

بهسازی گفتار مبتنی بر روش ترکیبی تفاضل طیفی و برنامه‌نویسی ژنتیک

سعیده سادات سدیدپور^۱، محمدمهدی همایون‌پور^{۲*}، مهدی فسقوری^۳

۱- کارشناس ارشد مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

۲- دانشیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر

۳- مربی پژوهشکده فناوری اطلاعات، مرکز تحقیقات مخابرات ایران

homayoun@aut.ac.ir

(دریافت مقاله: بهمن ۱۳۸۸، پذیرش مقاله: آبان ۱۳۸۹)

چکیده- در بیشتر محیط‌های زندگی، نویزهای مختلفی حضور دارد که داده‌های صوتی را تخریب می‌کند. در این مقاله روشی را معرفی می‌کنیم که در آن با استفاده از برنامه‌نویسی ژنتیک، نویز افزوده شده به داده‌های صوتی کاهش داده می‌شود تا داده‌های با کیفیت بهتری به دست آید. به این منظور ترکیب دو روش تفاضل طیفی و برنامه‌نویسی ژنتیک برای بهسازی گفتار پیاده‌سازی شده است. در این روش در مرحله اول، نویز به روش تفاضل طیفی کاهش می‌یابد. در گام بعدی، برای بهسازی بیشتر، از روش برنامه‌نویسی ژنتیک استفاده می‌شود. در این گام درخت‌هایی آموزش داده می‌شود که گفتار خروجی الگوریتم تفاضل طیفی را به داده‌های تمیزتری نگاشت می‌کند. برتری روش ترکیبی پیشنهادی در بهسازی گفتار اثبات شده و بهبودی در حدود ۲ تا ۶/۵ دسی‌بل در نسبت سیگنال به نویز حاصل شده است. مقایسه روشهای برنامه‌نویسی ژنتیک، شبکه عصبی، تفاضل طیفی و روش ترکیبی حاضر در بهسازی گفتار نشان می‌دهد که در مجموع، روش ترکیبی تفاضل طیفی - برنامه‌نویسی ژنتیک نسبت به سایر روشها نتایج بسیار بهتری را ارائه می‌کند.

کلید واژگان: برنامه‌نویسی ژنتیک، تفاضل طیفی، گفتار نویزی، بهسازی گفتار.

۱- مقدمه

گرفته است. فرایند بازیابی می‌تواند به شکل بازیابی خود سیگنال گفتار برای افزایش سطح ادراک در سیستم شنوایی انسان یا به شکل بهسازی ویژگی‌های استخراج شده از گفتار برای استفاده در سیستم‌های بازشناسی گفتار یا گوینده مقاوم

بهسازی گفتار به فرایند بازیابی گفتار تمیز از سیگنال نویزی گفتار گفته می‌شود و در مواردی به کار می‌رود که سیگنال گفتار تحت تأثیر نویز، انعکاس یا سایر عوامل تخریب قرار

در برابر نویز باشد. اعوجاج می‌تواند ناشی از نویز جمع‌شونده محیطی یا حاصل از اعوجاج خطی یا غیرخطی در مسیر^۱ انتقال باشد.

ورود نویز به سیستم ممکن است در محل منبع سیگنال، در مسیر مخابره یا در محل گیرنده رخ دهد. عوامل تخریبی وارد شونده در مسأله نیز می‌تواند ترکیبی از این سه حالت باشد [۱]. نویز ممکن است با سیگنال اصلی جمع، ضرب، و حتی در حالت کلی‌تر پیچیده^۲ شود و ممکن است نویز مستقل یا وابسته به سیگنال اصلی باشد.

کاربردهای متنوع بهسازی گفتار، اهداف متفاوتی را برای فرایند بهسازی ایجاد کرده است. افزایش قابلیت درک گفتار یا کاهش خستگی شنونده از جمله اهداف نهایی فرایند بهسازی گفتار هستند.

