wavelet neural network)

۴ شبکه ویونت (Wavelet Neural Network)

۱-۴ مقدمه

شبکه ویونت یک نوع شبکه عصبی پیش سو (Feed forward) می باشد که از توابع موجک به عنوان تابع فعال ساز استفاده می کند. این شبکه در سال ۱۹۹۲ توسط دانشمندان بنونیست (Benveniste) و ژانگ (Zhang) معرفی گردید.

استفاده ی این شبکه از توابع خانواده ی موجک به عنوان تابع فعالساز، کاهش قابل توجهی در پیچیدگی های محاسباتی در ساختار این شبکه ایجاد می کند. از کاربردهای این شبکه تقریب توابع و مدل کردن و شناسایی سیستم های دینامیکی می باشد. در ادامه به معرفی بیشتر این شبکه پرداخته خواهد شد.

۲-۴ معرفی شبکه ویونت

مدل مبتنی بر شبکه عصبی در مواجهه با مسائلی که طبیعتی بسیار غیرخطی دارند ، با مشکلاتی روبرو می شوند. این اشکال معمولا از عدم توانایی مدل شبکه عصبی – که از توابع فعالسازی نسبتا ساده استفاده می کنند- برای تطابق با ساختارهای دینامیکی سیستم واقعی ناشی می شود.

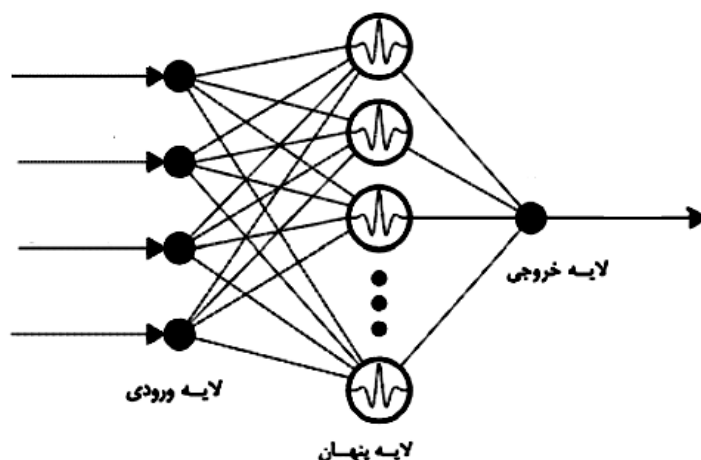
استفاده از تلفیق این مدلها با تبدیل موجک می تواند در راستای حل مشکل فوق، راهگشا واقع گردد. به کارگیری این شبکه تلفیقی به عنوان جایگزین مناسب در شبکه های عصبی پیش سو جهت تخمین و تقریب توابع غیرخطی اختیاری پیشنهاد می شود [۳۵].

در شبکه استاندارد پسخورد تابع فعال سازی نرون لایه پنهان یک تابع سیگموئید است.

حال آنکه شبکه های ویونت توابع موجک را به عنوان توابع فعالسازی نرون های لایه پنهان از شبکه پیش سو مورد استفاده قرار می دهد [۴۵].

این شبکه در تخمین و شناسایی سیستم های دینامیکی کاربرد دارد [۴۷]. شکل ۴-۱، شبکه ویونت را نشان

می دهد.



شکل ۴-۱: ساختار شبکه ویونت [۴۵]

دو روش اصلی برای ایجاد این شبکه وجود دارد:

در روش اول پردازش توسط شبکه عصبی و موجک بصورت مجزا انجام می گیرد. ابتدا سیگنال ورودی با مجموعه ای از ضرایب موجک در نرونهای لایه پنهان تجزیه می شود و سپس این ضرایب به عنوان ورودی نرونهای جمع کننده ای به کار میروند که وزندهای ورودی آنها توسط الگوریتم یادگیری تنظیم می شوند.

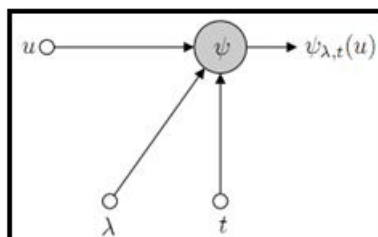
در روش دوم پارامترهای تبدیل موجک و وزندهای اتصالات شبکه به صورت همزمان توسط الگوریتم یادگیری تغییر یافته و به روز می شود. این دو مدل از نظر معماری شبکه مشابه می باشند [۲۶].

ساده ترین شکل مدل های ویونت ، مدلی با یک ورودی و یک خروجی است. هر کدام از نرونهای موجک را یک ویولون می نامیم. لایه پنهان شامل ویولون هایی است که پارامترهای ورودی اش ضرایب مقیاس و انتقال موجک هستند. ویولونها هنگامی که مقدار ورودی در محدوده کوچکی از بازه ورودی قرار دارد، خروجی غیر صفر تولید می کند.

خروجی هر ویولون به صورت زیر است:

$$\psi_{\lambda,t}(u) = \psi\left(\frac{u-t}{\lambda}\right) \quad (۱-۴)$$

در رابطه فوق λ پارامتر مقیاس و t پارامتر انتقال (تاخیر) است. شکل زیر ساختار یک ویولون تک ورودی را نشان می‌دهد [۲۰].



شکل ۲-۴: ساختار یک نرون موجک

به علت آنکه موجک‌ها توابعی هستند که به سرعت ناپدید می‌شوند:

در صورتیکه پارامتر مقیاس یک موجک بسیار کوچک در نظر گرفته شود، ممکن است بیش از حد به صورت محلی عمل نماید. در صورتیکه پارامتر انتقال یک موجک به صورت غیر متناسب انتخاب شود این امکان وجود دارد که خارج از محدوده موردنظر وارد عمل شود. بنابراین اختیار کردن تصادفی مقادیر ضرایب مقیاس و انتقال به دور از منطق است. یکی از روشهای انتخاب این مقادیر به این صورت است که ابتدا بازه $[a_k, b_k]$ که شامل مقادیر k امین مولفه بردارهای ورودی می باشد، مرکز تابع موجک j در مرکز متوازی السطوح تعریف شده $\{[a_k, b_k]\}$ با فاصله N_i است. برای k امین داده ورودی داریم [۱۵]:

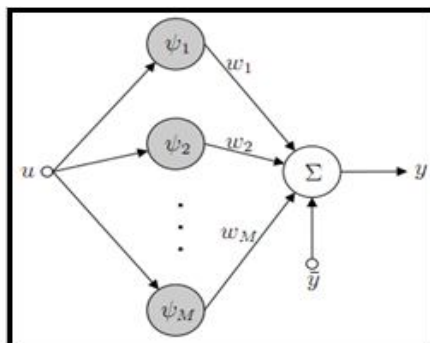
$$t_{jk} = 0.5(a_k + b_k) \quad , \quad \lambda_{jk} = 0.2(b_k - a_k) \quad (۲-۴)$$

۱-۲-۴ شبکه ویونت تک بعدی

برای یک ویونت تک ورودی تک خروجی ، خروجی ترکیب خطی وزنی توابع فعالساز ویولت می‌باشد و به صورت زیر قابل ارائه است:

$$y(u) = \sum_{i=1}^M w_i \psi_{\lambda_i, t_i}(u) + \bar{y} \quad (۳-۴)$$

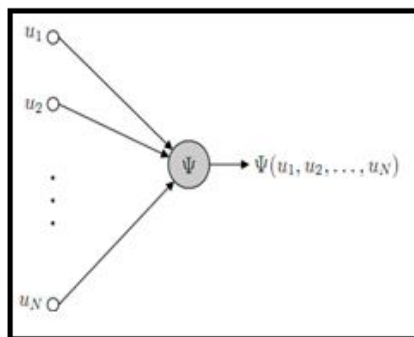
در این رابطه $\psi(u)$ تابع موجک و w وزنه‌های شبکه می باشد. عبارت $\bar{y}$ برای توابع تخمین دارای میانگین غیر صفر افزوده شده است، زیرا توابع موجک همگی دارای میانگین صفر هستند [۲۶]. در شکل ۳-۴ شبکه ویونت یک بعدی نشان داده شده است.



شکل ۳-۴: شبکه ویونت یک بعدی

۲-۲-۴ شبکه ویونت چند بعدی

در این شبکه ورودی به صورت بردار است.



شکل ۴-۴: شبکه ویونت چند بعدی

خروجی در این حالت به صورت ترکیبی از ویولونهای چندبعدی است. خروجی هر ویولون با رابطه زیر بیان می شود:

$$\psi(u_1, \dots, u_N) = \prod_{n=1}^N \psi_{\lambda_n, t_n}(u_n) \quad (4-4)$$

خروجی در این حالت از رابطه زیر بدست می آید:

$$y_j = \sum_{i=1}^M w_{ij} \psi_i(u_1, \dots, u_N) + \bar{y}_j \quad j = 1, \dots, K \quad (5-4)$$

رابطه خروجی و ورودی در این حالت بصورت زیر می باشد:

$$Y(U) = \sum_{i=1}^M w_i \psi_i(u) + \bar{Y} \quad (6-4)$$

که در آن [۲۶]:

$$Y = (y_1, \dots, y_K) \quad (7-4)$$

$$W_i = (w_{i1}, \dots, w_{iK})$$

$$U = (u_1, \dots, u_N)$$

$$\bar{Y} = (\bar{y}_1, \dots, \bar{y}_K)$$

۳-۲-۴ الگوریتم آموزش

یکی از کاربردهای شبکه عصبی موجک تخمین توابع می باشد.

Benveniste و Zhang در سال ۱۹۹۲ الگوریتمی برای آموزش این شبکه و تنظیم ضرایب ارائه دادند. در این الگوریتم از روش گرادیان تصادفی استفاده شده است. متوسط مربع اختلاف بین سیگنال اصلی (مطلوب) و سیگنال تخمین زده شده ، معیار خطا در نظر گرفته شده است.

چنانچه سیگنال تخمین y را بر اساس پارامترهای شبکه به صورت زیر تعریف کنیم، خواهیم داشت :

$$y_\theta(u) = \sum_{i=1}^M w_i \psi_i\left(\frac{u - t_i}{\lambda_i}\right) + \bar{y}_j \quad (8-4)$$

که θ بردار پارامترهای شبکه ، شامل w, t, λ و $\bar{y}$ می باشد.

تابع هدف زیر باید مینیم شود. در این رابطه $f(u)$ تابع مطلوبست که باید تخمین زده شود.

$$c(\theta) = \frac{1}{2} E\{(y_\theta(u) - f(u))^2\} \quad (9-4)$$

گرادیان هر پارامتر θ به صورت مشتق جزئی تعریف می شود. براین اساس روابط زیر خواهیم داشت:

$$\frac{\partial c}{\partial \bar{y}} = e_K \quad (10-4)$$

$$\frac{\partial c}{\partial \mathcal{W}_i} = e_K \psi(z_{Ki})$$

$$\frac{\partial c}{\partial t_i} = -e_K \mathcal{W}_i \frac{1}{\lambda_i} \psi(z_{Ki})$$

$$\frac{\partial c}{\partial \lambda_i} = -e_K \mathcal{W}_i \left(\frac{u_K - t_i}{\lambda_i^2} \right) \psi(z_{Ki})$$

که در آن:

$$e_K = y_\theta(u_K) - f(u_K) \quad , \quad z_{Ki} = \frac{u_K - t_i}{\lambda_i} \quad , \quad \psi(z) = \frac{d\psi(z)}{dz} \quad (11-4)$$

می باشد [۲۳].

در پیاده سازی این الگوریتم نرخ یادگیری و تعداد تکرارها باید انتخاب شود. نرخ یادگیری γ مقداری در بازه [۰ و ۱) می باشد که با سرعت همگرایی الگوریتم متناسب است. گرادیان هر پارامتر از استفاده برای به روز شدن ، در این مقدار ضرب می شود. تعداد تکرارها مشخص کننده چند بار ورودیها باید برای آموزش وارد شوند. هرچه این مقدار بزرگتر باشد همگرایی نتیجه شبکه به تابع بیشتر خواهد بود ولی زمان اجرای الگوریتم افزایش خواهد یافت [۲۶].

شبکه ی پیشنهادی و بکارگیری آن در بهسازی گفتار

۵ شبکه‌ی پیشنهادی و بکارگیری آن در بهسازی گفتار

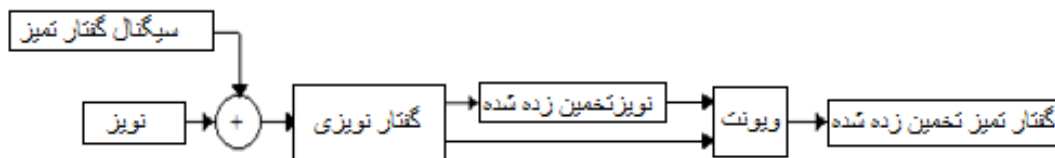
۵-۱ مقدمه

پس از توضیحاتی که در فصل‌های قبل درباره اصول روش‌های بهسازی گفتار و معرفی روش‌های متداول بیان شد، در این فصل روش پیشنهادی مورد بررسی قرار خواهد گرفت.

حذف و یا کاهش نویز یکی از شاخص‌ترین کاربردهای عملی فیلترهای فعال - وفقی می‌باشد. زمینه‌های زیادی وجود دارد که در آنها تکنیک‌های قدیمی فیلتر کردن نمی‌تواند در زمینه‌ی کاهش نویز استفاده گردد. تعداد ضرایب در حدود چند صد ضریب و عدم عملکرد مطلوب در فرکانس‌های پایین از معایب آنها می‌باشد. فیلترهای فعال - وفقی دارای این قابلیت بوده که پارامترهایشان را به صورت وفقی تنظیم نموده و طراحی آنها نیاز چندانی به دانستن اطلاعات اولیه در زمینه‌ی ویژگی‌های منبع نویز و سیگنال ورودی ندارد. این تکنیک با تولید سیگنالی برابر با سیگنال نویز به حذف و یا کاهش نویز می‌پردازد. همانطور که بیان شد، یک روش قدیمی و پایه برای کاهش نویز استفاده از تئوری فیلتر وینر بوده که همانند الگوریتم‌های وفقی مانند کمترین میانگین مربعات و یا کمترین مربعات بازگشتی، می‌تواند با محیط نایستاتطبق یابد. در این پایان نامه با استفاده از ویونت و کاربرد آن به صورت فیلتر وفقی، عمل کاهش نویز بر روی سیگنال‌های نویزی گفتار انجام شده است. در فصل آخر نیز به بررسی نتایج ارزیابی روش پیشنهادی و مقایسه آن با روش‌های دیگر بهسازی گفتار براساس معیارهای ارزیابی بیان شده، پرداخته خواهد شد.

بلوک دیاگرام روش پیشنهادی در شکل ۵-۱ آورده شده است. کلیه بخش‌های این بلوک در ادامه فصل توضیح داده خواهند شد. در شبکه ویونت هدف ما حداقل کردن خطا می‌باشد. ویونت با تنظیم پارامترهای خود نویزی را مدل می‌کند که به نویز سیگنال آلوده‌ی نویزی شبیه بوده و با حذف آن از سیگنال نویزی، سیگنال گفتار تمیز به خوبی تخمین زده می‌شود.

در ادامه فصل با محاسبه‌ی معیارهای ارزیابی، میزان بهبود سیگنال گفتار و عملکرد شبکه‌ی پیشنهادی را بررسی خواهیم کرد.



شکل ۵-۱: بلوک دیاگرام روش پیشنهادی

۵-۲ شبکه ویونت

ساختار شبکه ویونت را می‌توان به دو بخش کلی زیر تقسیم کرد:

- معماری شبکه ویونت

- الگوریتم آموزش شبکه ویونت

از نظر معماری، شبکه در لایه‌ی پنهان خود دارای توابع موجک می‌باشد. هر تابع موجک شامل پارامتر مقیاس و انتقال می‌باشد. لایه‌ها توسط وزنهای شبکه به هم متصل می‌شوند. در این قسمت هدف یافتن پارامتر مقیاس و انتقال و وزنهای شبکه می‌باشد. در قسمت الگوریتم آموزش هدف پیاده کردن الگوریتمی است که به کمک آن بتوان تابع خطا را مینیمم کرد [۳۲].

۵-۲-۱ شبکه پیشنهادی

شبکه ویونت (WNN) پیشنهادی مطابق شکل ۵-۲ از ۴ لایه تشکیل شده است، که شامل یک لایه ورودی (لایه i)، لایه موجک مادر (لایه j)، یک لایه موجک (لایه k) و لایه خروجی (لایه o) می‌باشد. ورودی شبکه e_1 و e_2 می‌باشد. برای هر گره i در لایه ورودی، ورودی شبکه و خروجی های شبکه مطابق زیر است:

$$net_i^1 = x_i^1 \quad (۵-۱)$$

$$y_i^1 = f_i^1(net_i^1) = net_i^1, \quad i = 1, 2$$

در حالی که $x_1^l = e1$ و $x_2^l = e2$ است.

گروهی از موجکها توسط مقیاس و انتقالات اعمالی روی یک تابع ثابت که موجک مادر نامیده می‌شود، ساخته می‌شود. در لایه موجک مادر، هر گره نقش موجک ϕ_j را که از موجک مادرش استخراج شده است اجرا می‌کند.

موجک مادری که در این روش بکار می‌رود مشتق اول تابع گوسین $\phi(x) = -x e^{-x^2/2}$ می‌باشد [۳۳].
برای گره ژم داریم:

$$net_j^2 = \frac{x_i^2 - m_{ij}}{\sigma_{ij}} \quad (۲-۵)$$

$$y_j^2 = f_j^2(net_j^2) = \phi_j(net_j^2), j = 1, \dots, n$$

m_{ij} و σ_{ij} انتقال و مقیاس در ترم ژام از ورودی x_i^2 به گره لایه موجک مادر می‌باشد و n تعداد کلی موجکها با توجه به گره های ورودی است. بعلاوه، هر گره k در لایه موجک توسط Π نشان داده شده است که سیگنالهای ورودی و خروجی ها را ضرب می‌کند.

برای k امین گره قانون:

$$net_k^3 = \prod_j w_{jk}^3 x_j^3, \quad y_k^3 = f_k^3(net_k^3) = net_k^3, \quad k = 1, \dots, l \quad (۳-۵)$$

x_j^3 ورودی ژام به گره لایه موجک

w_{jk}^3 وزنهای بین لایه موجک مادر و لایه موجک، فرض شده است تا واحد باشند.

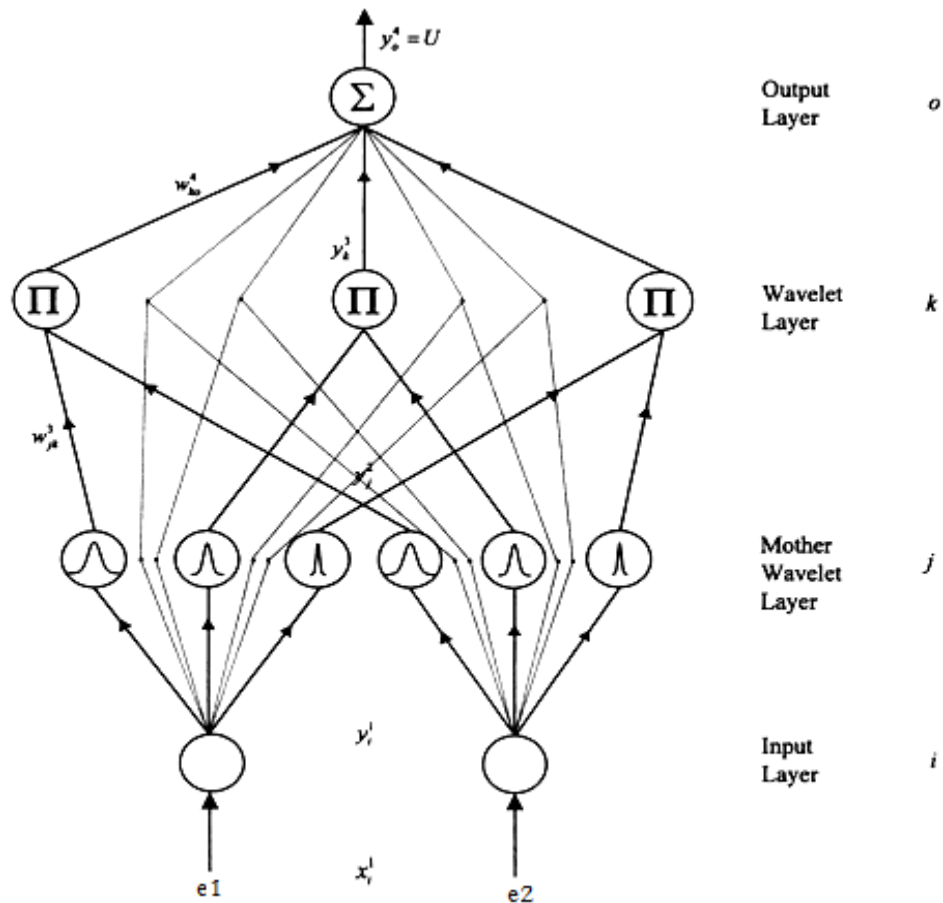
$l = n/i$ تعداد موجکها است که هر گره ورودی همان تعداد گره های موجک مادر را دارد.

تک گره o در لایه خروجی بصورت Σ علامت گذاری شده است، که خروجی کلی را با جمع همه سیگنالهای ورودی محاسبه می‌کند:

$$net_o^4 = \sum_k w_{ko}^4 x_k^4, \quad y_o^4 = f_o^4(net_o^4) = net_o^4, \quad o = 1 \quad (۴-۵)$$

وزن متصل w_{ko}^4 نیروی کار خروجی o ام بدست آمده توسط موجک k ام است.

x_k^4 ورودی k ام به گره لایه خروجی، و $y_o^4 = U$ خروجی می‌باشد.



شکل ۵-۲: ساختار ویونت چهار-لایه [۳۳]

۵-۲-۱-۱ مکانیزم آموزش on-line

روش مورد استفاده برای بروز رسانی وزنها و پارامترها در شبکه به این صورت می‌باشد که ابتدا خطا را از تابع هدف استخراج می‌کنیم و سپس با قانون یادگیری بازگشتی، خطا را به همه گره‌ها و پارامترها اعمال می‌کنیم. مکانیزم آموزش on-line شبکه ویونت با استفاده از روش شیب گرادیان با سرپرست انجام شده است. ابتدا تابع هدف بصورت $E = 0.5 e^2$ تعریف می‌شود. و خطا (e) به صورت رابطه‌ی (۵-۵) تعریف می‌شود.

$$e = \theta_t - \theta_o \quad (5-5)$$

در این رابطه، θ_t سیگنال تارگت و θ_o سیگنال خروجی شبکه می باشد.

مکانیزم آموزش on-line مبتنی بر پس انتشار مطابق زیر توصیف می شود. در لایه خروجی، عبارت خطا مطابق زیر انتشار می شود:

$$\delta_o^4 = -\frac{\partial E}{\partial y_o^4} = \left[-\frac{\partial E}{\partial e} \frac{\partial e}{\partial y_o^4} \right] = \left[-\frac{\partial E}{\partial e} \frac{\partial e}{\partial \theta_o} \frac{\partial \theta_o}{\partial y_o^4} \right] \quad (6-5)$$

و وزنهایش به اندازه معادله زیر بروز می شود:

$$\Delta w_{ko}^4 = -\eta_w \frac{\partial E}{\partial w_{ko}^4} = \left[-\eta_w \frac{\partial E}{\partial y_o^4} \right] \left(\frac{\partial y_o^4}{\partial net_o^4} \frac{\partial net_o^4}{\partial w_{ko}^4} \right) = \eta_w \delta_o^4 x_k^4 \quad (7-5)$$

وزنهای لایه خروجی مطابق معادله زیر بروز می شوند:

$$w_{ko}^4(N+1) = w_{ko}^4(N) + \Delta w_{ko}^4 \quad (8-5)$$

ضریب η_w ضریب یادگیری وزن می باشد، و N تعداد تکرارها را نشان می دهد. در حالی که وزنهای لایه موجک واحد می باشند، تنها عبارت خطا نیاز دارد بصورت زیر محاسبه شود و منتشر شود:

$$\delta_k^3 = -\frac{\partial E}{\partial net_k^3} = \left[-\frac{\partial E}{\partial y_o^4} \right] \left(\frac{\partial y_o^4}{\partial net_o^4} \frac{\partial net_o^4}{\partial y_k^3} \frac{\partial y_k^3}{\partial net_k^3} \right) = \delta_o^4 w_{ko}^4 \quad (9-5)$$

علاوه بر آن، عملیات تکثیر در لایه موجک مادر انجام می شود. در ادامه عبارت خطا مطابق زیر محاسبه می شود:

$$\delta_j^2 = -\frac{\partial E}{\partial net_j^2} = \left[-\frac{\partial E}{\partial y_o^4} \frac{\partial y_o^4}{\partial net_o^4} \frac{\partial net_o^4}{\partial y_k^3} \frac{\partial y_k^3}{\partial net_k^3} \right] \left[\frac{\partial net_k^3}{\partial y_j^2} \frac{\partial y_j^2}{\partial net_j^2} \right] = \sum_k \delta_k^3 y_k^3 \quad (10-5)$$

و m_{ij} به اندازه زیر بروز می شود:

$$\Delta m_{ij} = -\eta_m \frac{\partial E}{\partial m_{ij}} = \left[-\eta_m \frac{\partial E}{\partial y_j^2} \frac{\partial y_j^2}{\partial net_j^2} \frac{\partial net_j^2}{\partial m_{ij}} \right] = -\eta_m \frac{\delta_j^2}{\sigma_{ij}} \quad (11-5)$$

و σ_{ij} به اندازه زیر بروز می شود:

$$\Delta\sigma_{ij} = -\eta_{\sigma} \frac{\partial E}{\partial \sigma_{ij}} = \left[-\eta_{\sigma} \frac{\partial E}{\partial y_j^2} \frac{\partial y_j^2}{\partial net_j^2} \frac{\partial net_j^2}{\partial \sigma_{ij}} \right] = \eta_{\sigma} \delta_j^2 \frac{(m_{ij} - x_i^2)}{(\sigma_{ij})^2} \quad (12-5)$$

انتقال و مقیاس موجک مادر بصورت زیر بروز می شود:

$$m_{ij}(N+1) = m_{ij}(N) + \Delta m_{ij} \quad (13-5)$$

$$\sigma_{ij}(N+1) = \sigma_{ij}(N) + \Delta \sigma_{ij}$$

در اینجا η_m و η_{σ} ضرایب یادگیری انتقال و مقیاس موجک مادر می باشد.

یک تقریب برای خطای لایه خروجی بصورت زیر در نظر می گیریم [35]:

$$\delta_o^4 \cong e + e(1 - z^{-1}) \quad (14-5)$$

۲-۱-۲-۵ آنالیز همگرایی

انتخاب مقادیر برای ضرایب یادگیری اثر زیادی روی کارایی شبکه دارد. به منظور آموزش شبکه ویونت بصورت موثر، ضرایب یادگیری تطبیقی در این قسمت استخراج شده است که همگرایی خطای ردیابی را تضمین می کند و مبتنی بر آنالیز تابع لیپانوف نوع گسسته می باشد.

تئوری ۱. فرض کنید η_w ضریب یادگیری وزنه های WNN باشد و $P_{w \max}$ بصورت

$P_{w \max} \equiv \max_N \|P_w(N)\|$ تعریف می شود در حالی که $P_w(N) = \partial y_o^4 / \partial w_{ko}^4$ و $\|\cdot\|$ نرم اقلیدسی در

$\mathbb{R}^n$ است. اگر η_w بصورت $\eta_w = \lambda / P_{w \max}^2 = \lambda / R_u$ انتخاب شود؛ همگرایی تضمین می شود. λ یک بهره ثابت

مثبت است؛ R_u تعداد گره ها در لایه موجک WNN است. که ما در اینجا بخاطر تمرکز روی موضوع اصلی از

اثبات آن خودداری می کنیم [51].

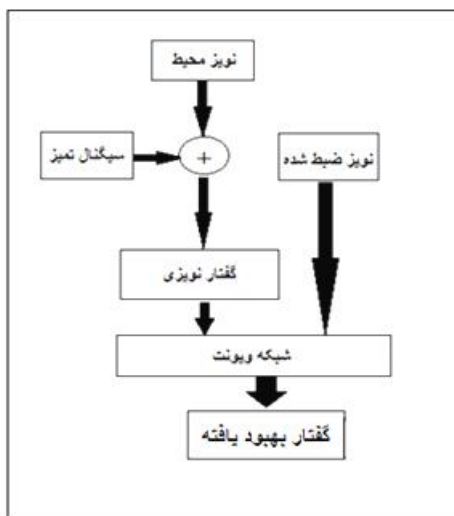
۳-۵ منبع نویز

در این بخش برای منبع نویز دو فرض وجود دارد:

نویز با دسترسی مستقیم: در محیط هایی که نویز وجود دارد و گوینده در حال صحبت می باشد؛ با ضبط نویز

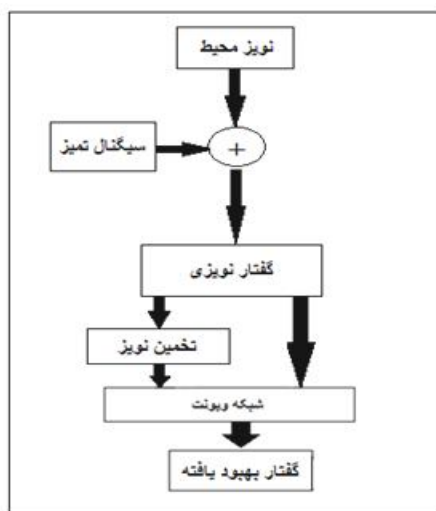
پس زمینه در لحظات سکوت گوینده، می توانیم نویز محیط را بدست آوریم. برای مثال در اتاقی که گوینده در

حال صحبت می‌باشد و همزمان صدای کولر یا فن رایانه به صورت نویز زمینه وجود دارد؛ با سکوت در این اتاق و ضبط نویز زمینه، به منبع نویز دست خواهیم یافت. شکل ۳-۵ بلوک دیاگرام حذف نویز با استفاده از نویز ضبط شده را نشان می‌دهد.



شکل ۳-۵: بلوک دیاگرام حذف نویز با استفاده از نویز ضبط شده

نویز بدون دسترس مستقیم: در مواردی که به عنوان مثال گفتار ضبط شده‌ی نویزی در دست داریم و هدف حذف یا کاهش نویز آن می‌باشد، دستیابی مستقیم به نویز زمینه در گفتار امکان ندارد باید با استفاده از روشهای تخمین نویز، به تخمین نویز زمینه پرداخت.



شکل ۴-۵: بلوک دیاگرام حذف نویز با استفاده از تخمین نویز

در این قسمت تخمین نویز با استفاده از تبدیل موجک صورت گرفته است. شکل ۴-۵ بلوک دیاگرام حذف نویز با استفاده از تخمین نویز نشان می‌دهد.

۵-۳-۱ تخمین نویز

در این پایان نامه یک روش تخمین نویز براساس تبدیل موجک گسسته پیشنهاد شده است. قبل از بیان این روش به معرفی تبدیل موجک گسسته می‌پردازیم.

۵-۳-۱-۱ تبدیل موجک گسسته

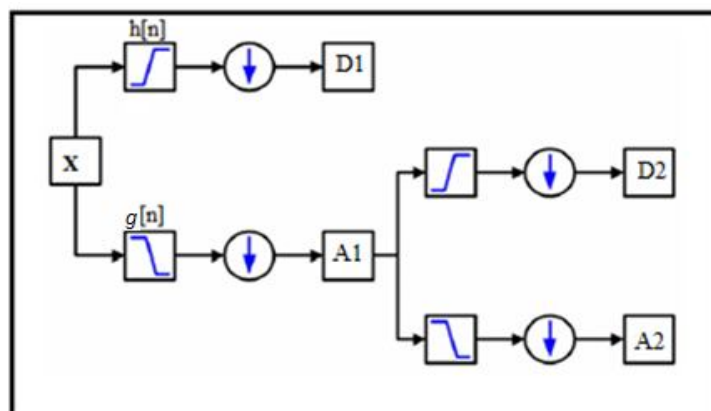
تبدیل موجک گسسته از یک سیگنال x با عبور آن از یک مجموعه از فیلترها محاسبه می‌گردد. برای این امر ابتدا سیگنال گسسته مورد نظر از یک فیلتر پایین گذر با پاسخ ضربه g عبور داده می‌شود که با استفاده از کانولوشن سیگنال و پاسخ ضربه فیلتر انجام می‌گیرد.

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{+\infty} x[k]g[n-k] \quad (۵-۱۵)$$

همچنین سیگنال به طور همزمان با استفاده از فیلتر بالاگذر با پاسخ ضربه h نیز تجزیه می‌گردد. در نتیجه این امر دو باند خروجی حاصل می‌گردد. باند اول شامل ضرایب جزئیات ناشی از عبور از فیلتر بالاگذر و باند دوم شامل ضرایب کلیات ناشی از عبور از فیلتر پایین گذر می‌باشد. برای اینکه تجزیه سیگنال به دو مجموعه ضرایب کلیات و جزئیات به طور کامل و بدون از دست دادن هرگونه اطلاعات صورت گیرد، فیلترهای بالا و پایین گذر باید از ویژگی‌های مربوط به خود برخوردار باشند. از آنجا که پس از عبور از هر فیلتر نیمی از فرکانسهای سیگنال حذف شده‌اند، پس مطابق با معیار نایکویست نیمی از نمونه‌ها می‌توانند حذف گردند. یعنی خروجی فیلترها می‌توانند با ضریب دو باز نمونه برداری گردند. روند تجزیه سیگنال به دو باند جزئیات و کلیات می‌تواند به صورت سلسله‌وار بر روی باند کلیات در هر مرحله ادامه داشته باشد. رابطه (۵-۱۶) سیگنال حاصل از تجزیه به باندهای فرکانسی مختلف را نشان می‌دهد.

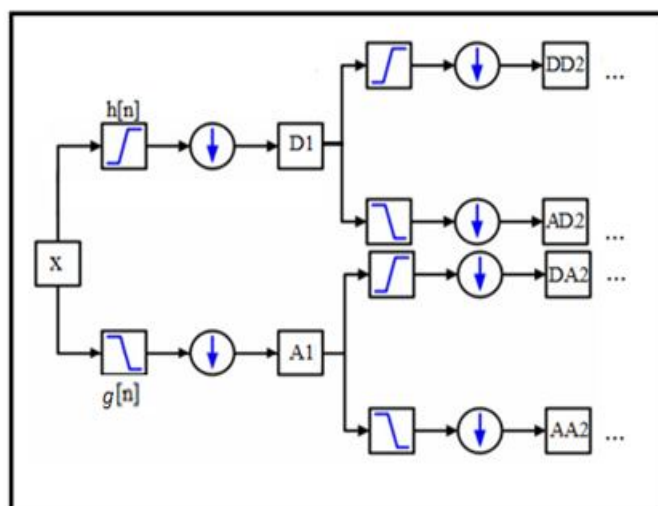
$$\hat{x} = [A_n, D_n, D_{n-1}, \dots, D_1] \quad (۵-۱۶)$$

در این معادله D جزئیات و A تقریب و n آخرین مرحله‌ی تجزیه $\hat{x}$ حاصل تبدیل موجک گسسته از x می‌باشد. در شکل ۵-۵ روند تجزیه سیگنال در تبدیل موجک نشان داده شده است.



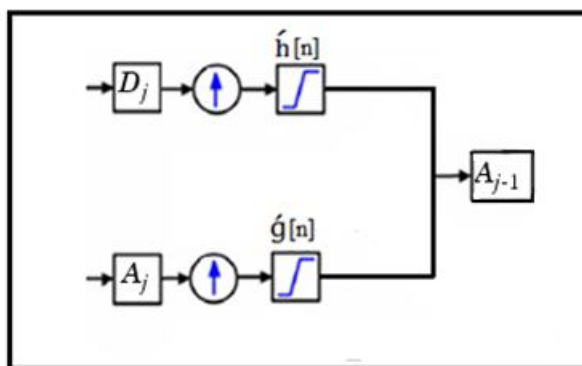
شکل ۵-۵: شمای کلی تجزیه سیگنال در تبدیل موجک گسسته

در بعضی کاربردها علاوه بر تجزیه‌ی متوالی سیگنال تقریب، سیگنال جزئیات نیز تجزیه می‌شود. در این حالت تبدیل بسته‌ی موجک را خواهیم داشت. شکل ۵-۶ شمای کلی تجزیه سیگنال در تبدیل بسته موجک را نشان می‌دهد.



شکل ۵-۶: شمای کلی تجزیه سیگنال در تبدیل بسته موجک گسسته

در ادامه به کمک موجک می‌توان سیگنال اولیه را بازسازی کرد. ابتدا تقریب و جزئیات با ضریب ۲ افزایشی نمونه برداری شده و برای بازسازی از دوفیلتر دوگان $\hat{h}$ و $\hat{g}$ عبور داده می‌شوند. شکل ۷-۵ شمای کلی بازسازی در تبدیل موجک گسسته را نشان می‌دهد [۵۰].

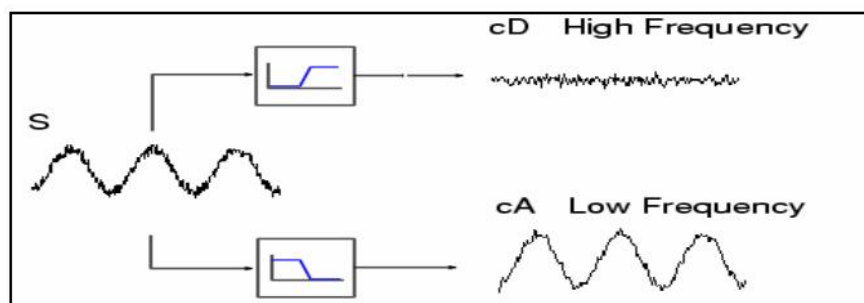


شکل ۷-۵: شمای کلی بازسازی در تبدیل موجک گسسته

۲-۱-۳-۵ تخمین نویز با استفاده از تبدیل موجک گسسته

روشهای فیلترینگ مختلفی بر اساس تبدیل موجک گسسته بنا نهاده شده اند که هریک از آنها با توجه به کاربرد موردنظر می‌تواند مورد استفاده قرار گیرد. تبدیل موجک یک تبدیل زمان-فرکانس می‌باشد. یعنی پس از انتقال سیگنال توسط تبدیل موجک به طور همزمان می‌توان به ویژگیهای فرکانسی و زمانی آن دسترسی داشت. برای تخمین نویز سیگنال بر اساس تبدیل موجک، قبل از هر چیز باید سیگنال نویزی موردنظر به فضای موجک منتقل گردد. همانطور که گفته شد در تبدیل موجک گسسته این امر به وسیله تجزیه سیگنال به باندهای فرکانسی مختلف صورت می‌گیرد. برای این امر در هر مرحله سیگنال به دو زیر باند فرکانسی مختلف تجزیه می‌گردد. نیم‌باند فرکانسی پایین حاوی ضرایب تقریبی می‌باشد و مشخصات کلی سیگنال را مشخص می‌کند. نیم باند فرکانس بالا حاوی ضرایب جزئیات می‌باشد و جزئیات ریز سیگنال را مشخص می‌کند. پس از این مرحله دوباره باند حاوی ضرایب فرکانس بالا تحت تبدیل موجک قرار گرفته و خود به دو باند فرکانسی تجزیه می‌گردد. این کار به طور متناوب صورت می‌گیرد تا رزولوشن فرکانسی موردنظر حاصل گردد. پس از انتقال سیگنال به فضای موجک، چندین مجموعه ضرایب بدست می‌آید که هر مجموعه مربوط به یک محدوده

فرکانسی خاص می‌باشند. با در نظر گرفتن حد آستانه بر روی این ضرایب و صفر کردن بعضی از آنها می‌توان نویز را تخمین زد [۴۹]. شکل ۸-۵ روند فیلتر نمودن سیگنال با استفاده از تبدیل موجک را نشان می‌دهد.



شکل ۸-۵: عملیات فیلتر نمودن سیگنال با استفاده از تبدیل موجک

۳-۱-۳-۵ آستانه گذاری روی ضرایب موجک

همانطور که در بحث تبدیل موجک اشاره شد، انرژی سیگنال غالباً در تعداد کمی از ضرایب تبدیل موجک متمرکز می‌باشند. این ضرایب نسبت به ضرایب تبدیل موجک نویز، بزرگتر هستند. بنابراین می‌توان با صفر کردن ضرایب بزرگتر و اعمال عکس تبدیل موجک، سیگنال نویز را تخمین زد.

۲-۳-۵ انتخاب آستانه

روش‌های مختلفی را می‌توان برای انتخاب حد آستانه در نظر گرفت، اما در اغلب کاربردها روش پیشنهادی Donoho که در سال ۱۹۹۵ معرفی گردید مورد استفاده قرار می‌گیرد که به فرم رابطه (۱۷-۵) محاسبه می‌شود [۴۰].

$$T = \hat{\sigma} \sqrt{2 \log(N)} \quad (۱۷-۵)$$

که N طول سیگنال نویزی و $\hat{\sigma}$ تخمینی از انحراف معیار کلی نویز است.

در مورد نویز سفید گوسی می‌توان انحراف معیار $\hat{\sigma}$ را با توجه به رابطه‌ی (۱۸-۵) بدست آورد. این رابطه در

سال ۱۹۹۴ توسط Donoho و Johnston معرفی گردید [۴۱].

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{Median(|c|)}{0.6745} \quad (۱۸-۵)$$

در این رابطه c ضریب تبدیل موجک نویز و منظور از MAD ، میانه می باشد.