مسأله مهم دیگری که کیفیت عملکرد سیستم‌های بهسازی گفتار را تحت تأثیر قرار می‌دهد وجود تنوع در نویز است [۲] و [۳]. نویزهایی که داده‌های گفتاری را تحت تأثیر قرار می‌دهند، ویژگی‌های مختلفی دارند. نویز ممکن است ایستان یا غیرایستان باشد. در نویز ایستان، چگالی طیف توان نویز، ثابت است و با زمان تغییر نمی‌کند. در نویز غیرایستان، مشخصات آماری نویز با زمان تغییر می‌کند. بیشتر نویزها در محیط‌های واقعی غیرایستان هستند. به‌عنوان مثال نویز مهمه^۳ غیر ایستان است [۴]. این نویز با سیگنال اصلی همبستگی دارد و برای کاهش آن از روش‌هایی نظیر حذف وفقی انعکاس صدا^۴ (AEC) استفاده می‌شود.

نویز یا سیگنال تداخل‌کننده ممکن است ناشی از همهمه، موسیقی، باد یا ترافیک باشد و لذا طیف توان آن می‌تواند دارای شکل‌های مختلفی باشد. از سوی دیگر

نویزهای مختلف، تأثیر متفاوتی بر شنونده می‌گذارند. برای مثال نویز در محدوده فرکانسی ۴۰۰ هرتز تا ۵ کیلو هرتز می‌تواند قابلیت درک گفتار را کاهش دهد. اما نویزهای فرکانس بالاتر، تأثیر کمتری بر قابلیت فهم گفتار داشته و بیشتر باعث خستگی شنونده می‌شوند.

در ادامه، به مرور کارهای پیشین می‌پردازیم و سپس ساختار سیستم ارائه شده در مقاله بیان و روش تفاضل طیفی و برنامه‌نویسی ژنتیک^۵ (GP) توضیح داده می‌شود. در بخش بعدی مروری کلی بر اجزای تشکیل دهنده سیستم بهسازی گفتار خواهیم داشت. سپس روش پیشنهادی تشریح شده و نتایج استفاده از آن برای بهسازی گفتار، با روش‌هایی مانند تفاضل طیفی و شبکه عصبی مقایسه می‌شود. در پایان درباره نتایج به‌دست آمده بحث شده و روش‌های بهسازی از نظر کارایی و محاسبات مورد نیاز مقایسه خواهند شد. در بخش آخر به نتیجه‌گیری می‌پردازیم.

۲- مرور کارهای پیشین

هدف اصلی سیستم‌های بهسازی گفتار، ارتقای یک یا چند جنبه ادراکی گفتار مانند کیفیت عمومی و قابلیت درک برای ماشین است.

به دلیل اهمیت و کاربردهای متنوع سیستم‌های بهسازی گفتار، روش‌های مختلفی در سه دهه گذشته بدین منظور توسعه یافته است. در سیستم‌های تک‌مسیره، تخمین طیف نویز اهمیت زیادی دارد. از جمله روش‌هایی که برای این منظور استفاده می‌شود، تعیین قاب‌های فاقد گفتار و سپس به‌روز کردن تخمین طیف نویز از روی آنها است. چنانچه اطلاعات پیشینی از گفتار در دست باشد؛ یک راه حذف نویز، کم کردن تخمین طیف نویز به‌دست آمده از قاب‌های فاقد گفتار از سیگنال نویزی در حوزه فرکانس^۶ است.

1. Channel
2. Convolve
3. Babble Noise
4. Adaptive Echo Canceling

5. Genetic Programming
6. Spectral Domain

بعضی از روشهای بهسازی گفتار عبارتند از روش تفاضل طیفی^۱ [۵] و [۶] و استفاده از فیلترهای تطبیقی [۷] مختلفی مانند فیلتر کالمن، وینر و شانه‌ای^۲ به صورت مجزا یا در ترکیب با سایر روشها که به منظور حذف نویز از سیگنال گفتار مورد استفاده قرار گرفته‌اند [۸].