اگر از تبدیل بسته موجک برای بدست آوردن ضرایب موجک استفاده کرده باشیم، حد آستانه از رابطه-
ی (۱۹-۵) محاسبه می شود.

$$T = \hat{\sigma} \sqrt{2 \log(N \log_2 N)} \quad (۱۹-۵)$$

در سال ۱۹۹۷، Silverman روش آستانه بندی پیشنهاد شده توسط Donoho را توسعه داده و حد آستانه را
با توجه به رابطه ی (۲۰-۵) به طور مجزا برای هر سطح در تبدیل بسته موجک بیان کرد.

$$T_j = \hat{\sigma}_j \sqrt{2 \log(N_j)} \quad (۲۰-۵)$$

که N_j تعداد نمونه ها در سطح j ام و $\hat{\sigma}_j$ انحراف معیار نویز در سطح j ام می باشد. انحراف معیار نویز در این
حالت برای هر سطح، از رابطه ی (۲۱-۵) تخمین زده می شود.

$$\hat{\sigma}_j = \frac{MAD_j}{0.6745} \quad (۲۱-۵)$$

در این رابطه MAD_j در هر سطح محاسبه می شود [۳۴].

نتایج شبیه سازی و پیشنهادات

۶ نتایج شبیه سازی

۱-۶ مقدمه

در این فصل مراحل پیاده سازی روش پیشنهادی به تفصیل بیان خواهد شد. نتایج حاصل از آزمایشات عملکرد مطلوب روش پیشنهادی را نشان می‌دهند.

۲-۶ معرفی پایگاه داده

در این پایان نامه از دادگان فارسی استفاده شده است. این پایگاه داده مجموعه ای از ۶۰۰۰ جمله می‌باشد که توسط ۳۰۰ گویشور فارسی زبان تولید شده اند. گویشوران بر حسب سن، جنس، لهجه و میزان تحصیلات انتخاب گردیده و هر گویشور ۲۰ جمله در اطاقک ضد صدای آزمایشگاه زبان شناسی دانشگاه تهران بیان نموده‌اند. فرکانس نمونه برداری این جملات ۲۲۰۵۰ هرتز می‌باشد.

۳-۶ فرضیات

قبل از هر چیز لازم است فرضیات اساسی به کار رفته در روش پیشنهادی مطرح شود. در روش پیشنهادی فرض می‌شود که مولفه های سگینال تمیز مستقل از همدیگر هستند. فرض دیگر این است که نویز از سیگنال تمیز مستقل می‌باشد و همچنین نویز به صورت سیگنالی جمع شونده بر سیگنال گفتار سوار شده و گفتار آلوده‌ی نویزی را بوجود آورده است.

سیگنال گفتار تمیز $s(n)$ در اثر جمع شدن با نویز $n(n)$ سیگنال نویزی $y(n)$ را طبق رابطه (۱-۶) تولید کرده است.

$$y(i) = s(i) + n(i) \quad (۱-۶)$$

روش بهسازی گفتار معرفی شده در این پایان نامه شامل چندین مرحله می‌باشد که در این قسمت مراحل به

همراه نتایج آن بیان خواهد شد.

۴-۶ مراحل بهسازی گفتار در روش پیشنهادی

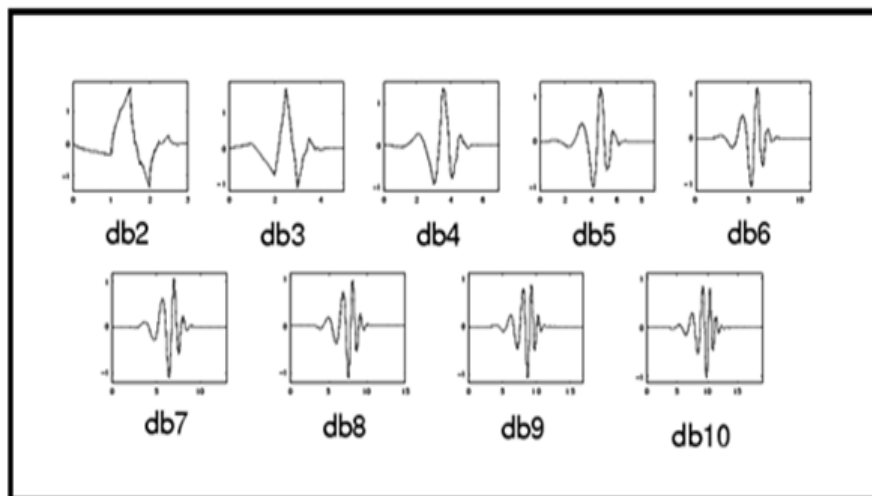
۱-۴-۶ نویز:

همانطور که در بلوک دیاگرام شکل ۵-۱ در فصل قبل نشان داده شد؛ اولین قدم دست یافتن به نویز سوار شده بر سیگنال نویزی می‌باشد. در این مرحله دو روش زیر بیان شد.

الف: نویز با دسترسی مستقیم

ب: نویز بدون دسترسی مستقیم

در حالت الف که نیاز به معرفی الگوریتم نخواهد بود و دست یافتن به نویز به روشی که قبلاً بیان شد؛ کار ساده‌ای خواهد بود. در حالت ب برای تخمین نویز از تبدیل موجک گسسته استفاده کرده ایم. در اینجا گفتار نویزی y به کمک موجک دابیشز ۵ تجزیه شده است. شکل زیر نمونه هایی از موجک دابیشز را نشان می‌دهد.



شکل ۱-۶: مثالی از موجکهای دابیشز

برای انتخاب نوع دابیشز، روش پیشنهادی را بر روی گفتار آلوده به نویز سفید گوسی با طول ۱۱۱۰۰۰ نمونه

آزمایش کرده و به ازای n های مختلف سیگنال به نویز خروجی را محاسبه کرده ایم. نتایج در جدول ۶-۱ نشان داده شده است.

جدول ۶-۱: SNR سیگنال خروجی به ازای n های مختلف $db(n)$					
Input SNRs(dB) ↓ $db(n): n \rightarrow$	Output SNRs(dB) ↓				
	۱	۳	۵	۷	۹
-۵	۱.۹	۲.۵۹	۲.۶۳	۲.۶	۲.۶۱
-۱	۳	۴	۴.۹	۴.۷	۴.۷
+۴	۶.۵	۷.۸	۸.۰۷	۸.۰۱	۸.۰۴

همانطور که در جدول مشاهده می شود، افزایش سیگنال به نویز خروجی برای دابیشز ۵ ($n=5$) نسبت به سایر انواع موجک بیشتر است. برای انتخاب تعداد مرحله ی تجزیه (j)، بر روی سیگنال نویزی روش پیشنهادی را آزمایش کرده و به ازای j های مختلف سیگنال به نویز خروجی را محاسبه کرده ایم. نتایج بدست آمده در جدول ۶-۲ آمده است.

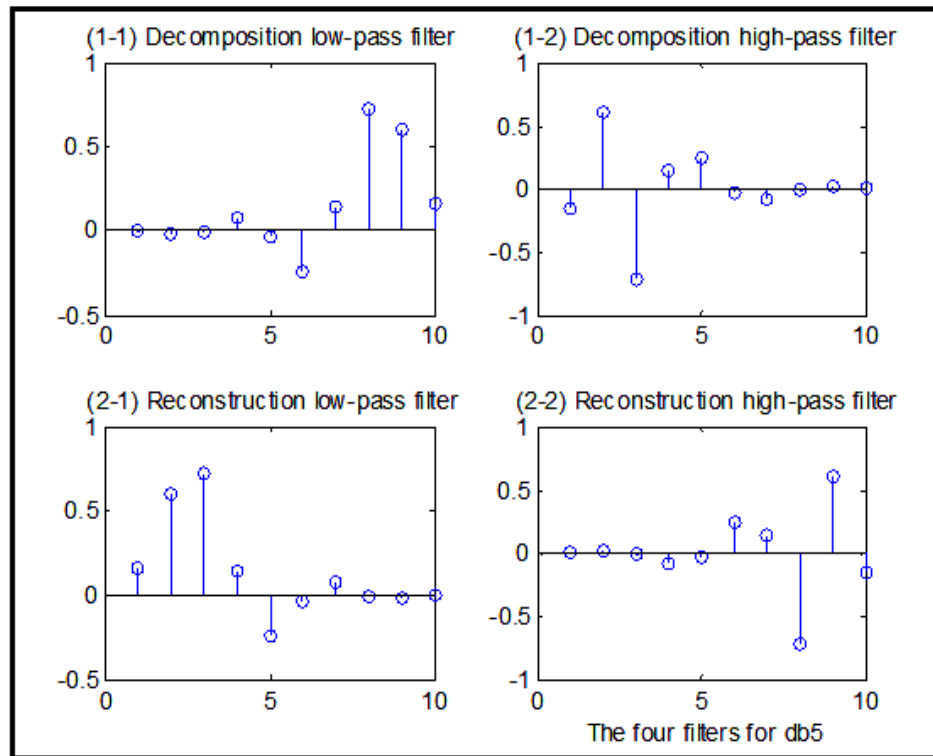
جدول ۶-۲: SNR سیگنال خروجی به ازای تعداد مراحل تجزیه (j)					
Input SNRs(dB) ↓ $j \rightarrow$	Output SNRs(dB) ↓				
	۱	۳	۵	۷	۹
-۵	۰.۲۷	۱.۵	۲.۵	۲.۶	۲.۶۱
-۱	۱.۰۱	۳.۷	۴.۷	۴.۹۵	۴.۸
+۴	۳.۱۹	۷	۷.۹	۸.۰۹	۸.۱۴

همانطور که مشاهده می شود، به ازای $j=9$ بیشترین افزایش سیگنال به نویز را خواهیم داشت. افزایش مرحله ی تجزیه از حدی بیشتر در نتیجه ی آزمایش بی تاثیر می باشد. پس موجک انتخابی ما در این پایان نامه دابیشز ۵

(db5) و تعداد مرحله‌ی تجزیه (j) ۹ می‌باشد. حاصل تبدیل موجک سیگنال y در رابطه (۲-۶) آمده است.

$$\hat{y} = [A_9, D_9, D_8, D_7, D_6, D_5, D_4, D_3, D_2, D_1] \quad (2-6)$$

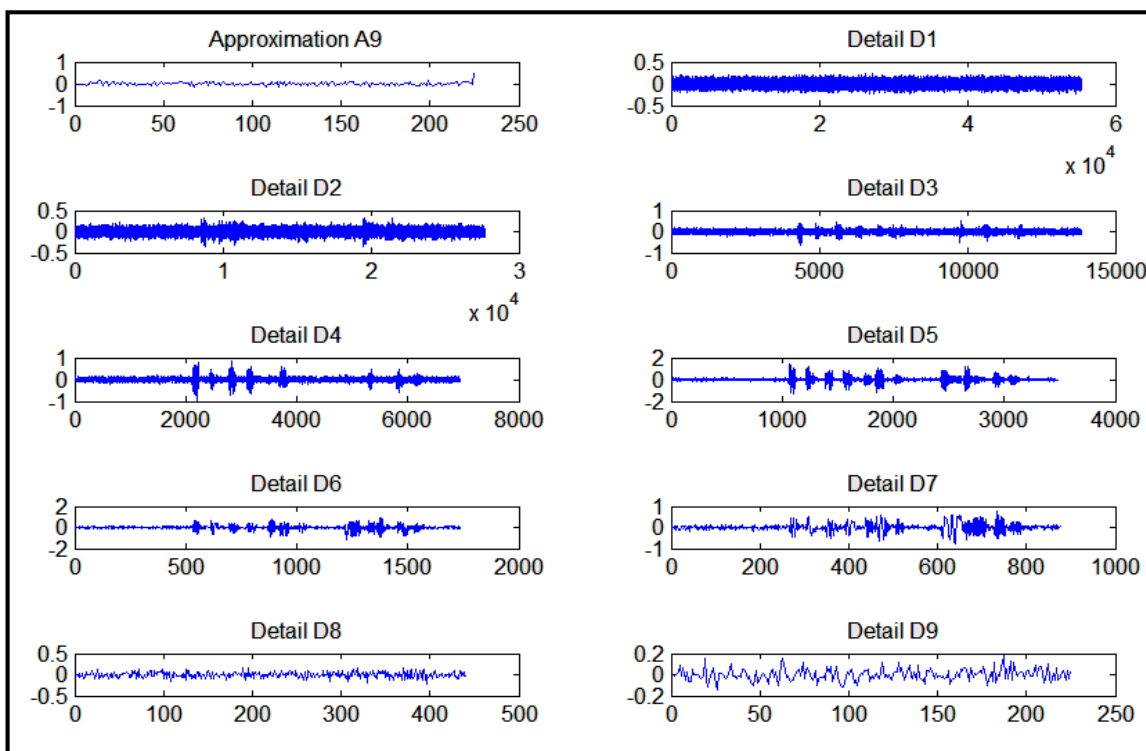
در شکل ۲-۶: فیلترهای بالاگذر و پایین گذر و دوگان آنها که به ازای موجک دابیشز بدست آمده اند، مشاهده می‌شود. این فیلترها با استفاده از جعبه ابزار نرم افزار matlab بدست آمده‌اند.



شکل ۲-۶: فیلترهای بالاگذر و پایین گذر و دوگان آنها

در این شکل بخش (۱-۱) ضرایب فیلتر پایین گذر (g) و بخش (۱-۲) ضرایب فیلتر بالا گذر (h) در مرحله تجزیه می‌باشد. همچنین بخش (۲-۱) فیلتر پایین گذر ($\tilde{g}$) و بخش (۲-۲) ضرایب فیلتر بالا گذر ($\tilde{h}$) در مرحله بازسازی را نشان می‌دهد. شکل ۳-۶: تجزیه این گفتار نویزی را نشان می‌دهد.

در این شکل تجزیه به کمک موجک دابیشز تا ۹ مرحله صورت گرفته است. همانطور که قبلاً بیان شد، منظور از D_j جزئیات است که خروجی حاصل از فیلتر بالاگذر و A_j خروجی حاصل از فیلتر بالاگذر مرحله j می‌باشد.



شکل ۳-۶: مرحله تجزیه گفتار نویزی به کمک موجک دابیشز

انتخاب آستانه: ۱-۱-۴-۶

در این قسمت با استفاده از حد آستانه‌ی پیشنهادی Donoho، با توجه به رابطه‌ی (۵-۱۹) آستانه را بدست آورده و آنرا بر روی ضرایب $\hat{y}$ اعمال کرده‌ایم. به این ترتیب که ضرایب بزرگتر از حد آستانه را صفر در نظر گرفته‌ایم. اعمال این حد آستانه در روابط زیر آورده شده است. با این فرض که c در معادله (۵-۱۸) را ضریب تبدیل موجک سیگنال نویزی در نظر گرفته‌ایم.

$$\text{THR}(\hat{y}, T) = \begin{cases} \hat{y}, & |\hat{y}| < T \\ 0, & |\hat{y}| \geq T \end{cases} \quad (۳-۶)$$

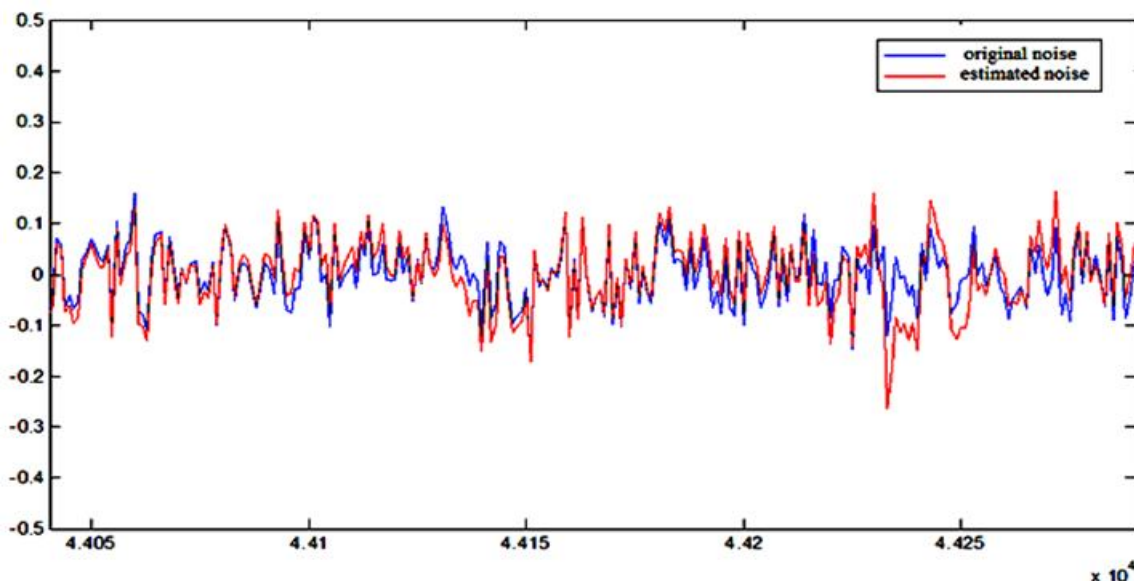
$$\hat{n} = \text{THR}(\hat{y}, T) = \begin{cases} \hat{y}, & |\hat{y}| < T \\ 0, & |\hat{y}| \geq T \end{cases} \quad (۴-۶)$$

که THR آستانه گذاری را انجام می‌دهد. $\hat{n}$ ضریب نویز تخمین زده شده در حوزه‌ی موجک می‌باشد و در نهایت از سیگنال باقی‌مانده عکس تبدیل موجک گرفته خواهد شد. بدین ترتیب نویز سوار بر سیگنال را تخمین زده‌ایم. خلاصه‌ی روند طی شده در شکل ۴-۶ مشاهده می‌شود.



شکل ۴-۶: خلاصه روند طی شده در تخمین نویز روش پیشنهادی

در شکل زیر قسمتی از نویز سفید گوسی و نویز تخمینی با روش بیان شده، نشان داده شده است.



شکل ۵-۶: مقایسه نویز اصلی و نویز تخمین زده شده

همانطور که مشاهده می‌گردد، نویز تخمین زده شده شباهت زیادی با نویز واقعی خواهد داشت و این نشان

می‌دهد که روش تخمین نویز استفاده شده در این قسمت، روشی مطلوب می‌باشد.

دومین قدم، اعمال نویز تخمینی و سیگنال گفتار آلوده نویزی به شبکه‌ی ویونت می‌باشد. این شبکه وزنه‌ای

خود را طوری تطبیق می‌دهد که بتواند نویز را به خوبی مدل کرده و آنرا از سیگنال آلوده حذف کند. ورودی

شبکه، نویز تخمینی می‌باشد و گفتار نویزی، به عنوان تارگت انتخاب می‌شود.

همانطور که بیان شد، شبکه ویونت (WNN) پیشنهادی ۴ لایه تشکیل شده است، که شامل یک لایه ورودی

(لایه i)، لایه موجک مادر (لایه j)، یک لایه موجک (لایه k) و لایه خروجی (لایه o) می‌باشد. نکته‌ی مهم در

طراحی این شبکه، نوع تابع موجک، تعداد نرونها در هر لایه و تعداد تکرارهای الگوریتم شبکه می‌باشد. در

بخش بعد، با توجه به نتایج آزمایش حاصل از تغییر پارامترهای شبکه‌ی ویونت، مطلوب ترین شبکه را انتخاب

کرده‌ایم.

۵-۶ نتایج آموزش شبکه ویونت با توابع موجک مختلف

در این بخش از توابع فعالساز مختلف برای آموزش شبکه ویونت استفاده می‌کنیم و مشاهده می‌کنیم چه تغییری در سرعت آموزش و کاهش اندازه تابع انرژی (خطای پارامترها) بوجود می‌آید. نتایج در جدول ۳-۶ آمده است.

جدول ۳-۶: نتایج MSE شبکه ویونت با توابع موجک مختلف در ۵ مرتبه تکرار الگوریتم آموزش

Name	f (x)	Sum Squared error in iteration5
Morlet	$f(x) = \cos(x) \exp\left(-\frac{x^2}{2}\right)$	0.0029
RASP1	$f(x) = \frac{x}{(x^2 + 1)^2}$	0.0119
RASP2	$f(x) = \frac{x \cos(x)}{x^2 + 1}$	0.0091
RASP3	$f(x) = \frac{\sin(\pi x)}{x^2 - 1}$	0.0072
SLOG	$f(x) = \frac{1}{1 + e^{-x+1}} - \frac{1}{1 + e^{-x+3}} - \frac{1}{1 + e^{-x-3}} + \frac{1}{1 + e^{-x-1}}$	8.0492e-004
SLOG2	$f(x) = \frac{3}{1 + e^{-x-1}} - \frac{3}{1 + e^{-x+1}} - \frac{1}{1 + e^{-x-3}} + \frac{1}{1 + e^{-x+3}}$	0.0023
POLYWOG1	$f(x) = x \exp\left(-\frac{x^2}{2}\right)$	0.0047
POLYWOG2	$f(x) = (x^3 - 3x) \exp\left(-\frac{x^2}{2}\right)$	0.0220
POLYWOG3	$f(x) = (x^4 - 6x^2 + 3) \exp\left(-\frac{x^2}{2}\right)$	0.0026
POLYWOG4	$f(x) = (1 - x^2) \exp\left(-\frac{x^2}{2}\right)$	5. 7e-006
POLYWOG5	$f(x) = (3x^2 - x^4) \exp\left(-\frac{x^2}{2}\right)$	8.25e-005
Shannon	$f(x) = \frac{\sin(2\pi x) - \sin(\pi x)}{\pi x}$	0.0035
Gaussian	$\varphi(x) = -x \exp\left(-\frac{x^2}{2}\right)$	0.00149

جدول ۳-۶ مجموع مربعات خطا (MSE) بین سیگنال مطلوب و سیگنال خروجی را به ازای موجکهای مختلف در ۵ تکرار نشان می‌دهد. با توجه به جدول زیر می‌بینیم که در ۵ تکرار اولیه آموزش توابع SLOG1 و POLYWOG4 و POLYWOG5 خطای مجموع کمتری نسبت به بقیه توابع دارند در نتیجه پارامترهای خروجی آنها شبیه تر به خروجی مطلوب می‌باشد.

در نتیجه برای آموزش on-line بهترین تابع موجک، موجک POLYWOG4 می‌باشد.

۶-۶ نتایج آموزش شبکه ویونت با تعداد نورونهای متفاوت در لایه ها

اگر از تابع POLYWOG4 در شبکه ویونت استفاده کنیم و تعداد نورونهای لایه دوم ۶ و لایه سوم ۳ نورون باشد نتایج زیر حاصل می‌شود. در جداول زیر n تعداد گره‌ها در لایه موجک مادر و l تعداد گره‌ها در لایه موجک را نشان می‌دهد.

جدول ۴-۶ : اطلاعات شبکه ویونت با ۶ نورون در لایه دوم و ۳ نورون در لایه سوم

Threshold Error ↓	Number of Iterations↓
	POLYWOG4 , n=6, l=3
۰.۸	۷
۰.۵	۷
۰.۳	۷
۰.۱	۲۴
۰.۰۵	۴۸
۰.۰۴	۸۴
۰.۰۳	۱۰۷
۰.۰۲	۷۸

در این جداول ، تعداد تکرارهای لازم برای رسیدن به حد خطاهای ذکر شده، نشان داده شده است. نتیجه ای که از این جدول می توان گرفت این است که با توجه به تعداد نورونها در لایه های مشخص شده، چند بار باید شبکه آموزش ببیند تا به حد خطای قابل قبول برسد.

نتایج این جدول و جدولهای بعدی می تواند منجر به یک روش نظام مند شود تا بتوانیم تشخیص دهیم برای رسیدن به خطای خاصی از چند بار تکرار آموزش و چه تعداد نورون در لایه های مختلف استفاده کنیم که این خود نتیجه ای ارزشمند است. در جدول ۵-۶ تعداد تکرارهای الگوریتم آموزش تا رسیدن به حد خطای آستانه آورده شده است.

جدول ۵-۶ : اطلاعات شبکه ویونت با ۸ نورون در لایه دوم و ۴ نورون در لایه سوم

Threshold Error ↓	Number of Iterations↓
	POLYWOG4 , n=8, l=4
۰.۸	۵
۰.۵	۱۱
۰.۳	۱۳
۰.۱	۲۹
۰.۰۵	۶۳
۰.۰۴	۵۳
۰.۰۳	۴۱
۰.۰۲	۳۸

جدول ۶-۶: اطلاعات شبکه ویونت با ۲۴ نورون در لایه دوم و ۱۲ نورون در لایه سوم

Threshold Error ↓	Number of Iterations↓
	POLYWOG4 ,n=24 ,l=12
۰.۸	۷
۰.۵	۱۱
۰.۳	۱۰
۰.۱	۶۳
۰.۰۵	۱۲۸
۰.۰۴	۹۶
۰.۰۳	۱۷۰
۰.۰۲	۲۴۵

جدول ۷-۶: اطلاعات شبکه ویونت با ۴۸ نورون در لایه دوم و ۲۴ نورون در لایه سوم

Threshold Error ↓	Number of Iterations↓
	POLYWOG4 ,n=48 ,l=24
۰.۸	۸
۰.۵	۱۵
۰.۳	۴۰
۰.۱	۶۸
۰.۰۵	۱۱۴
۰.۰۴	۱۶۳
۰.۰۳	۳۴۲
۰.۰۲	۳۹۹

آوردیم به این صورت می باشد که برای رسیدن به خطای قابل قبول ۰.۸ بهترین حالت، انتخاب ۸ نورون در لایه دوم و ۴ نورون در لایه سوم می باشد چون دارای کمترین تکرار برای آموزش می باشد. و برای رسیدن به خطای قابل قبول ۰.۵، ۰.۳ و ۰.۱ انتخاب ۶ نورون در لایه دوم و ۳ نورون در لایه سوم، می تواند بهترین انتخاب برای ما باشد. و برای رسیدن به خطای ۰.۰۴ و ۰.۰۳ و ۰.۰۲ انتخاب ۸ نورون در لایه دوم و ۴ نورون در لایه سوم مناسب می باشد.

۶-۷ بررسی نمودار و منحنی های حاصل از پیاده سازی روش پیشنهادی با پارامترهای مطلوب

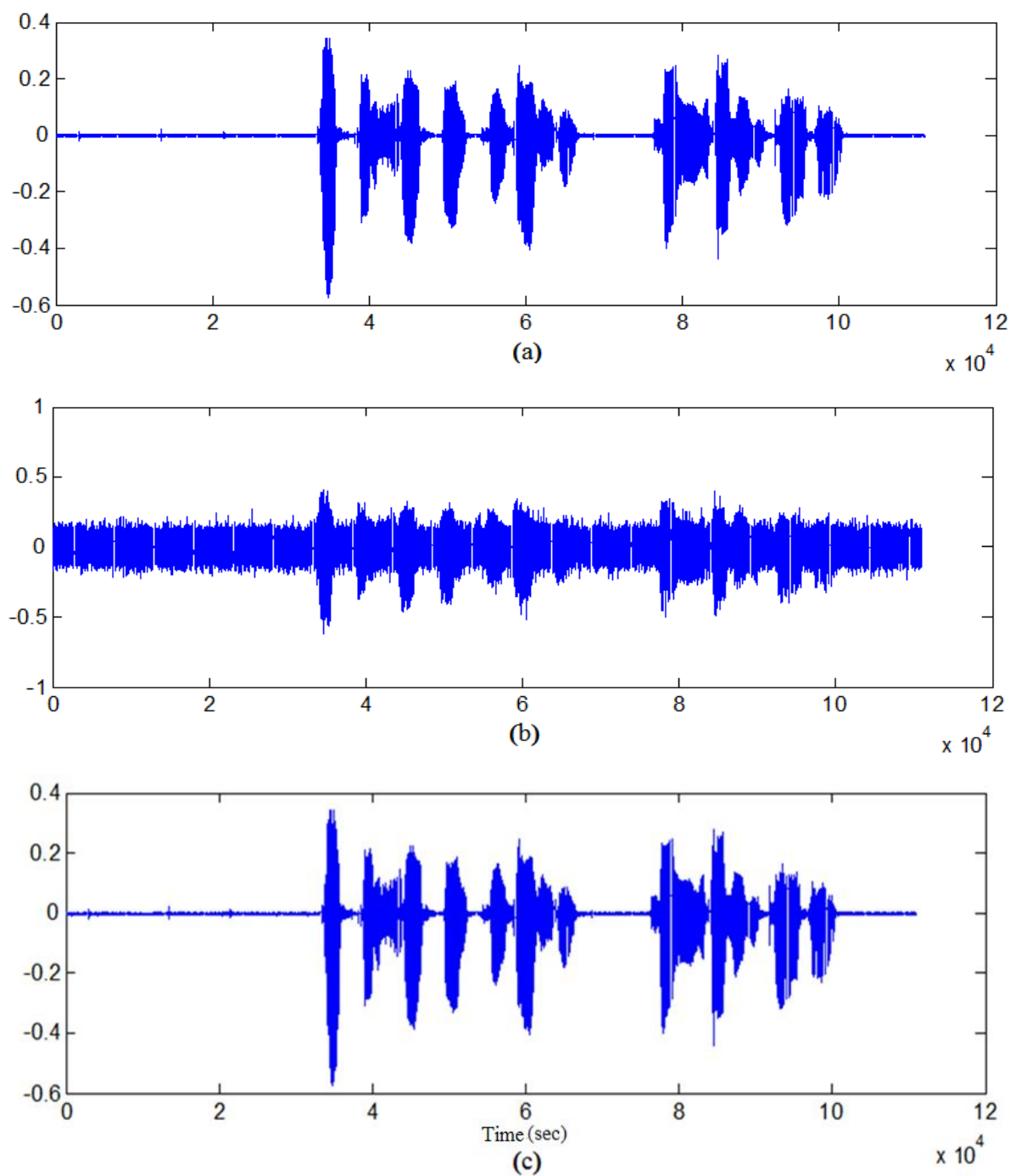
در شبکه ی پیاده شده، ۲ گره در لایه ورودی، ۸ گره در لایه ی موجک مادر، ۴ گره در لایه موجک در نظر گرفته شده است. ضریب یادگیری انتقال (η_m) اولیه برابر ۰.۰۰۴۵ و ضریب مقیاس اولیه (η_σ) برابر ۰.۰۰۱۵ فرض شده است. ضریب یادگیری وزن η_w برابر ۰.۰۱۴ فرض شده است. همانطور که بیان شد موجک انتخابی ما db5 می باشد. در شکل زیر یک جمله از دادگان فارسی انخاب شده است. این سیگنال گفتار را یکبار با نویز سفید گوسی، بار دیگر با نویز همهمه و نویز F۱۶ آلوده کرده ایم. مرحله ی بعد نتایج گرافیکی حاصل از بهسازی گفتار توسط روش پیشنهادی را به دو نحو نشان داده ایم.

اولین نحو نمایش، دامنه ی سیگنال در حوزه زمان را نشان می دهد و دومین نحو نمایش، نمایش Spectrogram سیگنال می باشد.

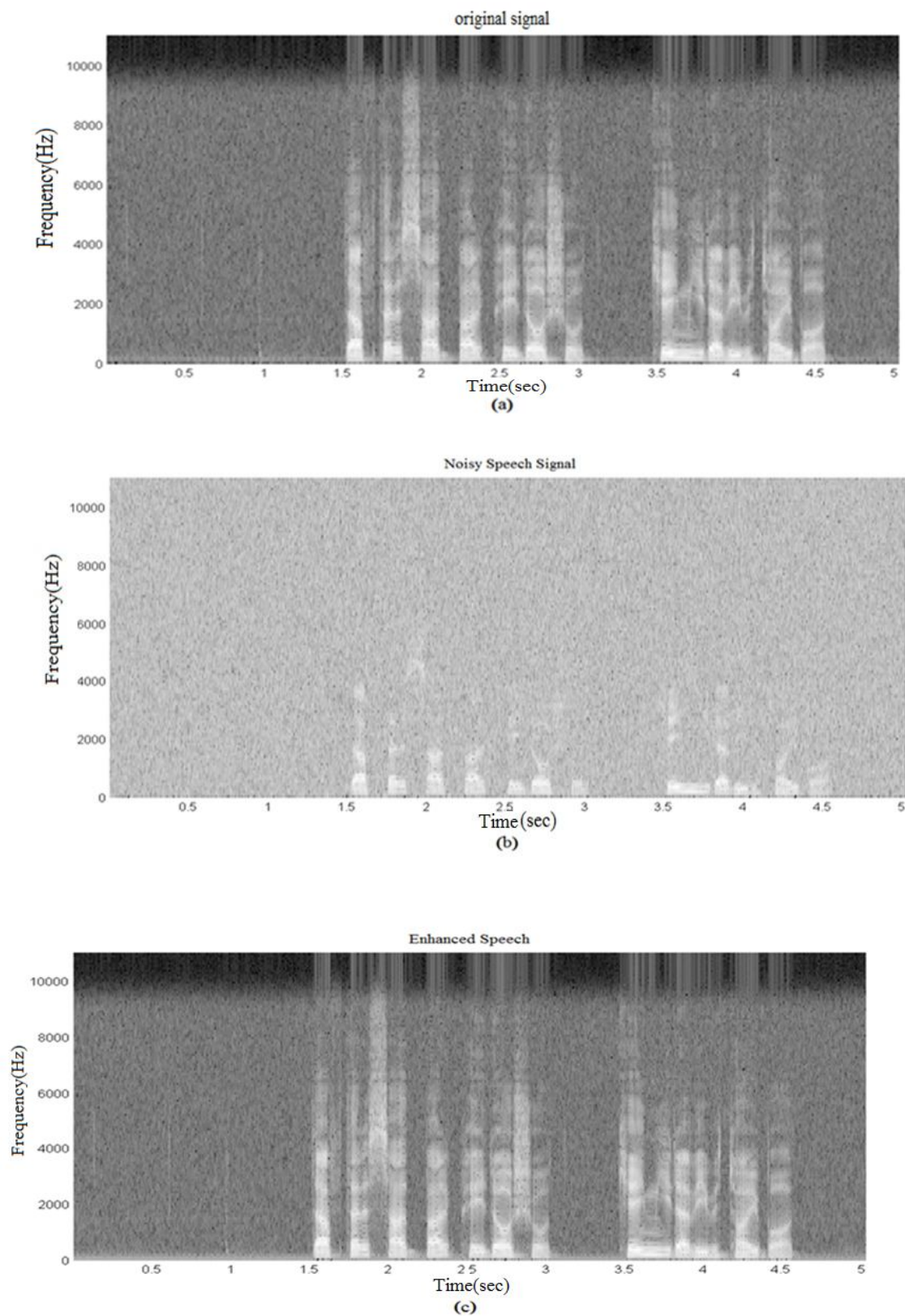
Spectrogram نمایش گرافیکیست که دوبعد هندسی دارد. محور افقی نشان دهنده زمان و محور عمودی، فرکانس است. بعد سوم نشان دهنده ی دامنه از یک فرکانس و زمان خاص است.

شکل ۶-۶ در قسمت (a) یک سیگنال گفتار بدون نویز را نشان می دهد. در قسمت (b) این سیگنال با نویز سفید گوسی آغشته شده است. قسمت (c) گفتار بهبود یافته حاصل از روش پیشنهادی را نشان می دهد. همانطور که مشاهده می شود سیگنال این قسمت با سیگنال قسمت (a) شباهت داشته و این نشان دهنده ی

کاهش نویز از سیگنال گفتار نویزی می باشد.



شکل ۶-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز سفید گوسی (b) و گفتار بهبود یافته (c)



شکل ۶-۷: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز سفید گوسی (b) و گفتار بهبود یافته (c)

در شکل ۶-۷ برای نشان دادن شباهت سیگنال گفتار تمیز و گفتار بهبود یافته، Spectrogram برای گفتار تمیز در قسمت (a) و برای گفتار بهبود یافته در قسمت (c) نمایش داده شده است.

همانطور که مشاهده می شود شباهت این دو Spectrogram نشان دهنده کارایی روش پیشنهادی در حذف نویز می باشد.

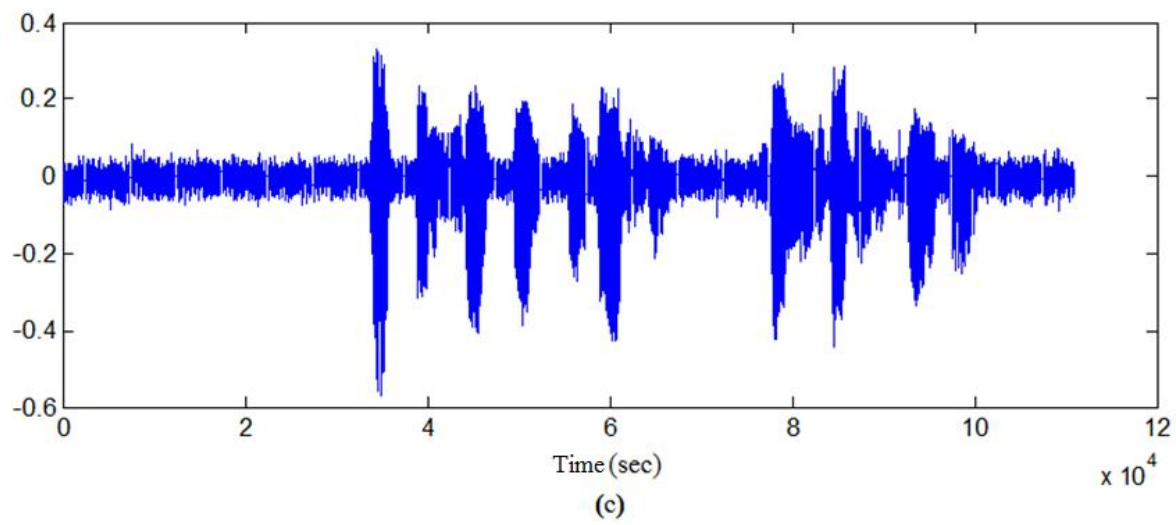
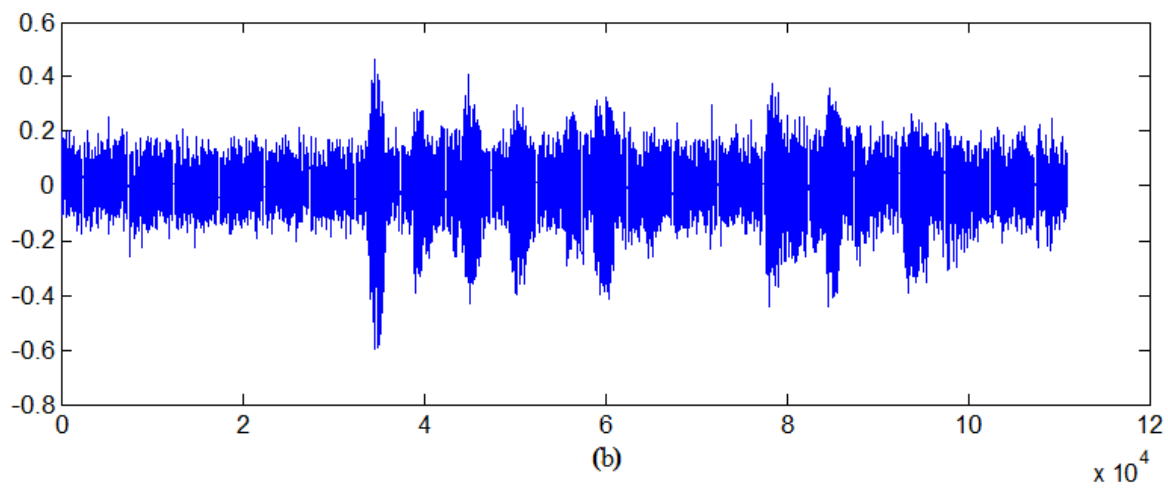
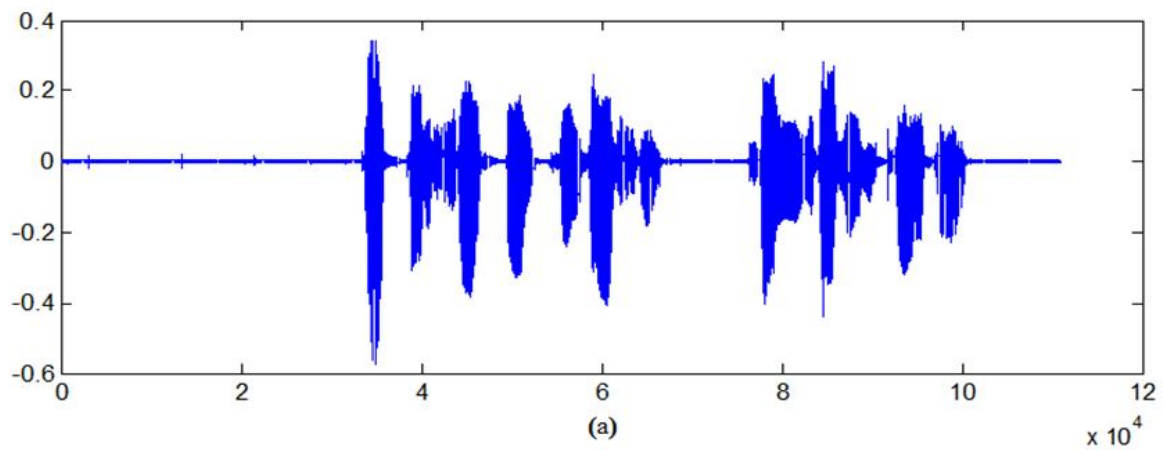
شکل ۶-۸ در قسمت (a) یک سیگنال گفتار بدون نویز را نشان می دهد.

در قسمت (b) این سیگنال با نویز F16 آغشته شده است. قسمت (c) گفتار بهبود یافته حاصل از روش پیشنهادی را نشان می دهد.