تفاضل طیفی، جزو اولین الگوریتم‌های پیشنهادی برای کاهش نویز [۵] و [۹] و از جمله روشهای موفق مبتنی بر پردازش سیگنال است [۵]. این روش، مبتنی بر اصول ساده‌ای بوده و پیاده‌سازی آن ساده و بار محاسباتی آن اندک است. در این روش، نویز جمع‌شونده فرض می‌شود و از تفاضل طیف نویز تخمین زده شده از گفتار نویزی، تخمینی از طیف سیگنال تمیز به دست می‌آید. اگر مقدار کم شده زیاد باشد، بخش‌هایی از گفتار حذف می‌شود و اگر کم باشد، مقداری نویز باقی می‌ماند. از معایب اصلی این روش - که در سایر روشهای پردازش طیف سیگنال نیز مشاهده می‌شود - ایجاد نویز موزیکال^۳ ناراحت‌کننده در سیگنال بهبودیافته است؛ لذا به منظور بهبود نتایج از روشهای کمکی استفاده می‌شود [۱۰]. به دلیل سادگی پیاده‌سازی، در بیشتر پژوهشها، نتایج روش پیشنهادی با نتایج حاصل از این روش مقایسه می‌شود [۱۱] تا [۱۳].

در کاربردهایی مانند تلفن بانک و ارتباطات تلفنی با مراکز مختلف، امکان اضافه شدن انواع نویز به سیگنال وجود دارد. به ویژه اگر ماشین پاسخ‌دهنده باشد، نویز باعث بروز مشکلات و کاهش کارایی سیستم می‌شود. بیشتر روشهای به کار رفته برای بهسازی سیگنال گفتار محدودیت‌های زیادی دارند که نیاز به میکروفون اضافی یکی از آنها است. برای داده‌هایی که از خطوط تلفن به دست می‌آیند، دسترسی به میکروفون اضافی ممکن نیست. در تبدیل گوینده که یکی از زیرشاخه‌های پردازش گفتار است، از روشهای نگاشت

استفاده می‌شود اما در بهسازی گفتار، تاکنون از نگاشت داده نویزی به داده تمیز استفاده نشده بلکه فقط به منظور یافتن بخش‌های از دست رفته از آن استفاده شده است [۱۴]. فکر استفاده از نگاشت در بهسازی گفتار، به دلیل مشابهت بهسازی با تبدیل گوینده مطرح می‌شود. در هر دو مورد، سیگنال خروجی، شکل تبدیل یافته سیگنال ورودی است. در این مقاله، روش جدیدی بررسی می‌شود که در آن از برنامه‌نویسی ژنتیک که یکی از قدرتمندترین روشهای تخمین‌زننده به شمار می‌رود، به منظور نگاشت داده نویزی به داده تمیز و بهسازی گفتار تلفنی استفاده می‌شود. در روش پیشنهادی بعدی به منظور بهبود بیشتر کیفیت، ترکیب برنامه‌نویسی ژنتیک با روش مرسوم تفاضل طیفی مطرح می‌شود تا بدین ترتیب از مزایای هر دو روش استفاده شود. تفاضل طیفی، سیگنال مناسبی را در نواحی گفتار ایجاد می‌کند و برنامه‌نویسی ژنتیک، تأثیر نویز در نواحی سکوت را به طور قابل ملاحظه‌ای کاهش می‌دهد. این روش ترکیبی، کیفیت سیگنال را بسیار بهبود می‌دهد.

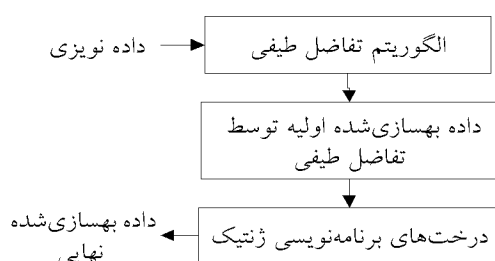
۳- ساختار سیستم

هدف این مقاله، تحقیق و بهبود شیوه‌های کنونی بهسازی گفتار است. پیشنهاد اول، استفاده از برنامه‌نویسی ژنتیک برای بهسازی گفتار است. برنامه‌نویسی ژنتیک نوعی نگاشت از ضرایب طیف فوریه هر قاب از گفتار نویزی را به طیف فوریه گفتار بهسازی شده انجام می‌دهد که این نگاشت توسط درخت‌های GP آموزش دیده برای همین منظور صورت می‌گیرد.