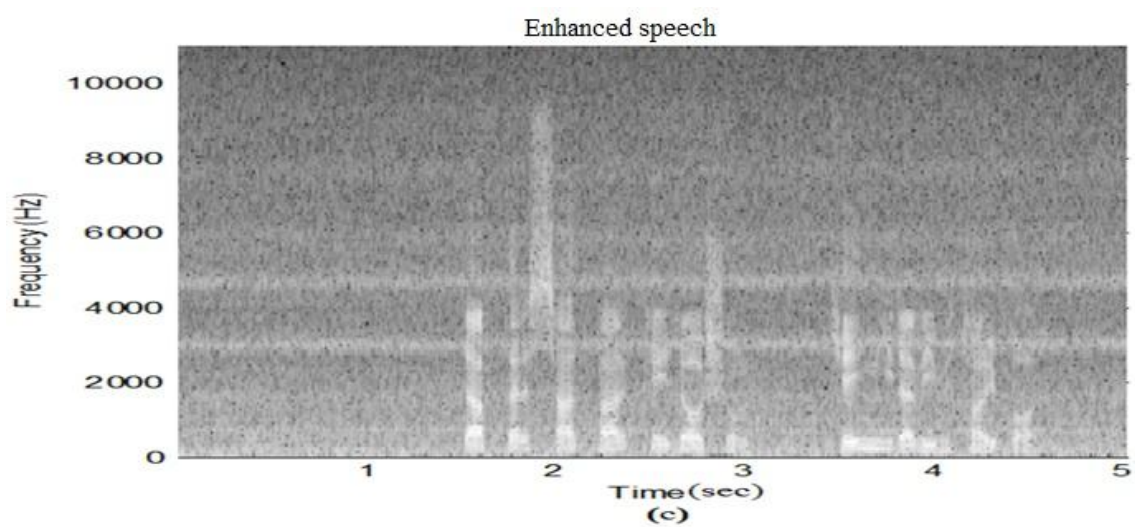
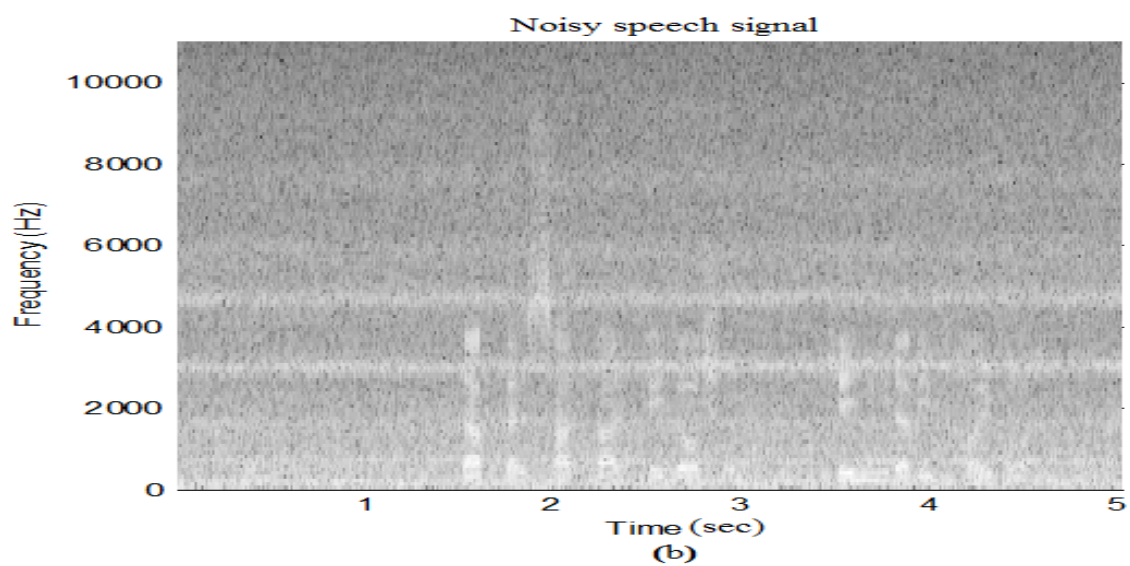
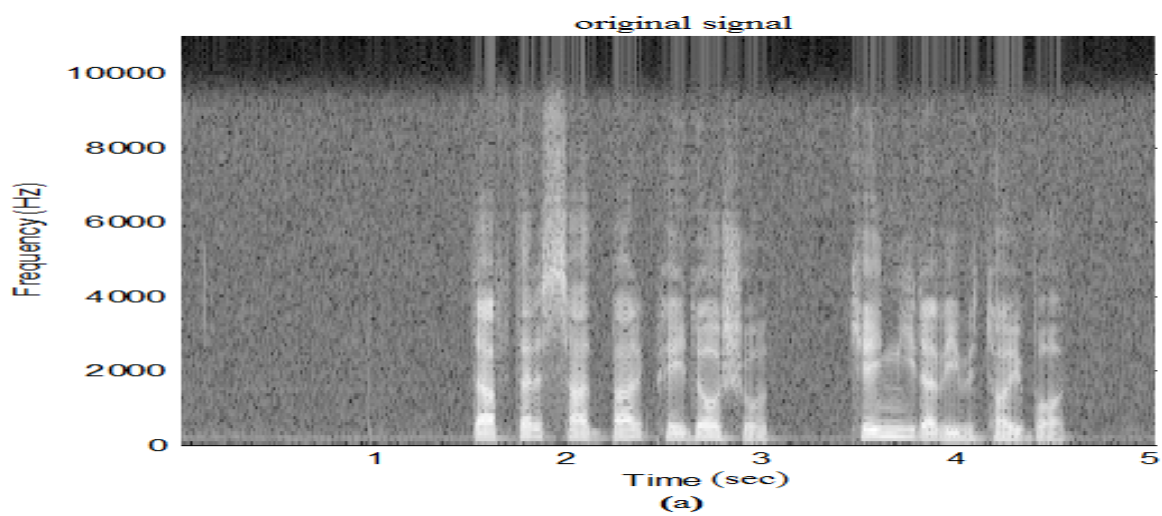
همانطور که در سیگنال (c) مشاهده می شود، روش پیشنهادی تا حد مطلوبی نویز سوار بر سیگنال گفتار را کاهش داده است.

در شکل ۶-۹، Spectrogram برای گفتار تمیز و گفتار آلوده به نویز F16 و گفتار بهبود یافته را در قسمت (a)، (b) و (c) را مشاهده خواهید کرد.

Spectrogram در گفتار بهبود یافته، با تقریب خوبی، مشابه با Spectrogram در گفتار تمیز می باشد.

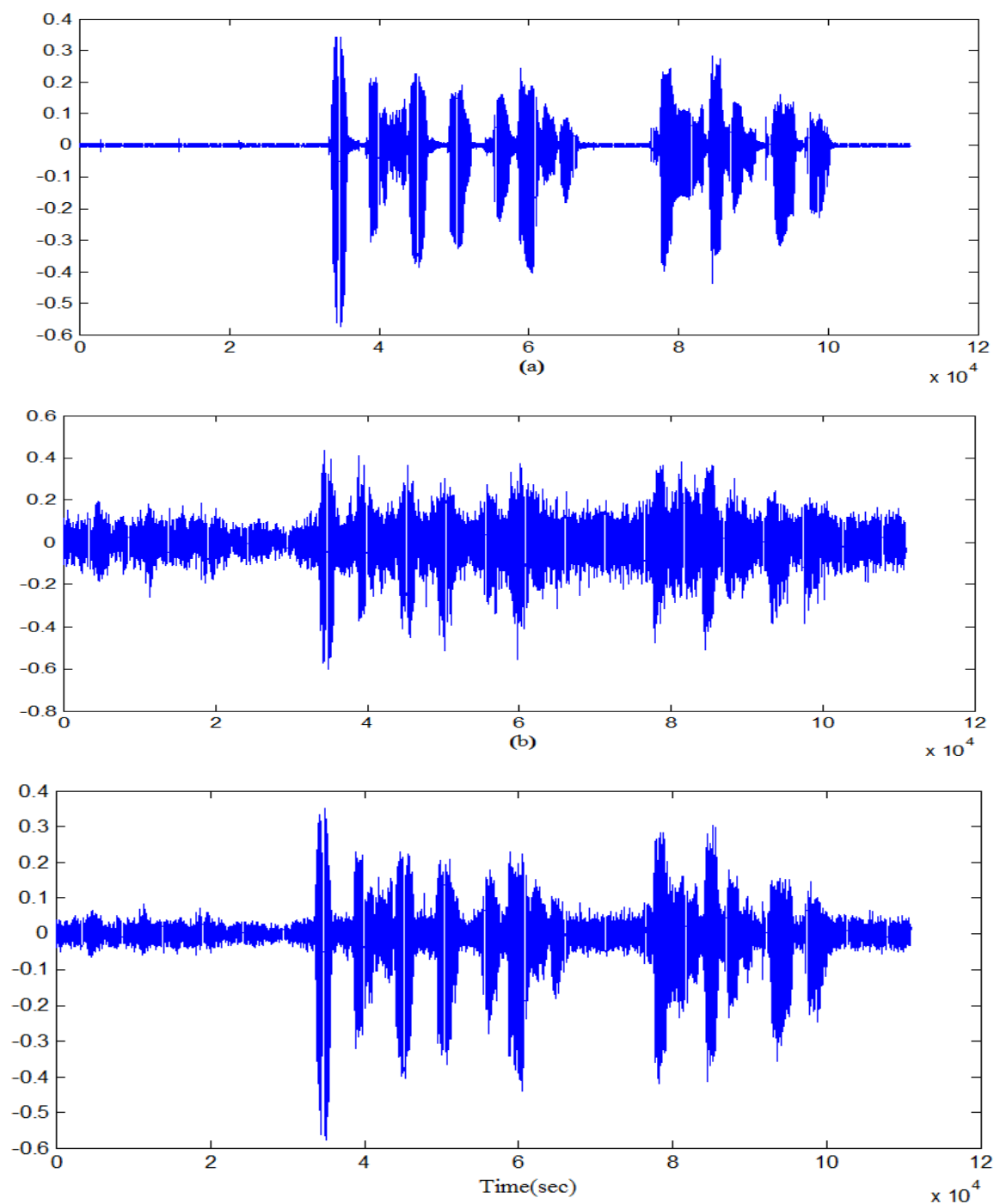


شکل ۸-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c)

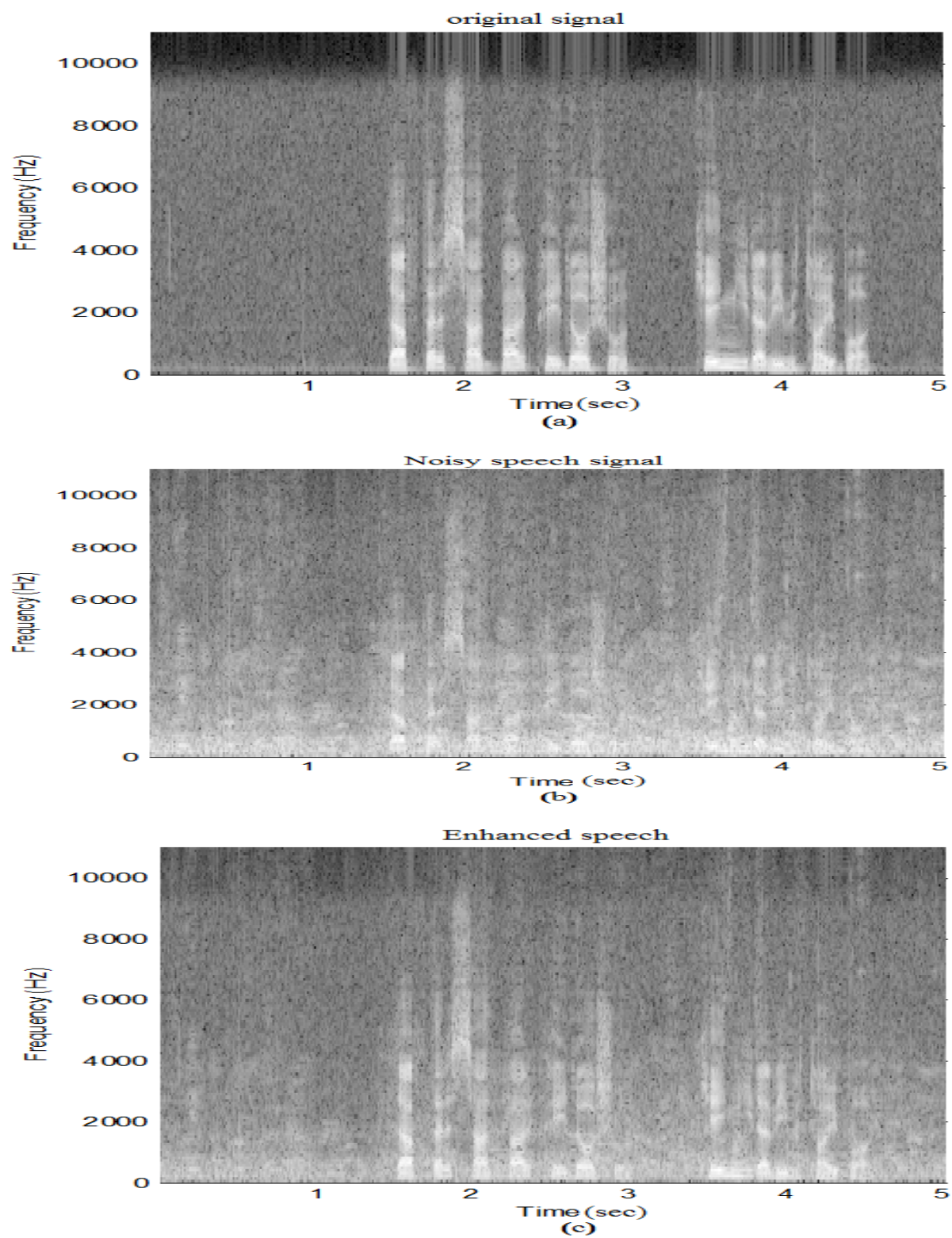


شکل ۹-۶: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c)

در شکل ۱۰-۶ سیگنال گفتار، آلوده به نویز همهمه می‌باشد. در قسمت (a) سیگنال گفتار بدون نویز نشان داده شده است. در قسمت (b) این سیگنال با نویز همهمه آغشته شده است.



شکل ۱۰-۶: نمایش سیگنال گفتار تمیز (a) و آغشته به نویز همهمه (b) و گفتار بهبود یافته (c)



شکل ۶-۱۱: نمایش Spectrogram برای گفتار تمیز (a) و آغشته به نویز F16 (b) و گفتار بهبود یافته (c)

گفتار بهبود یافته حاصل از روش پیشنهادی در قسمت (c) نشان داده شده است. همانطور که در سیگنال (c) مشاهده می‌شود، روش پیشنهادی تا حد مطلوبی نویز سوار بر سیگنال گفتار را کاهش داده است. در شکل ۱۱-۶، Spectrogram برای گفتار تمیز و گفتار آلوده به نویز همهمه و گفتار بهبود یافته را در قسمت (a) و (b) و (c) را مشاهده خواهید کرد.

۸-۶ ارزیابی روش پیشنهادی

پس از پیاده سازی روش پیشنهادی، برای بررسی و سنجش عملکرد روش، باید آن را با معیارهای مشخص ارزیابی کرد. روش پیشنهادی، ابتدا با تخمین نویز روش الف (ضبط صدای نویز در لحظات سکوت) و سپس با تخمین نویز روش ب (تخمین نویز با تبدیل موجک) پیاده سازی شده است و ما در ادامه معیارهای ارزیابی را برای هر دو حالت بررسی خواهیم کرد.

در این پایان نامه، برای ارزیابی روش پیشنهادی ده جمله از پایگاه دادگان فارسی که به وسیله پنج زن و پنج مرد گفته شده بودند، انتخاب شده و در آزمایش های مختلف، سه نویز سفید گوسی، جنگنده F16 و همهمه در چهار سطح سیگنال به نویز ۵-، ۱۰-، ۵ و ۱۰ به سیگنال اضافه گردید. (نویز سفید گوسی به وسیله نرم افزار MATLAB تولید شده است).

۱-۸-۶ معیار ارزیابی سیگنال به نویز با استفاده از روش الف

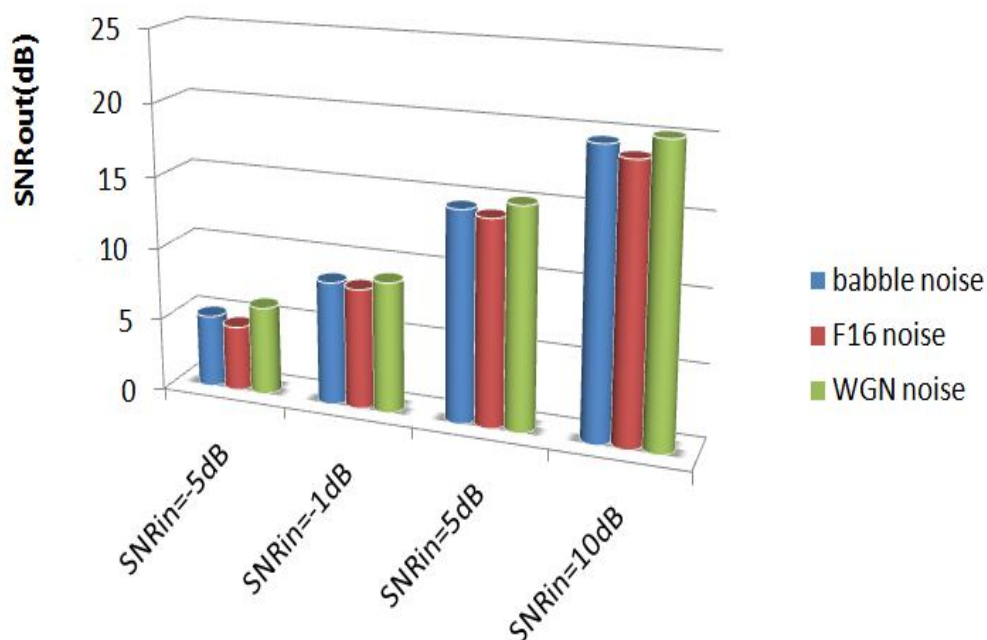
در این قسمت، با استفاده از ضبط نویز در لحظات سکوت، نویز تخمینی برای شبکه ویونت را محاسبه کرده ایم. در ادامه با استفاده از سه معیار SNR و SNRseg و LLR، به ارزیابی روش پیشنهادی خواهیم پرداخت.

۱-۱-۸-۶ معیار ارزیابی سیگنال به نویز

همانطور که قبلا بیان شد، SNR یکی از متداولترین معیارهای ارزیابی کیفیت سیگنال بهبود یافته می باشد که مقدار آن به صورت رابطه ی (۵-۶) بدست می آید.

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N S^2(n)}{\sum_{n=1}^N [s(n) - \hat{S}(n)]^2} \quad (5-6)$$

در این رابطه s سیگنال گفتار تمیز و $\hat{S}$ سیگنال گفتار بهبود یافته می‌باشد. هر اندازه این مقدار بیشتر باشد نشانگر کارایی و عملکرد مطلوب شبکه در حذف نویز می‌باشد. معیاری که در این قسمت مورد بررسی قرار داده شده است، سیگنال به نویز می‌باشد. مقادیر این معیار، در نمودار شکل ۶-۱۲ و هم در جدول ۶-۸ برای سه نویز سفید، همهمه و F16، در چهار سطح سیگنال به نویز ۵-، -۱، ۵ و ۱۰ دسی بل ذکر شده است. مقایسه SNR سیگنال خروجی با توجه به انواع نویز و SNR سیگنال ورودی در شکل ۶-۱۲ آمده است.



شکل ۶-۱۲: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف

همانطور که در جدول ۶-۸ آمده است، روش پیشنهادی، SNR سیگنال آلوده به نویز سفید ۵dB- را به ۵dB ۶.۱۲ می‌رساند. این مقدار در مورد سیگنال گفتار آلوده به نویز همهمه به ۵.۰۲dB و در مورد سیگنال گفتار آلوده به نویز F16 به ۴.۵dB رسیده است.

با اعمال گفتار آلوده به نویز سفید ۱dB- به شبکه پیشنهادی، این معیار در خروجی به ۹.۰۱dB خواهد رسید. این مقدار در مورد سیگنال گفتار آلوده به نویز همهمه به ۸.۵۳dB و در مورد سیگنال گفتار آلوده به نویز F16

به ۸.۳dB رسیده است. در مورد سیگنال گفتار آلوده به نویز سفید ورودی با SNR برابر با ۵dB و ۱۰dB، این

معیار در سیگنال گفتار بهبود یافته به ترتیب برابر با ۱۵.۱۴dB و ۲۰.۱۵dB شده است.

جدول ۶-۸: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف

Input SNRs(dB) ↓ Kind Of Noise →	Output SNRs(dB) ↓		
	babble	F16	WGN
-۵	۵.۰۲	۴.۵	۶.۱۲
-۱	۸.۵۳	۸.۳	۹.۰۱
۵	۱۴.۴	۱۴.۱	۱۵.۱۴
۱۰	۱۹.۵۱	۱۸.۷۳	۲۰.۱۵

در مورد سیگنال گفتار آلوده به نویز همهمه و نویز F16 با سیگنال به نویز ۵dB، روش پیشنهادی این معیار را

در گفتار خروجی به ترتیب به ۱۴.۴۸dB و ۱۴.۱dB می رساند. همانطور که مشاهده می شود، بهترین جواب

ها مربوط به حذف نویز سفید و سپس همهمه می باشد.

۶-۸-۱-۲ معیار ارزیابی سیگنال به نویز قطعه ای

معیار که در این قسمت مورد بررسی قرار داده شده است، سیگنال به نویز قطعه ای می باشد. قبلا بیان شد که

در این معیار SNR در قاب های کوچک تر محاسبه شده و میانگین این مقادیر به عنوان ارزیابی نهایی ارائه

می گردد.