در آموزش درخت‌های GP، طیف فوریه هر فریم داده نویزی، به عنوان ورودی درخت‌های حاصل از برنامه‌نویسی ژنتیک و طیف فوریه داده تمیز به عنوان خروجی آن استفاده می‌شود. تعداد درخت‌هایی که باید بدین منظور آموزش داده شود، به اندازه تعداد ضرایب فوریه قاب تحلیل در نظر گرفته شده است. هر ضریب فوریه سیگنال حاصل از بهسازی، به

1. Spectral Subtraction
2. Comb Filtering
3. Musical Noise

پیشنهاد دیگر به منظور بهبود بیشتر سیگنال بهسازی شده، ترکیب سری دو روش تفاضل طیفی و برنامه نویسی ژنتیک است. در این روش، بهسازی اولیه با الگوریتم تفاضل طیفی صورت گرفته و سپس داده به دست آمده به درخت های برنامه نویسی ژنتیک ارائه می شود. مراحل روش پیشنهادی در شکل ۲ نشان داده شده است. روش ترکیبی، مزایای هر دو روش را داشته و بهبود مناسبی را ایجاد می کند.



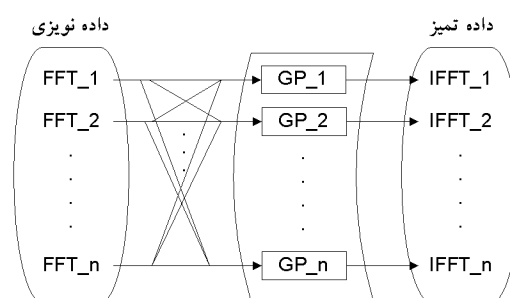
شکل ۲ روش بهسازی پیشنهادی مقاله

۳-۱- تفاضل طیفی

الگوریتم تفاضل طیفی، یکی از قدیمی ترین و ساده ترین روش های بهسازی گفتار است که برای بازیابی طیف توان سیگنال آلوده به نویز جمع شونده استفاده می شود. این روش طیف توان نویز تخمین زده شده را از طیف توان سیگنال نویزی کم کرده و سیگنال بهسازی شده را تولید می کند. در این روش، باید نواحی گفتار و سکوت از هم تشخیص داده شوند. همچنین فرض می شود که نویز در کل نواحی سکوت ایستا باقی می ماند و با سیگنال اصلی ناهمبسته^۲ است.

در الگوریتم تفاضل طیفی، ابتدا طیف اولیه ای از نویز تخمین زده می شود. سپس این تخمین با استفاده

تمامی ضرایب فوریه سیگنال نویزی ورودی وابسته است یعنی هر درخت GP، تمامی نمونه های طیف قاب ورودی را به یک ضریب طیف خروجی می نگارد. شکل ۱، بهسازی گفتار به روش GP را نشان می دهد.



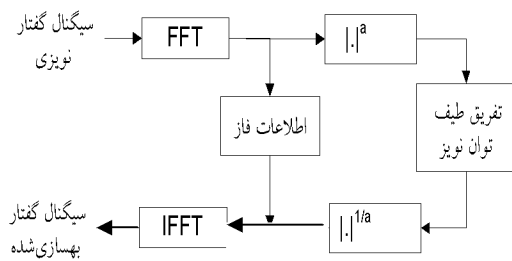
شکل ۱ روش پیشنهادی بهسازی گفتار با استفاده از GP

نوعی تعادل در تعداد نمونه های گفتار قاب ورودی و در نتیجه در تعداد نقاط تبدیل فوریه سریع (FFT) به کار رفته لازم است وجود داشته باشد، زیرا تعداد کم نمونه ها، باعث افت شدید کیفیت سیگنال بهسازی شده می شود و تعداد زیاد نمونه ها نیز، تعداد درخت هایی را که باید آموزش داده شوند و در نتیجه، زمان آموزش را افزایش می دهد.

می توان بردار ویژگی را با طول بیشتری مانند ۱۲۸ انتخاب کرد و با اعمال روش هایی طول بردار را کاهش داد. برای کاهش بعد بردار ویژگی، استفاده از میانگین گیری روی چند نمونه یا استفاده از کاهش فرکانس نمونه برداری^۱ پیشنهاد می شود. در هر دو مورد، پس از اعمال GP بر داده آزمایشی، درونیایی انجام می شود تا تعداد ضرایب طیف و در نتیجه تعداد نمونه های فریم حاصل از بهسازی به اندازه اولیه باز گردد. برای آنکه کیفیت بهسازی افت نکند، طول قاب اولیه کمی کاهش داده شده و زمان بیشتری برای آموزش در نظر گرفته می شود. هر دو روش کاهش نمونه برداری و میانگین گیری، کیفیت داده های بهسازی شده را کاهش می دهند.

2 Uncorrelated

1 Down-Sampling



شکل ۳ روش تفاضل طیفی

۳-۲- برنامه نویسی ژنتیک

برنامه نویسی ژنتیک (GP)، زیرمجموعه‌ای از الگوریتم‌های ژنتیک است که توسط کزا^۱ گسترش یافت [۱۵] و [۱۶]. در این روش اعضای جمعیت، درخت‌هایی (گراف‌هایی) هستند که با استفاده از عملگرهای ژنتیکی بازترکیبی و جهش تولید می‌شوند. هر فرد از جمعیت توسط گره‌های غیرپایانه‌ای یا توابع و گره‌های پایانه‌ای یا برگ‌ها که متغیرها یا ثابت‌ها هستند نشان داده می‌شود [۱۷].

برنامه نویسی ژنتیک را می‌توان در مدل‌سازی معادلات ریاضی و شبیه‌سازی خطوط و منحنی‌های مختلف به کار برد. برنامه نویسی ژنتیک، شامل مراحل زیر است:

(۱) با تولید تصادفی جمعیت راه حل‌های با اندازه N ، کار شروع می‌شود.

(۲) هر راه حل از جمعیت، ارزیابی شده و مقدار شایستگی آن به دست می‌آید.

(۳) یک عملگر ژنتیک به صورت احتمالی انتخاب می‌شود.

- اگر عملگر انتخابی تکثیر باشد، فردی از جمعیت جاری انتخاب و به جمعیت جدید کپی می‌شود. عملگر تکثیر، اصول انتخاب طبیعی و بقای شایسته‌ترین‌ها را تکرار می‌کند.

از دوره‌های زمانی سکوت - که فقط نویز حضور دارد - به‌روز می‌شود. در این روش فرض بر آن است که نویز با سیگنال اصلی جمع شده است. اگر $s(n)$ سیگنال تمیز، $y(n)$ سیگنال نویزی، $w(n)$ نویز جمع‌شونده و $\hat{s}(n)$ سیگنال بهسازی شده باشند، معادله زیر برقرار است:

$$y(n) = s(n) + w(n) \quad (1)$$

با فرض جمع‌شونده و ناهمبسته بودن نویز با سیگنال اصلی، رابطه زیر برقرار است [۵]:

$$Y(\omega) = S(\omega) + W(\omega) \quad (2)$$

که $Y(\omega)$ ، $S(\omega)$ و $W(\omega)$ به ترتیب طیف توان سیگنال نویزی، سیگنال بدون نویز و نویز است. در روش تفاضل طیفی، تخمین اندازه طیف توان نویز، با استفاده از نواحی سکوت و از طریق میانگین‌گیری اندازه طیف توان نویز در پنجره‌های مختلف این نواحی انجام می‌شود.