رابطه این معیار به صورت زیر است که در آن M تعداد قاب ها و L طول هر قاب است.

$$SNR_{seg} = \frac{10}{M} \sum_{f=1}^M \log_{10} \left[\frac{\sum_{n=L.F}^{L.f+L-1} s^2(n)}{\sum_{n=L.F}^{L.f+L-1} [s(n) - \hat{s}(n)]^2} \right] \quad (6-6)$$

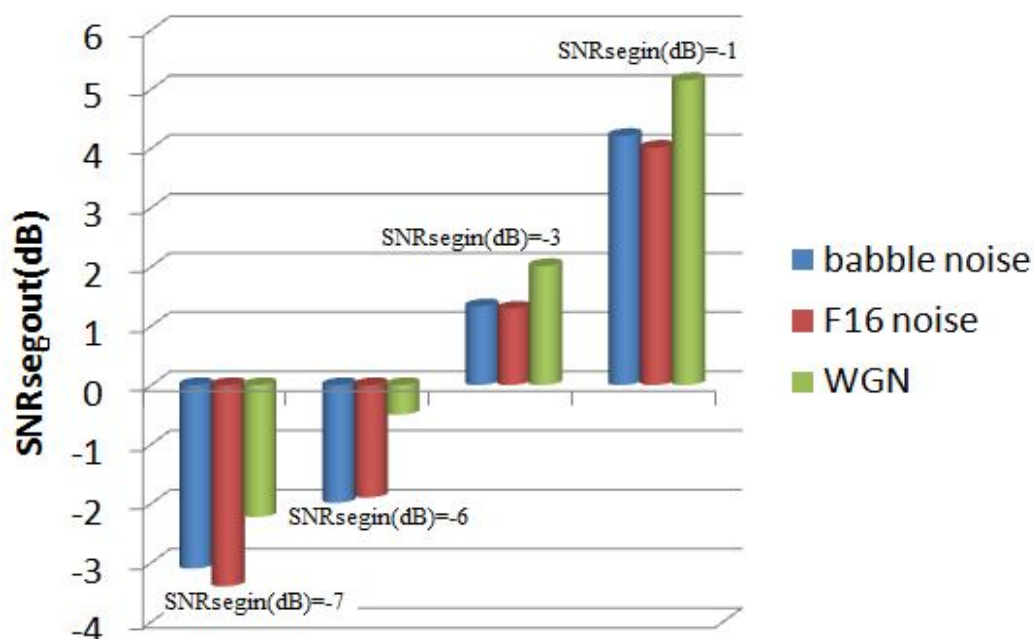
مقادیر این معیار، در نمودار شکل ۶-۱۳ و هم در جدول ۶-۹ برای سه نویز سفید، همهمه و F16، در چهار

سطح سیگنال به نویز قطعه ای ۷-، ۶-، ۳- و ۱- دسی بل ذکر شده است.

جدول ۶-۹: مقایسه SNRseg سیگنال خروجی به ازای SNRseg ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف

Input SNRseg(dB) ↓ Kind Of Noise →	Output SNRseg(dB) ↓		
	babble	F16	WGN
-۷	-۳.۱	-۳.۴	-۲.۲۳
-۶	-۱.۹۹	-۱.۹	-۰.۵
-۳	۱.۲۳	۱.۲۹	۲.۰۱
-۱	۴.۲۱	۴.۰۱	۵.۱۵

لازم به ذکر است که در این حالت محاسبه سیگنال به نویز قطعه ای ورودی نیز الزامی است. زیرا به ازای یک سیگنال به نویز ورودی، اگر سیگنال به نویزهای قطعه ای ورودی متفاوت باشد، سیگنال به نویزهای قطعه ای خروجی نیز متفاوت خواهند بود.



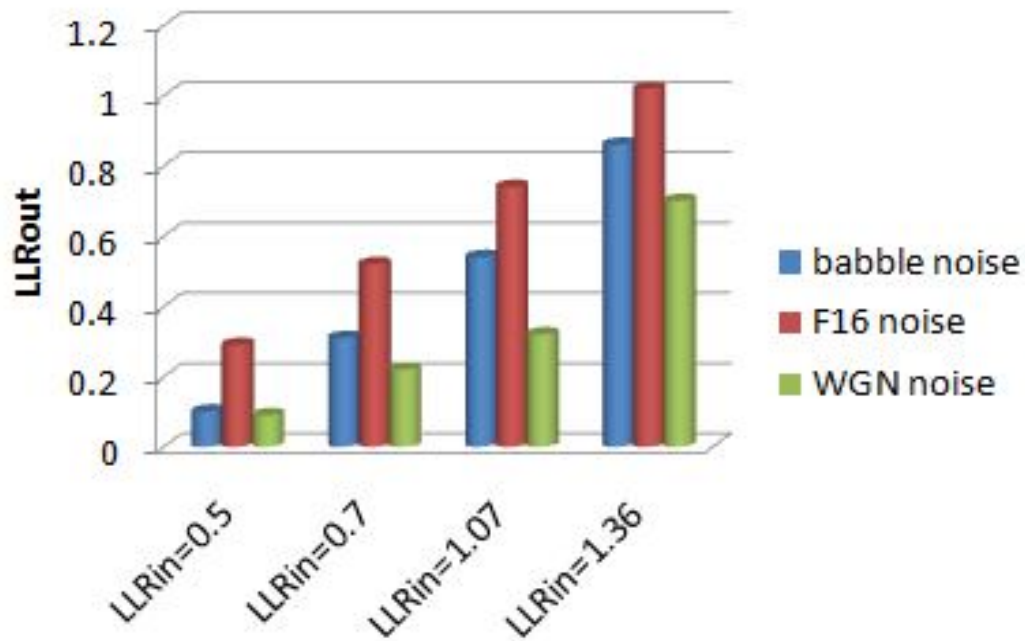
شکل ۶-۱۳: مقایسه SNRseg سیگنال خروجی به ازای segSNR ورودی مختلف

۳-۱-۸-۶ معیار ارزیابی LLR

پس از بررسی معیار سیگنال به نویز و سیگنال به نویز قطعه ای، معیار LLR برای سه نویز مذکور و چهار سطح سیگنال به نویز قبل بررسی شده است. مقادیر معیار LLR برای این سه نویز و چهار سطح سیگنال به نویز، در نمودار شکل ۱۴-۶ و جدول ۱۰-۶ رسم شده است.

جدول ۱۰-۶: مقایسه LLR سیگنال خروجی به ازای LLR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی الف

Input LLR ↓ Kind Of Noise →	Output LLR ↓		
	babble	F16	WGN
۰.۵	۰.۱	۰.۲۹	۰.۰۹
۰.۷	۰.۳۱	۰.۵۲	۰.۲۲
۱.۰۷	۰.۵۴	۰.۷۴	۰.۳۲
۱.۳۶	۰.۸۶	۱.۰۲	۰.۷



شکل ۱۴-۶: مقایسه LLR سیگنال خروجی به ازای LLR ورودی مختلف

قبلا بیان شد که کمتر بودن مقدار LLR به معنی بهتر بودن کیفیت سیگنال خروجی است. در این بخش نیز بهترین جواب ها مربوط به حذف نویز سفید و سپس همه می باشد.

۲-۸-۶ معیار ارزیابی سیگنال به نویز با استفاده از روش ب

در این قسمت، برای تخمین نویز از تبدیل موجک استفاده کرده ایم. در ادامه با استفاده از دو معیار SNR و SNRseg، به ارزیابی روش پیشنهادی خواهیم پرداخت.

۱-۲-۸-۶ معیار ارزیابی سیگنال به نویز

همانطور که در جدول ۱۱-۶ مشاهده می شود، روش پیشنهادی ب، SNR سیگنال آلوده به نویز سفید -5dB را به 2.57dB می رساند. این مقدار در مورد سیگنال گفتار آلوده به نویز همه به -4.8dB و در مورد سیگنال گفتار آلوده به نویز F16 به -3.17dB رسیده است. با اعمال گفتار آلوده به نویز سفید -1dB به شبکه پیشنهادی، این معیار در خروجی به 4.9dB خواهد رسید. این مقدار در مورد سیگنال گفتار آلوده به نویز همه به -0.8dB و در مورد سیگنال گفتار آلوده به نویز F16 به 2.64dB رسیده است.

در مورد سیگنال گفتار آلوده به نویز سفید ورودی با SNR برابر با 5dB و 10dB ، این معیار در سیگنال گفتار بهبود یافته به ترتیب برابر با 8.3dB و 8.63dB شده است. در مورد سیگنال گفتار آلوده به نویز همه و نویز F16 با سیگنال به نویز 5dB ، روش پیشنهادی این معیار را در گفتار خروجی به ترتیب به 5.2dB و 6.46dB می رساند.

همانطور که مشاهده می شود، بهترین جواب ها مربوط به حذف نویز سفید و سپس نویز F16 می باشد.

جدول ۶-۱۱: مقایسه SNR سیگنال خروجی به ازای SNR ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی ب

Input SNRs(dB)↓ Kind Of Noise →	Output SNRs(dB) ↓		
	babble	F16	WGN
-۵	-۴.۸	-۳.۱۷	۲.۵۷
-۱	-۰.۸	۲.۶۴	۴.۹
۵	۵.۲	۶.۴۶	۸.۳
۱۰	۱۰.۵۹	۱۱.۷۰	۸.۶۳

۶-۲-۸-۲ معیار ارزیابی سیگنال به نویز قطعه ای

همانطور که در جدول ۶-۱۲ مشاهده می‌شود، روش پیشنهادی ب، معیار سیگنال به نویز قطعه‌ای را در گفتار بهبود یافته، به خوبی افزایش داده است و افزایش این معیار در روش پیشنهادی، نشان‌دهنده عملکرد مناسب و مطلوب این روش می‌باشد.

همانطور که مشاهده می‌شود، روش پیشنهادی، حذف نویز سفید را از گفتار آلوده نسبت به سایر نویزها، بهتر انجام داده است.

جدول ۶-۱۲: مقایسه SNRseg سیگنال خروجی به ازای SNRseg ورودی مختلف در گفتار آلوده به نویز سفید گوسی و F16 و همهمه در روش پیشنهادی ب

Input SNRseg(dB)↓ Kind Of Noise →	Output SNRseg(dB) ↓		
	babble	F16	WGN
-۷	-۶.۱	-۶.۲	-۴.۹
-۶	-۵.۲	-۴.۹	-۱.۹
-۳	-۲.۲	-۲.۶	-۱.۱
-۱	۰.۲۱	-۰.۴۱	۰.۵

۹-۶ مقایسه روش پیشنهادی با روش های دیگر

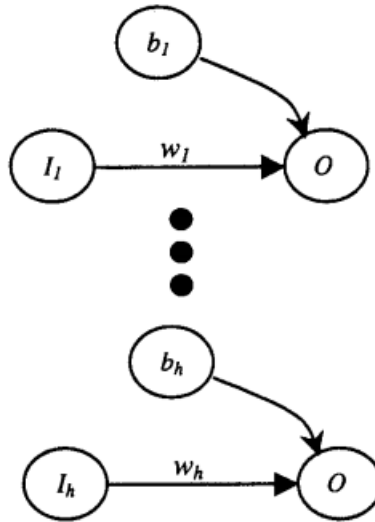
برای بررسی بهبود یا عدم بهبود روش پیشنهادی، سه روش دیگر بهسازی گفتار پیاده سازی شدند و روش پیشنهادی با این سه روش مقایسه گردیده است. نتایج این مقایسه ها در این بخش آورده شده است. در ابتدا با استفاده از شبکه عصبی آدالاین [۲] و استفاده از الگوریتم های کمترین میانگین مربعات خطا (LMS) و بیشترین شیب نزول (Steepest Descent)، عمل حذف نویز بر روی سیگنال نویزی گفتار فارسی انجام و نتایج آن مورد بررسی قرار می گیرد.

سپس در روش دوم، از ترکیب شبکه ی عصبی و تبدیل موجک استفاده کرده؛ به این صورت که از ورودی های شبکه قبل از ورود به شبکه، تبدیل موجک گرفته خواهد شد؛ و نتایج آن بررسی خواهد شد. این دو روش به طور مختصر در فصل سوم توضیح داده شده اند. روش سوم استفاده از تبدیل موجک می باشد.

همانطور که قبلا بیان شد، شبکه عصبی آدالاین، شبکه عصبی مصنوعی ساده ای می باشد که از یک لایه تشکیل شده است و تابع فعالساز آن تابعی خطی می باشد. این شبکه براساس الگوریتم وفقی عمل کرده و از لحاظ کاربرد در سه گروه فیلترینگ، هموارسازی و پیش گویی دسته بندی می شود [۵۲].

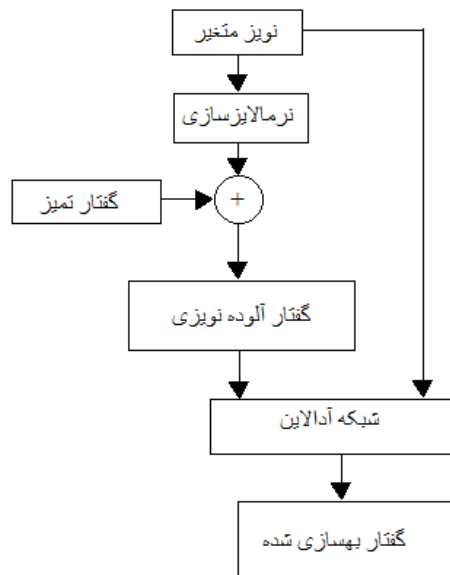
در ابتدا سیگنال گفتار فارسی آمیخته با نویز، به عنوان ورودی به سیستم اعمال شده و سیستم با استفاده از الگوریتم های بیان شده، سعی در شناسایی ویژگی های منبع نویز و تولید نویزی مشابه می نماید و سپس با تفریق سیگنال نویز تولیدی از سیگنال گفتار نویزی، منجر به حذف ویا کاهش نویز سیگنال می گردد.

ورودی در وزن های مرتبط ضرب گشته و خروجی را تشکیل می دهند. خروجی واقعی با خروجی مطلوب مقایسه گشته و سیگنال خطا را تشکیل می دهد. هدف رسیدن به کمترین خطای ممکن می باشد. در شکل ۱۵-۶ شبکه عصبی آدالاین نشان داده شده است.



شکل ۶-۱۵: شمای کلی شبکه عصبی آدالاین [۲۲]

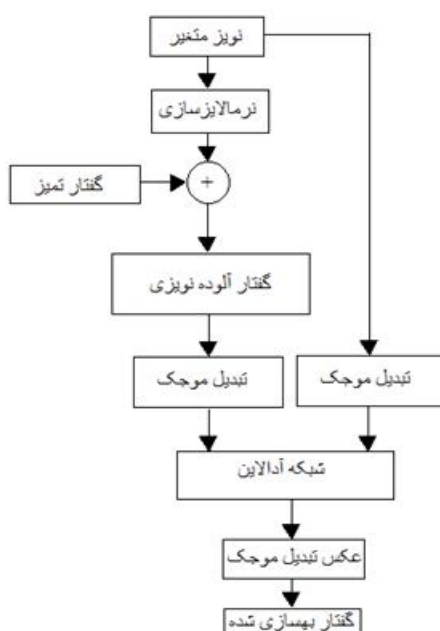
روند طی شده برای رسیدن به سیگنال گفتار بهسازی شده در شکل ۶-۱۶ آمده است. همانطور که گفته شد، گفتار آلوده به نویز و نویز ورودیهای این شبکه می باشند.



شکل ۶-۱۶: روند طی شده در بهسازی گفتار با شبکه عصبی آدالاین [۲]

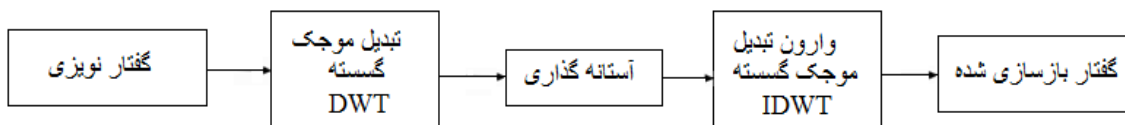
در روش دوم از ترکیب شبکه عصبی و تبدیل موجک استفاده شده است. به این ترتیب که از ورودیهای شبکه

عصبی، نویز و سیگنال گفتار نویزی، ابتدا تبدیل موجک گسسته گرفته می‌شود و نتیجه‌ی آن به عنوان ورودی به شبکه آدالاین اعمال خواهد شد. روند طی شده در این روش در شکل ۶-۱۷ آمده است. همانطور که در شکل ۶-۱۷ مشاهده می‌شود، برای رسیدن به سیگنال گفتار مطلوب از خروجی این شبکه تبدیل موجک عکس گرفته می‌شود. هدف از این روش، حذف نویز از سیگنال گفتار نویزی تجزیه شده در زیرباندهای فرکانسی مجزا می‌باشد. با استفاده از این روش می‌توان فرکانس‌های خاص را حذف کرد و فرکانسهای ضروری برای کاربردهایی همچون شناسایی گفتار را نگه داشت. همچنین با حذف تعدادی نمونه در سیگنال گفتار ورودی، فرایند حذف نویز را می‌توان تسریع بخشید [۲]. در روش سوم، بهسازی گفتار و کاهش نویز با استفاده از مشخص کردن حد آستانه به کمک ضرایب تبدیل موجک نویز و آستانه گذاری بر روی ضرایب تبدیل موجک سیگنال گفتار نویزی، انجام شده است. روند طی شده در این روش در شکل ۶-۱۸ آمده است.



شکل ۶-۱۷: روند طی شده در بهسازی گفتار با شبکه عصبی آدالاین و موجک [۲]

لازم به ذکر است که در روش‌های بیان شده، موجک استفاده شده، دابیشز ۵ (db5) می‌باشد. این نوع موجک با تکرار آزمایش و مقایسه‌ی نتایج مختلفی که با موجک‌های دیگر حاصل می‌شد، نتیجه شده است.



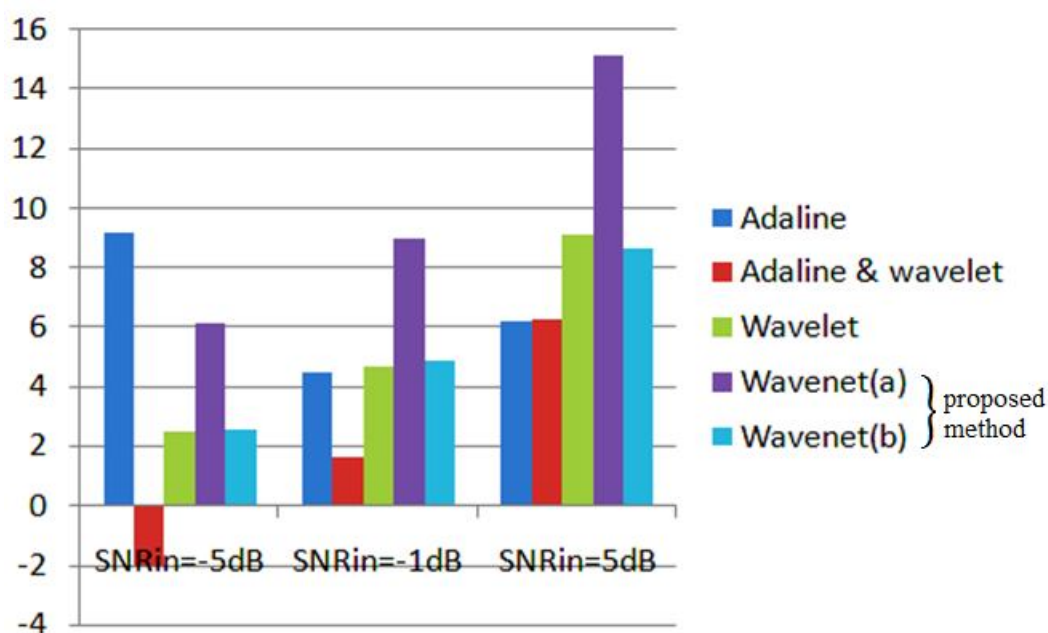
شکل ۶-۱۸: روند طی شده در بهسازی گفتار با روش تبدیل موجک

برای مقایسه عملکرد روش پیشنهادی ما در این پایان نامه با این روشهای پیشنهاد شده از معیارهای ارزیابی استفاده کرده ایم که نتایج آن در قسمت بعد آمده است. لازم به ذکر است که دادگان ما در همه ی روشهای بیان شده، همان ده جمله از پایگاه دادگان فارسی است که به وسیله پنج زن و پنج مرد گفته شده است و در آزمایش های مختلف، سه نویز سفید گوسی، جنگنده F16 و همهمه در سه سطح سیگنال به نویز ۵-، ۱۰- و ۵ به سیگنال اضافه گردیده است. نکته ی قابل توجه در شکلهایی که در قسمتهای بعدی آمده است اینست که منظور ما از روش wavenet(a)، روش پیشنهادی ما در حالت الف و از wavenet(b)، روش پیشنهادی در حالت ب می باشد. هر دسته از نمودارهای میله ای، در شکل های ترسیمی، مقدار معیار مربوطه را، به ترتیب از چپ به راست، برای روش آدالاین، ترکیب آن با موجک، موجک و روش پیشنهادی الف و ب نشان می دهد.