برای بازیابی سیگنال، طیف دامنه لحظه‌ای با فاز سیگنال نویزی ترکیب شده و با استفاده از تبدیل معکوس فوریه به حوزه زمان بازگردانده می‌شود. یکی از اشکال‌های اصلی روش تفاضل طیفی، تولید نویز باقی‌مانده‌ای به نام نویز موزیکال است. دلیل به وجود آمدن نویز موزیکال، آن است که $|\hat{S}(\omega)|$ به دست آمده ممکن است منفی باشد. بنابراین در بسیاری از سیستم‌های ساده، به جای این مقادیر منفی، از صفر یا مقدار کوچک مثبتی استفاده می‌شود. در نتیجه در سیگنال بهسازی شده، برخی مؤلفه‌های طیف به تناوب خاموش و روشن شده و به صورت نویز موزیکال شنیده می‌شوند. شکل ۳، فرایند پردازشها را در این روش نشان می‌دهد.

1 Koza

روشهای ارزیابی کمی است. برای محاسبه SNR بهبودیافته، از رابطه زیر استفاده می‌شود [۱۸].

$$SNR_{IMP} = SNR_{OUT} - SNR_{IN} \quad (۳)$$

که در آن SNR_{IN} ، SNR_{OUT} و SNR_{IMP} ، به ترتیب نشان‌دهنده سیگنال به نویز سیگنال نویزی، سیگنال به نویز سیگنال بهسازی شده و مقدار بهبود سیگنال به نویز است. SNR_{OUT} و SNR_{IN} از معادلات زیر بدست می‌آیند [۱۸]:

$$SNR_{IN} = 10 \log_{10} \frac{\sum s^*(n)}{\sum (y(n) - s(n))^2} \quad (۴)$$

$$SNR_{OUT} = 10 \log_{10} \frac{\sum s^*(n)}{\sum (\hat{s}(n) - s(n))^2} \quad (۵)$$

که در آن، $s(n)$ ، $y(n)$ و $\hat{s}(n)$ ، به ترتیب سیگنال گفتار بدون نویز، سیگنال گفتار نویزی و سیگنال بهسازی شده است.

۴-۲- میزان بهبود در ارزیابی ادراکی کیفیت گفتار

ویژگی PESQ سعی می‌کند ارزیابی کیفی را بدون نیاز به ارزیابی‌های شنیداری، به صورت کمی بیان کند و نیاز دارد که سیگنال تمیز اولیه را در اختیار داشته باشد. نتیجه ارزیابی PESQ، عددی بین صفر تا ۴/۵ است. هر چه این عدد بالاتر باشد، سیگنال بهسازی شده به سیگنال تمیز اولیه بیشتر شبیه است. امتیاز PESQ، به صورت ترکیبی خطی از میانگین مقدار اعوجاج D_{ind} و میانگین مقادیر اعوجاج نامتقارن A_{ind} به شکل زیر بیان می‌شود:

- اگر عملگر انتخابی بازترکیبی باشد، دو فرد انتخاب می‌شوند. سپس به صورت تصادفی، یک زیر درخت در هر فرد انتخاب و با یکدیگر جابه‌جا می‌شوند. دو فرزند تولید شده، در جمعیت جدید قرار می‌گیرند. بازترکیبی نقشی حیاتی را در فرایند تکامل ایفا می‌کند.

- اگر عملگر انتخاب شده جهش باشد، فردی از جمعیت جاری و سپس زیر درختی از آن انتخاب و عمل جهش بر روی آن انجام می‌شود. جهش، تنوع را حفظ می‌کند.

(۴) مرحله ۳ تا رسیدن جمعیت جدید به N عضو، ادامه می‌یابد.

(۵) به خلاف الگوریتم ژنتیک، برنامه‌نویسی ژنتیک همگرا نمی‌شود. بنابراین، مراحل ۲ تا ۴، تا رسیدن به راه حل مطلوب ادامه می‌یابد و اگر به نتیجه مطلوب دست نیابد، پس از تعداد تکرار معینی خاتمه می‌یابد.

۴- معیار ارزیابی

بر حسب نوع کاربرد سیستم بهسازی گفتار، ممکن است بهبود نظری^۱ یا عینی^۲ سیگنال گفتار مورد توجه باشد. در نتیجه، ارزیابی‌های متنوعی برای بهسازی گفتار مطرح است. در این مقاله از دو روش: میزان بهبود نسبت سیگنال به نویز (SNR) و ارزیابی ادراکی کیفیت گفتار (PESQ) استفاده شده است.

۴-۱- میزان بهبود نسبت سیگنال به نویز

معیار سیگنال به نویز، روشی معمول برای اندازه‌گیری کیفیت گفتار است و همانگونه که از نام آن پیداست نسبت توان سیگنال به توان نویز در واحد دسیبل را می‌سنجد. نسبت سیگنال به نویز بهبودیافته یکی از

1. Subjective
2. Objective

3. Average Disturbance Value

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \quad (6)$$

که در آن $a_0 = 4/5$ ، $a_1 = -0/1$ و $a_2 = -0/0309$ [۱۹].

۵- تفاضل طیفی - برنامه نویسی ژنتیک

برای انجام عملیات بهسازی گفتار، ابتدا فایل صوتی به قاب‌هایی تفکیک می‌شود. در این مرحله، در این مقاله ابتدا بهسازی به روش تفاضل طیفی که روشی مرسوم در بهسازی گفتار است [۲۰] و [۲۱]، انجام می‌شود. اصولاً دو روش تفاضل طیفی وجود دارد: یکی با استفاده از تفریق بزرگی طیف توان نویز از بزرگی طیف توان سیگنال نویزی (روش تفاضل طیفی توان^۱) و دیگری با استفاده از تفریق بزرگی طیف نویز از بزرگی طیف سیگنال نویزی (روش تفاضل طیفی بزرگی^۲). در این مقاله فقط از روش اول استفاده شده است؛ یعنی به منظور بهسازی به روش تفاضل طیفی، توان نویز در بازه‌های سکوت تخمین زده شده و از توان طیف هر قاب از سیگنال نویزی کاسته می‌شود. سپس به همراه فاز اولیه سیگنال نویزی تبدیل فوریه معکوس گرفته شده و سیگنال بهسازی شده به دست می‌آید [۲۰].

در گام بعدی، با توجه به شکل (۲) از برنامه نویسی ژنتیک برای بهسازی بیشتر استفاده می‌شود. در این مرحله، در قسمت آموزش، تبدیل فوریه هر قاب از سیگنالی که در یک مرحله به روش تفاضل طیفی بهسازی شده، محاسبه می‌شود. سپس نتایج تبدیل فوریه از حالت دکارتی به قطبی تبدیل شده و اندازه و فاز بردار حاصل استفاده می‌شود. مقادیر اندازه به دست آمده، برای آموزش به سیستم برنامه نویسی

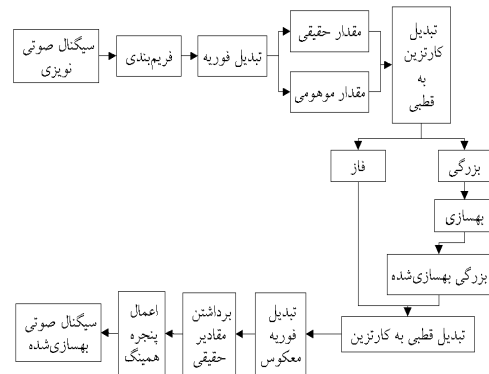
ژنتیک، ارائه و فاز نمونه‌های طیف داده نویزی نگاه داشته می‌شود تا بعداً به عنوان فاز نمونه بهسازی شده استفاده شوند. شکل (۱) چگونگی کار GP را در بهسازی گفتار نشان می‌دهد. وظیفه GP را مطابق شکل، نگاشت داده‌های نویزی به داده‌های تمیز در حوزه فرکانس است.