۱-۹-۶ معیار سیگنال به نویز

اولین معیاری که برای مقایسه ی عملکرد این روشها بیان شده است، معیار سیگنال به نویز (SNR) می باشد. ابتدا سیگنال آلوده به نویز سفید ۵dB+ به عنوان سیگنال نویزی ورودی در نظر گرفته شده است. همانطور که در شکل ۶-۱۹ نشان داده شده است، روش پیشنهادی در حالت الف بیشترین SNR را که برابر ۱۵.۱۴dB می باشد، ایجاد کرده و سیگنال به نویز ورودی را ۱۰.۱۴dB افزایش داده است. بعد از آن روش تبدیل موجک و سپس روش پیشنهادی ما در حالت ب بیشترین SNR را نشان می دهند. نتیجه ی دو روش دیگر آدالاین و ترکیب آن با موجک، تقریباً یکسان است. در مرحله بعد، سیگنال آلوده به نویز سفید گوسی ۱ dB- در ورودی اعمال شده است. همانطور که در جدول ۶-۱۳ مشاهده می شود، بیشترین افزایش SNR در روش ویونت حالت الف رخ داده است. در روش ویونت حالت الف پیشنهادی این مقدار سیگنال به نویز تا

۹.۰۱ dB و در حالت ب تا ۴.۹ dB افزایش یافته است.



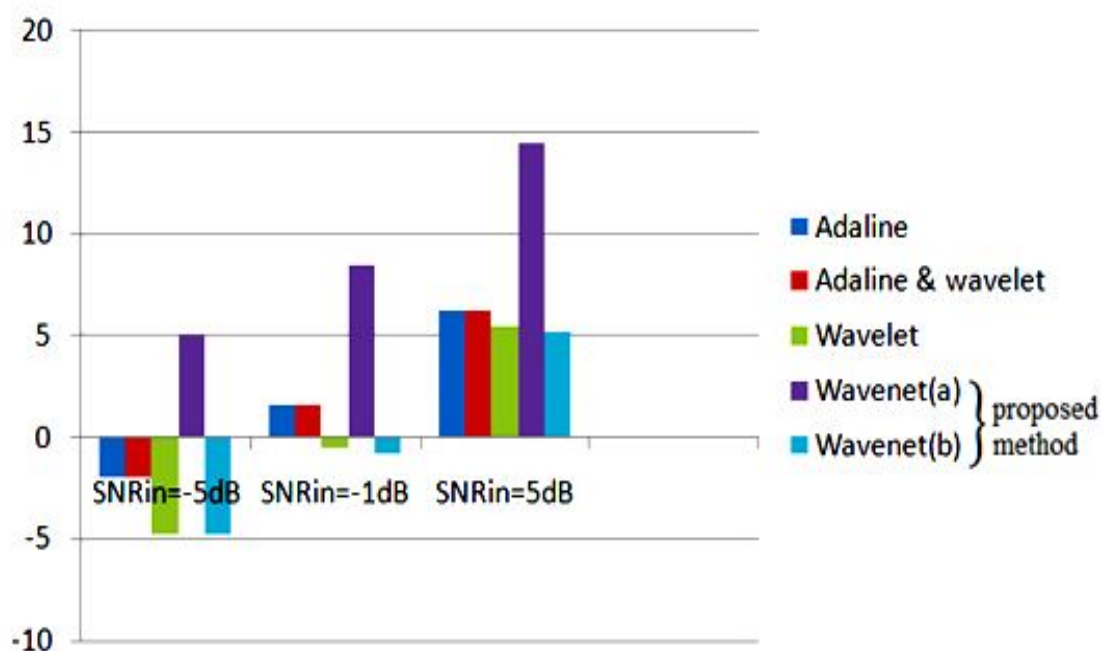
شکل ۶-۱۹: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNR برای نویز سفید گوسی

آزمایش بعدی را در شرایط سیگنال به نویز -۵dB پیاده کرده ایم. بیشترین افزایش سیگنال به نویز در روش آدالاین مشاهده شده است. این روش سیگنال به نویز خروجی را تا ۹.۱۶ dB افزایش داده است. بعد از این روش، افزایش سیگنال به نویز در روش ویونت حالت الف و ب به خوبی مشاهده می شود. ویونت پیشنهادی در حالت الف سیگنال به نویز ورودی را به اندازه ۱۱.۱۲dB و در حالت ب به اندازه ۷.۵۷dB افزایش داده است. نتایج بیان شده در جدول ۶-۱۳ آمده است.

جدول ۶-۱۳: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNR برای نویز سفید گوسی

Input SNRs(dB) ↓ Methodes →	Output SNRs(dB) ↓				
	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet (b)
-۵	۹.۱۶	-۲.۰۲	۲.۵	۶.۱۲	۲.۵۷
-۱	۴.۵	۱.۶۱	۴.۷	۹.۰۱	۴.۹
۵	۶.۲۴	۶.۲۵	۹.۱	۱۵.۱۴	۸.۶۳

تا بدین مرحله نتایج حاصل از روشهای بیان شده، در مورد سیگنال گفتار آلوده با نویز سفید را بیان کردیم. در ادامه به بررسی این روشها و عملکردشان بر روی سیگنال گفتار آلوده با نویز همهمه خواهیم پرداخت. در ابتدا سیگنال آلوده با سیگنال به نویز همهمه ۵dB را به عنوان ورودی در نظر گرفته ایم. در شکل ۶-۲۰ روش پیشنهادی در حالت الف، بیشترین SNR را نشان می دهد.



شکل ۶-۲۰: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف و ب بر حسب SNR برای نویز همهمه

همانطور که در جدول ۶-۱۳ بیان شده است، این روش سیگنال به نویز ۵dB را به ۱۴.۴۸dB رسانده است. این مقدار نشاندهنده کاهش مطلوب نویز در سیگنال خروجی می باشد. دو روش آدالاین و ترکیب آن با موجک در مقایسه با روش پیشنهادی ویونت در حالت ب عملکرد مطلوبتری داشته و SNR سیگنال در خروجی را بیشتر افزایش داده اند. SNR سیگنال خروجی در روش ویونت حالت ب برابر ۵.۲dB شده است. در مرحله بعد، از گفتار آلوده به نویز همهمه با SNR برابر ۱dB - استفاده شده است. همانطور که در جدول ۶-۱۳ بیان شده است، ویونت الف مقدار SNR در خروجی را به ۸.۵۳ dB و ویونت ب این مقدار را به ۰.۸dB - افزایش می -

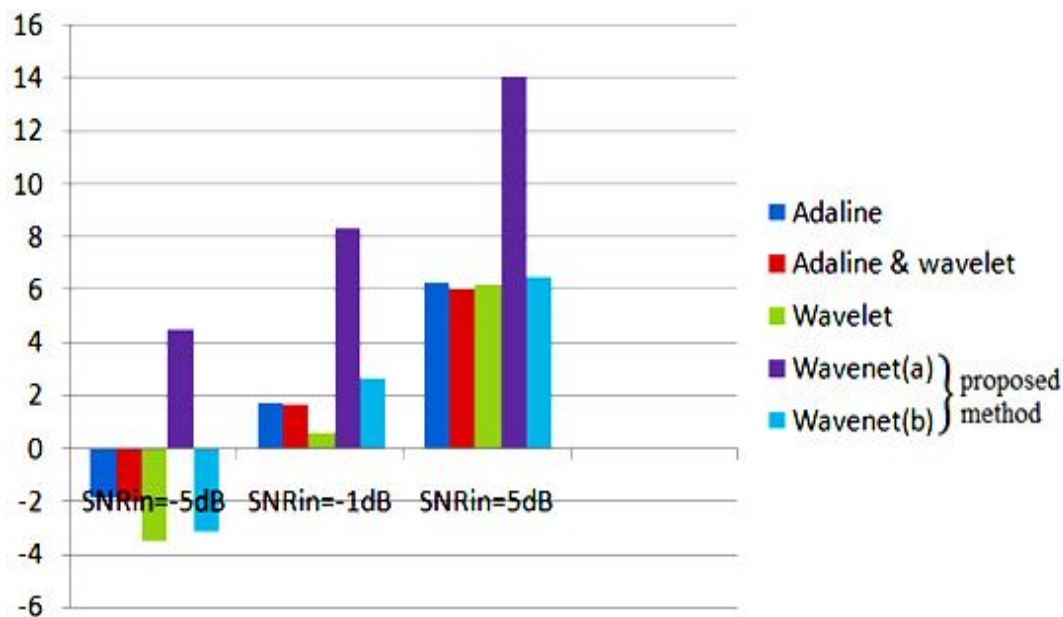
دهند. سیگنال به نویز خروجی در روش آدالاین ۱.۶۶ dB می باشد که بعد از روش پیشنهادی الف ما، بیشترین SNR را ایجاد کرده است. بعد از آن روش موجک و روش پیشنهادی ب، SNR را در سیگنال خروجی افزایش داده اند.

در آزمایش بعدی، از سیگنال آلوده به نویز هممه با سیگنال به نویز ۵dB- استفاده شده است. در این شرایط، SNR سیگنال گفتار بهبود یافته، با استفاده از روش پیشنهادی الف، برابر ۵.۰۲ dB شده است، که در مقایسه با سایر روشها بیشترین SNR می باشد. سایر مقادیر در جدول ۶-۱۳ بیان شده است.

جدول ۶-۱۴: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNR برای نویز هممه

Input SNRs(dB) ↓ Methodes →	Output SNRs(dB) ↓				
	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet (b)
-۵	-۱.۹۷	-۱.۹۳	-۴.۸	۵.۰۲	-۴.۸
-۱	۱.۶۶	۱.۶۵	-۰.۵	۸.۵۳	-۰.۸
۵	۶.۲۵	۶.۲۱	۵.۵	۱۴.۴۸	۵.۲

در ادامه به بررسی این روشها و عملکردشان بر روی سیگنال گفتار آلوده با نویز F16 خواهیم پرداخت. در این مرحله، ابتدا سیگنال گفتار آلوده به نویز F16 با SNR برابر با ۵dB را برای هر پنج روش آزمایش کرده ایم. شکل ۶-۲۱ مقایسه ای بین SNR در سیگنال بهبود یافته حاصل از این روشها را، نشان می دهد. در این حالت، روش پیشنهادی الف بیشترین SNR خروجی را ایجاد کرده است، که این مقدار برابر ۱۴.۱dB می باشد. در مرحله بعد، سیگنال آلوده به نویز F16 با SNR برابر ۱dB- در ورودی اعمال شده است. همانطور که در جدول ۶-۱۵ مشاهده می شود، بیشترین افزایش SNR در روش ویونت حالت الف رخ داده است. این مقدار برابر با ۸.۳dB می باشد.



شکل ۲۱-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف و ب بر حسب SNR برای نویز F16

در روش پیشنهادی ب این مقدار به ۲.۶۴dB افزایش یافته است. همانطور که مشاهده می شود، SNR در سیگنال گفتار بهبود یافته، در روش پیشنهادی ما نسبت به سایر روشها افزایش بیشتری داشته است. در آزمایش بعدی، از سیگنال آلوده به نویز F16 با سیگنال به نویز -۵dB استفاده شده است. در این آزمایش نیز روش پیشنهادی الف، بیشترین افزایش SNR را در خروجی داشته است. این نتایج در نمودار شکل ۲۱-۶ بیان شده است.

جدول ۱۵-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف و ب بر حسب SNR برای نویز F16

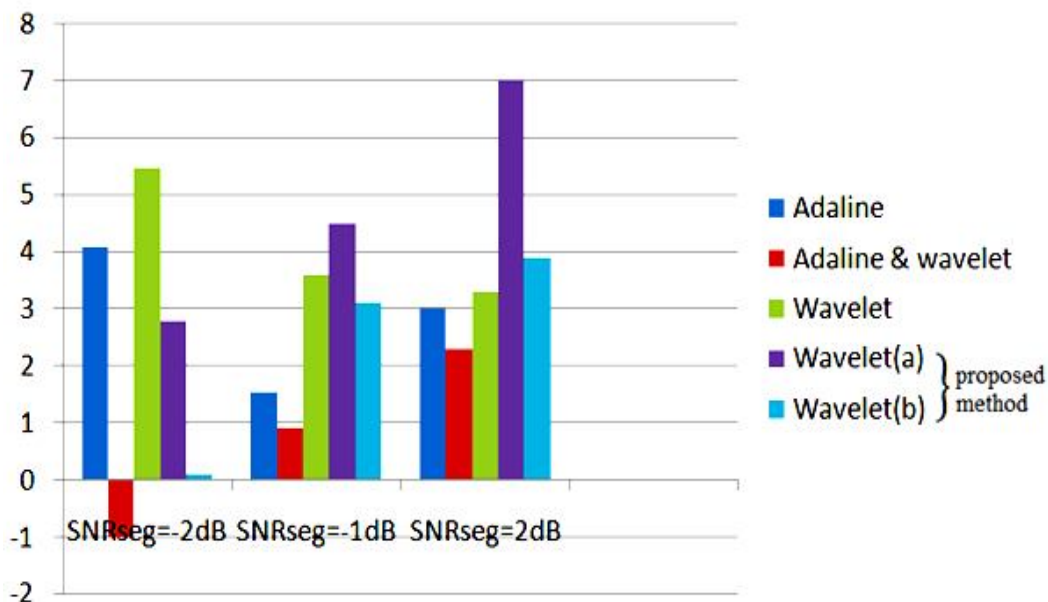
Input SNRs(dB)↓	Output SNRs(dB) ↓					
	Methodes→	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet(b)
-۵		-۱.۹	-۲.۰۱	-۳.۵	۴.۵	-۳.۱۷
-۱		۱.۷۲	۱.۶۸	۰.۵۹	۸.۳	۲.۶۴
۵		۶.۳	۶.۲۵	۶.۲۶	۱۴.۱	۶.۴۶

۲-۹-۶ معیار سیگنال به نویز قطعه‌ای

در این قسمت از معیار ارزیابی سیگنال به نویز قطعه‌ای (SNRseg)، برای مقایسه‌ی عملکرد روشهای بیان شده، استفاده شده است. ابتدا سیگنال آلوده به نویز سفید با سیگنال به نویز قطعه‌ای برابر با ۲dB به عنوان سیگنال نویزی ورودی در نظر گرفته شده است.

همانطور که در شکل ۲۲-۶ نشان داده شده است، روش پیشنهادی در این پایان‌نامه بیشترین SNRseg را نشان می‌دهد. طبق نتایج جدول ۱۶-۶ این معیار در گفتار بهبود یافته حاصل از ویونت الف برابر با ۷.۰۱dB و در ویونت ب برابر با ۳.۹dB شده است.

در مرحله‌ی بعد، SNRseg سیگنال گفتار آلوده ۱dB- می‌باشد. ویونت الف این مقدار را در سیگنال خروجی به ۴.۵dB و ویونت ب آنرا به ۳.۱dB افزایش داده اند که در مقایسه با سایر روشها، نتیجه‌ی مطلوبتری می‌باشد.



شکل ۲۲-۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف و ب بر حسب SNRseg برای نویز سفید گوسی

این روشها را برای گفتار آلوده به نویز سفید با سیگنال به نویز قطعه ای برابر -2dB پیاده کرده ایم. در این شرایط بیشترین SNRseg در سیگنال خروجی روش تبدیل موجک مشاهده شده است. بعد از آن روش آدالاین بیشترین SNRseg را نشان می دهد. ویونت الف و ب گفتار خروجی با SNRseg برابر با 2.77dB و 0.1 به ترتیب بوجود می آورند.

جدول ۶-۱۶: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف و ب بر حسب SNRseg برای نویز سفید گوسی

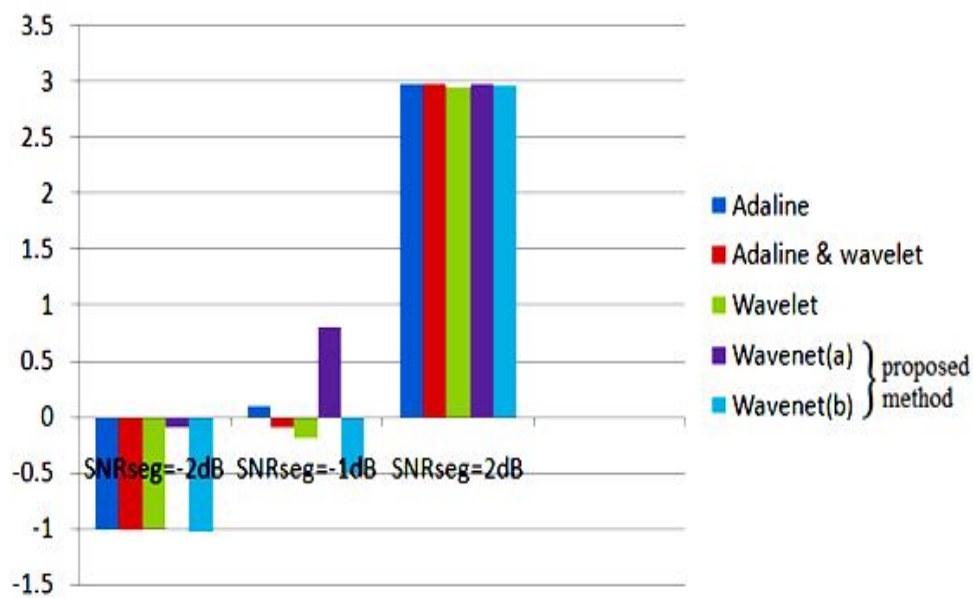
Input SNRseg(dB) ↓ Methodes →	Output SNRseg(dB) ↓				
	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet (b)
-۲	۴.۰۸	-۱.۰۱	۵.۴۸	۲.۷۷	۰.۱
-۱	۱.۵۳	۰.۹	۳.۶	۴.۵	۳.۱
۲	۲.۹۹	۲.۳	۳.۳	۷.۰۱	۳.۹

تا بدین مرحله SNRseg حاصل از روشهای بیان شده، در مورد سیگنال گفتار آلوده با نویز سفید را بیان کردیم. در ادامه به بررسی این معیار بر روی سیگنال گفتار آلوده با نویز همهمه خواهیم پرداخت. در ابتدا سیگنال آلوده با سیگنال به نویز همهمه 2dB را به عنوان ورودی در نظر گرفته ایم. همانطور که در شکل ۶-۲۳ مشاهده می شود، در این شرایط همه ی روشهای بیان شده تقریباً SNRseg یکسان ایجاد کرده اند.

در شرایطی که سیگنال به نویز ورودی -1dB می باشد، روش پیشنهادی حالت الف بیشترین SNRseg را در سیگنال گفتار بهبود یافته نشان می دهد که این مقدار برابر با 0.8dB می باشد. در سیگنال گفتار آلوده ی نویزی با SNRseg برابر با -2dB روش پیشنهادی الف SNRseg در گفتار بهبود یافته را به -0.1dB افزایش داده است.

نتایج حاصل از بررسی معیار سیگنال به نویز قطعه ای در روشهای بیان شده، در

جدول ۶-۱۷ آمده است.



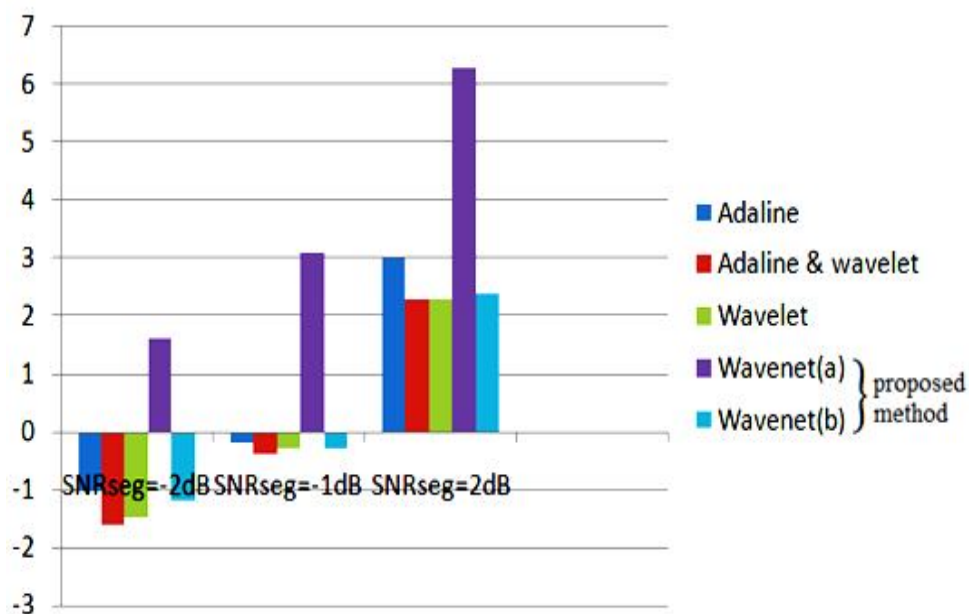
شکل ۶-۲۳: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNRseg برای نویز همه

جدول ۶-۱۷: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNRseg برای نویز همه

Input SNRseg(dB) ↓ Methodes →	Output SNRseg(dB) ↓				
	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet (b)
-۲	-۱.۰۱	-۱.۰۲	-۱	-۰.۱	-۱.۰۳
-۱	-۰.۱	-۰.۱	-۰.۲	۰.۸	-۰.۵
۲	۲.۹۹	۲.۹۸	۲.۹۵	۲.۹۸	۲.۹۷

در ادامه به بررسی معیار سیگنال به نویز قطعه‌ای، بر روی سیگنال آلوده به نویز F16 با روشهای گفته شده پرداخته‌ایم. همانطور که در شکل ۶-۲۴ مشاهده می‌شود، در شرایطی که سیگنال به نویز قطعه‌ای سیگنال ورودی ۲dB می‌باشد، روش ویونت الف و ب، بیشترین SNRseg را برای سیگنال گفتار بهبود یافته بوجود آورده‌اند. همانطور که در جدول ۶-۱۸ مشاهده می‌شود، ویونت الف این مقدار را به ۶.۲۹dB و ویونت ب این مقدار را به ۲.۴dB افزایش داده است. در شرایطی که سیگنال به نویز قطعه‌ای ورودی برابر با ۱dB - می‌باشد،

روش پیشنهادی الف این مقدار را به ۳.۱dB افزایش داده است.



شکل ۶-۲۴: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNRseg برای نویز F16

در روش پیشنهادی الف، در شرایط سیگنال گفتار آلوده به نویز F16 با SNRseg برابر با -۲dB، در SNRseg گفتار خروجی به ۱.۶ dB افزایش یافته است که این مقدار در بین نتیجه‌ی سایر روشها، بهترین نتیجه محسوب می‌شود. نتایج بیان شده در جدول ۶-۱۸ آورده شده است.

جدول ۶-۱۸: مقایسه روش پیشنهادی با روش های آدالاین، آدالاین و موجک، موجک و ویونت پیشنهادی الف وب بر حسب SNRseg برای نویز F16

Input SNRseg(dB) ↓ Methodes →	Output SNRseg(dB) ↓				
	Adaline	Adaline & wavelet	Wavelet	Wavenet (a)	Wavenet (b)
-۲	-۱.۰۲	-۱.۶۱	-۱.۵	۱.۶	-۱.۲
-۱	-۰.۲	-۰.۴	-۰.۳	۳.۱	-۰.۳
۲	۲.۹۹	۲.۳	۲.۳	۶.۲۹	۲.۴

همانطور که قبلاً بیان شد، برای روش پیشنهادی، سیگنال خطایی تعریف شد که هدف ما در اجرای این روش، به حداقل رساندن این سیگنال می‌باشد. معیار متوسط مربع خطا (MSE) در چهار روش شبکه عصبی آدالاین، ترکیب آن با ویولت، ویونت حالت الف وب، در جدول ۶-۱۹ آورده شده است. ابتدا الگوریتم آموزش همه‌ی روشها را تا ۵ بار تکرار کرده‌ایم. نتایج MSE بدست آمده نشان می‌دهد، هر دو روش پیشنهادی سیگنال خطای کوچکتری نسبت به دو روش دیگر ایجاد کرده اند و این نشان دهنده‌ی دقت شبکه‌ی پیشنهادی ما در تکرارهای پایین می‌باشد. همین فرایند را تا ده مرحله تکرار برای هر چهار روش انجام دادیم.

MSE ویونت پیشنهادی در حالت الف وب به ترتیب برابر با $3.1e-004$ و $4.13e-006$ شده است، حال آنکه این معیار در روش آدالاین و ترکیب آن با ویونت به ترتیب برابر با 0.0033 و 0.0034 شده است.

جدول ۶-۱۹: مقایسه MSE روش پیشنهادی با روش های آدالاین، آدالاین و موجک و ویونت پیشنهادی الف وب بر حسب تعداد تکرارهای مرحله آموزش

Number of Iterations↓ Methodes→	MSE (Mean squared normalized error) ↓			
	Adaline	Adaline & wavelet	Wavenet (a)	Wavenet(b)
۵	۰.۰۰۴۴	۰.۰۰۴۴	۳.۸e-۰۰۴	۵.۷e-۰۰۶
۱۰	۰.۰۰۳۳	۰.۰۰۳۴	۳.۱e-۰۰۴	۴.۱۳e-۰۰۶

از جدول فوق نتیجه می‌شود که با مشخص کردن خطای آستانه، روش پیشنهادی ما در تعداد تکرار کمتری به این خطا خواهد رسید و الگوریتم آموزش سریعتر متوقف می‌شود. با توجه به این موضوع، سرعت همگرایی بالا برای روش پیشنهادی را می‌توان، به عنوان ویژگی مهم آن اشاره کرد.

۶-۱۰ نتیجه گیری

در این پایان نامه به بررسی بهسازی گفتار، مفاهیم مربوط به آن و روش های مختلف بهسازی گفتار و روش پیشنهادی پرداخته شد. سپس دو حالت الف و ب برای نویز محیط در نظر گرفته شد. حالت الف مربوط به محیط های نویزی می باشد که دستیابی به نویز به طور مستقیم ممکن باشد. برای مثال در لحظات سکوت بتوان نویز را ضبط کرد و از آن در شبکه ی پیشنهادی استفاده کرد. حالت ب مربوط به شرایطیست که امکان دسترسی مستقیم به نویز وجود ندارد. در این حالت روش جدیدی برای تخمین سیگنال نویز از روی سیگنال نویزی (نویز با سیگنال جمع شده و مستقل از آن است) ارائه شد.

برای پیاده سازی روش پیشنهادی، سیگنال گفتار تمیز با سه نویز سفید گوسی، همهمه و F16 و در چهار سطح سیگنال به نویزهای ۵-، ۱۰-، ۵ و ۱۰ جمع گردید و سپس از سه معیار سیگنال به نویز و سیگنال به نویز قطعه ای و LLR برای ارزیابی روش استفاده شد. نتایج روش پیشنهادی با سه روش معروف دیگر بهسازی گفتار (شبکه عصبی آدالاین و ترکیب آدالاین و موجک و روش موجک) مقایسه شد. بررسی نمودارها و جداول موجود در این فصل، نشان می دهد که روش پیشنهادی حالت الف در تمامی انواع نویز، عملکرد خوبی در کاهش نویز در سیگنال گفتار نویزی داشته است. مطلب مهم دیگری که در روش پیشنهادی حالت ب، براساس معیارهای ارزیابی، می توان به عنوان نتیجه به آن اشاره کرد اینست که بهترین جواب ها، مربوط به نویز سفید گوسی است. بعد از نویز سفید، در مورد روش پیشنهادی الف، سیگنالهای آلوده به نویز همهمه، و در مورد روش پیشنهادی ب، سیگنال آلوده به نویز F16، بهترین جوابها را نتیجه داده اند.

در نهایت نیز مقایسه سرعت همگرایی روش پیشنهادی با سایر روشها بررسی گردید. روش پیشنهادی ما در این پایان نامه، سرعت همگرایی بالایی دارد. این نتیجه به این ترتیب حاصل شد که ابتدا سیگنال آلوده به نویز با SNR برابر با ۱dB- به عنوان ورودی به شبکه و یونت داده شد. بعد از بررسی تعداد تکرار الگوریتم آموزش در این شبکه، مشخص شد که تعداد تکرار لازم، برای ایجاد سیگنال گفتار بهبود یافته با سیگنال به نویز ۸dB

در خروجی، ۵ مرتبه می‌باشد و حال آنکه برای رسیدن به این نتیجه در روش شبکه عصبی آدالاین باید الگوریتم آموزش شبکه در این روش ۱۵ مرتبه تکرار شود و این نشانگر بالا بودن سرعت همگرایی در شبکه پیشنهادی می‌باشد.

از ویژگی‌های بارز این روش سرعت عمل بالای آن می‌باشد. این شبکه با توجه به ساختار چند مرحله‌ای که دارد، در مقایسه با سایر روشها، به سرعت به جواب نهایی می‌رسد و پاسخ گویی شبکه در اندک زمانی صورت می‌گیرد. نتایج مقادیر MSE بدست آمده نشان‌دهنده دقت بالای ویونت می‌باشد.

با بیان روش ویونت در این پایان نامه، ایده‌ای نو در ذهن ایجاد می‌شود که در شبکه های عصبی خاص، در لایه میانی، می توان از توابع موجک که انرژی ناپایدار و خاصیت نزولی دارند، استفاده کرد. همچنین مقایسه این روش با سایر روشها، نشان دهنده‌ی کارایی بالای این روش می‌باشد.

۱۱-۶ پیشنهادات

در روش پیشنهادی در حالت ب، برای رسیدن به نویز، به تخمین آن از طریق تبدیل موجک پرداختیم. در این حالت، روش پیشنهادی ما بهترین جواب را در سیگنال گفتار آلوده به نویز سفید از خود نشان داده است حال آنکه می‌توان از روشی برای تخمین نویز استفاده کرد که برای انواع نویز، جواب مطلوب حاصل گردد. تخمین نویز در این روش را می‌توان با سایر روشهای تخمین نویز از جمله روشهای VAD انجام داد. هرچه روش تخمین نویز دقیقتر عمل کند، بهسازی گفتار بهتر صورت خواهد گرفت.

تغییر الگوریتم شبکه‌ی پیشنهادی و ساختار آن از جمله تعداد ورودی های شبکه و یا تغییر در معیار خطا شاید نتایج بهتری را در حذف نویز از گفتار آلوده به نویز نشان دهد.

در این پایان نامه پیاده سازی شبکه ویونت، با استفاده از الگوریتم وفقی صورت گرفته است که پیاده سازی آن به فرم دیگر شاید مفیدتر واقع گردد.

- [1] Wiener. N, (1949), Extrapolation, Interpolation and Smoothing of Stationary Time Series: Engineering Application, MIT Press, Cambridge, Mass
- [2] kh. d. ibrahim, N. abu-isbeih and m. alfauri, speech Signal Enhancement Using Neural Network and Wavelet Transform, epartment of Communication and Electronics Engineering, Philadelphia University, P.O.Box 1, Amman .19392, Jordan, 2009 6th International Multi-Conference on Systems , Signals and Devices
- [3] Berouti. M and Schwartz. R, Makhoul. J and Cambridge . M. A, (1979) , Enhancement of Speech Corrupted by Acoustic Noise, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 208-211, Washington,DC, USA
- [4] Boll. S ,(1979), Suppression of Acoustic Noise in Speech Using Spectral Subtraction, IEEE Transaction on Aocostic, Speech and Signal Processing, Vol. 27, No. 2, pp. 113-120
- [5] Paliwal. K. K and Basu. A, (1987), A Speech Enhancement Method Based on Kalman Filtering, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 177-180, Dallas, Texas, USA
- [6] De Moor. B, (1993), The Singular Value Decomposition and Long and Short Space of Noisy Matrices, IEEE Transaction on Aocostic, Speech and Signal Processing, Vol. 41, No. 9, pp. 2826-2838
- [7] Dendrinios.M, Bakamidis.S, and Carayannis. G, (1991), Speech Enhancement from Noise: a regenerative approach, speech communication, Vol. 10, No. 1, pp. 45-76
- [8] Ephraim. Yand Van Trees. H. L, (1995), A Signal Subspace Approach for Speech Enhancement, IEEE Transaction on Aocostic, Speech and Signal Processing, Vol. 3, No. 12, pp. 252-266
- [9] Hendriks. R. C, (2008), Advances in DFT-Based Single-Microphon Speech Enhancement,Ph. D Thesis, Technischm Univerity, Delft
- [10] Ephraim. Y, Cohen. I, (2005), Recent Advancements in Speech Enhancement, In the Electrical Engineering HandBook, CRC Press
- [11] Ephraim. Y, (1992c), Statistical Model-Based Speech Enhancement System , Proceeding of the IEEE , Vol. 80, No. 10, pp. 1526-1555
- [12] Veisi.H, (2005), Model-Based Method for Noise Robust Speech Recognition Systems, M.S. Thesis, Computer Engineering Department, Sharif Universitie of Technology, Tehran, Iran
- [13] Loizou. P. C, (2007), Speech Enhancement:Theory and Practice, CRC Press
- [14] Rangachari. S, (2004), Noise Estimation Algorithms For Highly NON-Stationary Environment, Ph. D

Thesis, University of Texas at Dallas

- [15] OUSSAR.Y, DREYFUS.G(2000), Initialization by Selection for Wavelet Network Training, Laboratoire d'Électronique.École Supérieure de Physique et de Chimie Industrielles.10, rue Vauquelin.F - 75231 PARIS Cedex 05, FRANCE.
- [16] Chen. B, (2005), Speech Enhancement Using A Laplacian-Based MMSE Estimator, Ph. D. Thesis, Department of Electrical Engineering, Texas, USA, University of Texas
- [17] Talebi.M, Salehi.L, Barkouti.V,Cherif.A. Speech Enhancement with Bionic Wavelet Transform and Recurrent Neural Network. SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications
- [18] Veisi.H, (2005), Model-Based Method for Noise Robust Speech Recognition Systems, M.S. Thesis, Computer Engineering Department, Sharif University of Technology, Tehran, Iran
- [19] Hu. Y and Loizou. P, (2008), Evaluation of Objective Quality Measures for Speech Enhancement, , IEEE Transaction on Acoustic, Speech and Signal Processing, Vol. 16, No. 1, pp. 252-266
- [20] Paliwal. K and Alsteris. L, (2005), On the Usefulness of STFT Phase Spectrum in Human Listening Tests, Speech Communication, Vol. 45, No. 2, pp. 153-170
- [21] Vary. P, (1985), Noise Suppression by Spectral Magnitude Estimation Mechanism and Theoretical Limits, Signal Processing. Vol. 8, No. 4,
- [22] Ban Fah.L , Hussain.A & Abdul Samad.S.(2000 IEEE). Speech Enhancement by Noise Cancellation Using Neural Network. Multimedia Signal Processing Research Group Department of Electrical, Electronic & Systems Engineering Faculty of Engineering, Universiti Kebangsaan Malaysia 43600 UKM Bangi, Selangor D.E.
- [23] Lekutai.G. Adaptive Self-Tuning Neuro Wavelet Network Controllers.(March 31, 1997)
- [24] Soon Ing Yann . Transform Based Speech Enhancement Techniques .A thesis submitted to the Nanyang Technological University in fulfillment of the requirement for the degree of Doctor of Philosophy (2003)
- [25] Yul Y. Nazaruddin. Department of Engineering physics Institut Teknologi Bandung Yuliati. Electrical Engineering Department , Widya Mandala Surabaya Catholic University . Wavenet Based Modeling of Vehicle Suspension System . 2006 IEEE
- [26] Veitch.D. Wavelet Neural Networks and their application in the study of dynamical systems. Department of Mathematics University of York UK.(2005)
- [27] Tlucak.J , Juhar.J , Dobos.L, Clzmar.A "neural network based speech enhancement"(1999)
- [28] Rangachari. S, (2004), Noise Estimation Algorithms for Highly NON-Stationary Environment, Ph. D Thesis, University of Texas at Dallas
- [29] Xiaohui Li, Huaren Wu, Member, IEEE, and Jingjing Liu. Coordinated Design of Wide Area

Controllers and FACTS Controller for Enhancement of Power System Stability. 2010 International Conference on Power System Technology, IEEE

- [30] M.R. A. SHALLUF. Recurrent Neuro Fuzzy Controller Power System Stabilizer . Faculty of Electrical Engineering University technology Malaysia, MAY 2007
- [31] Lekutai . G .Adaptive self-tuning neuro wavelet network controllers . Virginia polytechnic institute and state university, Doctoral thesis, Blacksburg, Virginia, US, 1997.
- [32] Nazaruddin.y & Yuliaty,(2006).Wavenet Based Modeling of Vehicle Suspension System. Department of Engineering Physics , Institut Teknologi Bandung & Electrical Engineering Department,Widya Mandala Surabaya Catholic University
- [33] Jang Wai. R , Ming Chang . J, Intelligent control of induction servo motor drive via wavelet neural network, Elsevier, Electric Power Systems Research 61 (2002) 67–76.
- [34] Ghanbari.M. Karami-Mollaei, Elsevier, A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets, December2005
- [35] Pushpalatha.M , Nalini .N. A Compact Network with improved generalization using Wavelet basis function network for static Non-linear functions.Proceeding of international joint conference on neural networks.Atlanta.2009
- [36] Bahoura,M.,Rouat,J.,2001a. Wavelet speech enhancement using the teager energy operator. IEEE Signal Process.Lett.8,10–12.
- [37] Bahoura,M.,Rouat,J.,2001b. New approach for wavelet speech Enhancement. In:Euro speech 2001.Aalborg,Denmark,pp.1937–1940.
- [38] Chen,S.H.,Wang,J.F.,2004. Speech enhancement using Perceptual wavelet packet decomposition and Teager energy operator .J.VLSI Signal Process .System 36(2–3),125–139.
- [39]Cohen,I.,2001. Enhancement of speech using bark-scaled wavelet packet decomposition. In:Euro speech 2001.Aalborg,Denmark,pp.1933–1936.
- [40] Donoho,D.L.,Johnston,I.M.,1994. Ideal spatial adaptive via wavelet shrinkage .Biometrika 81,425–455.
- [41] Ghanbari,Y.,Karami,M.,2004. Spectral subtraction in the wavelet domain for speech enhancement. Internat . J.Software Inf.Technol.(IJSIT)1(1),26–30.
- [42] Johnston,I.M.,Silverman,B.W.,1997.Wavelet threshold estimators for data with correlated noise.J.Roy.Statist.Soc.Ser.B59,319–351.
- [43] Kamath,S.,Loizou,P,2002.Amulti-band spectral subtraction method for enhancing speech corrupted by colored noise .In: Proc. IEEE Internat . Conf. Acoust. Speech Signal Process.(ICASSP),Orlando, florida.

[۴۴] همایونپور.م و شریف نبوی .ا. " مقایسه و ارزیابی روشهای تشخیص گفتار از سکوت " آزمایشگاه سیستم های هوشمند

صوتی و گفتاری دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر

[۴۵] پورتنی. ا. لطف الهی یقین. م. امین فر. م. قلی زاده. س. "ارزیابی کارایی شبکه عصبی موجکی در تخمین نیروها

هیدرودینامیکی طولی و عرضی در اعضا استوانه فلز قائم با زبر جزئی "اولین کنفرانس ملی سازه و فولاد

[۴۶] محمدپوری. ز. "کاربرد مدل آماری مخلوط لاپلاس در بهسازی گفتار" پایان نامه کارشناسی ارشد. ۱۳۸۹

[۴۷] عبدالله زاده. م. و محبوب. م. و زرین قلم. ر. "شناسایی سیستم دینامیکی چندورودی چند خروجی" هفدهمین کنفرانس بین

المللی. ۱۳۸۸.

[۴۸] اسفندیان. ن. (۱۳۸۶)، "استفاده از آنالیز LPC در روش تبدیل موجک، جهت غنی سازی سیگنال گفتار"، پایان نامه

کارشناسی ارشد، دانشکده فنی مهندسی، دانشگاه مازندران.

[۴۹] خالوزاده. ح.، حفیظی. م.، شاهی. م. "نویز گیری سیگنالهای MRS با استفاده از تحلیل مولتی رزولوشنی در حوزه

موجک". چهارمین کنفرانس ژئوفیزیک تهران، اردیبهشت ۸۹

[۵۰]. روشن دل کاهوا، سیاهکوهی. ح. "تخمین موجک چشمه لرزه ای با استفاده از تبدیل موجک" سیزدهمین کنفرانس

ژئوفیزیک تهران

[۵۱] کبیریان. م.، "طراحی کنترلر ادوات FACTS برای بهبود پایداری دینامیکی سیستم قدرت با استفاده از تبدیل

موجک" پایان نامه ۱۳۹۰

[۵۲] شریفی تهرانی. ا.، عشوریان. م.، خالویی. ت. "کاهش نویز سیگنال صحبت فارسی به صورت فعال - وفقی بر اساس شبکه عصبی

مصنوعی جهت بهینه سازی تشخیص گفتار" دومین کنفرانس ملی مهندسی برق ۱۳۸۸

Summary

The name " speech enhancement " refers to large group of methodes that improve the quality and intelligibility of noisy speech by suppressing the background noise from noisy signal.

There are a lot of methodes and papers about speech enhancement.

In this paper,we propose a new speech enhancement system using the wavele neural network.

Wavenet is a special feedforward neural network, which uses wavelet basis function as an activation function.

The wavelet coefficients for the specific signal processing problems are obtained by updating the weights of the adaptive Wavenet using the coinjugate gradient method.WNN uses a feedforward network with a learning algorithm that optimizes the network parameters in such a way that the mean squared error (MSE) between the desirable signal and the output signal is minimized. Daubechies (db5) mother function is used in the estimation of noise .The proposed method was evaluated on several speakers and under various noise conditions including white Gaussian noise , babble noise and F16.

Finally, the proposed algorithm is evaluated in term of SNR, Segmental SNR and LLR(Log Likelihood Ratio) and then the proposed method is compared with neural network and composite of neural network and wavelet and wavelet methodes.

The results of this comparison and experimental results show that the proposed method has an acceptable performance.

Important words:

Speech Enhancement, Noise, Wavenet, Wavelet Transform