همان‌طور که در شکل (۴) دیده می‌شود، در مرحله آزمایش، عمل قاب‌بندی و به دست آوردن مقدار حقیقی و موهومی ضرایب تبدیل فوریه و سپس تبدیل از دکارتی به قطبی انجام می‌شود. برای بهبود بیشتر کیفیت بهسازی، مقداری هم‌پوشانی بین قاب‌ها در نظر گرفته می‌شود. بزرگی طیف فوریه، به مدل‌های GP آموزش داده شده ارائه شده و خروجی‌های هر درخت GP حاصل می‌شود. در اصل GP طیف سیگنال نویزی را به طیف سیگنال تمیز نگاشت می‌کند. خروجی‌ها به عنوان بزرگی طیف داده‌های بهسازی شده در نظر گرفته می‌شود. برای بازسازی داده صوتی، بزرگی طیف بهسازی شده همراه با فاز سیگنال نویزی، از قطبی به دکارتی تبدیل می‌شود. در ادامه با تبدیل معکوس فوریه نمونه‌های سیگنال تمیز شده به دست می‌آید. در انتها با توجه به همپوشانی در نظر گرفته شده برای قاب‌های داده نویزی اعمال شده به GP، نوعی همپوشانی و جمع^۳ بر روی نمونه‌های قاب‌های بازسازی شده صورت می‌گیرد. برای این منظور قاب‌های بازسازی شده در پنجره همپینگ ضرب و با در نظر گرفتن همپوشانی، با یکدیگر جمع می‌شوند. نتیجه به دست آمده، فایل صوتی بهسازی شده نهایی است.

1. Power Spectral Subtraction
2. Magnitude Spectral Subtraction

3. Overlap And Add (OLA)

میانگین باشد، مقدار نمونه بزرگی طیف با صفر جایگزین می‌شود. از آنجا که هر مقدار بزرگی طیف توسط یک درخت GP به دست می‌آید، این بدان معناست که درخت به دست آمده، برای نمونه مورد نظر مطلوب نبوده و بهتر است حذف شود. شکل (۶) نمونه‌ای از فایل بهسازی شده به روش GP پس از حذف اثر گفته شده را نشان می‌دهد.



شکل ۴ مراحل بهسازی گفتار به روش برنامه‌نویسی ژنتیک پیشنهادی

۶- ویژگی‌های برنامه‌نویسی ژنتیک

به دلیل آن‌که برنامه‌نویسی ژنتیک، روش مدل‌سازی پایه در این مقاله است، با به دست آوردن بهترین تنظیمات، می‌توان نتایج حاصل را بهبود داده و بهینه کرد. جدول (۱) مقادیر بهینه ویژگی‌های به دست آمده در تولید جمعیت را نشان می‌دهد.

جمعیت اولیه زیاد، امکان دیدن جوابهای بیشتری را ممکن می‌سازد اما مدت زمان محاسبه تابع شایستگی را افزایش می‌دهد؛ بنابراین ۱۰۰۰ به عنوان جمعیت اولیه انتخاب می‌شود. برای دیدن حالت‌های مختلف با این تعداد جمعیت، تعداد نسل بیشتری باید در نظر گرفته شود که ۱۵۰ است.

چهار عمل اصلی حسابی، به عنوان توابع پایه در مجموعه توابع در نظر گرفته شد. برای بهبود نتایج، از لگاریتمها، جذر، سینوس و کسینوس و تانژانت نیز استفاده شد. با توجه به استفاده از توابع سینوسی، بهتر است عدد π نیز بتواند در برگ‌های درخت GP انتخاب شود. محدوده عددی ۰ تا ۱۰، برای پوشش محدوده اعداد مناسب است. بازترکیبی، از نوع «یک نقطه‌ای یکنواخت» در نظر گرفته شده که بیانگر جابه‌جایی والدین از یک نقطه تصادفی در هر یک از آنها است. انتخاب بازماندگان، به

بردارهای خروجی حاصل از درخت‌های GP، به پردازش‌هایی نیاز دارند تا کیفیت فایل‌های بازسازی شده بهبود یابد. یکی از مشکلات نتایج به دست آمده، منفی شدن برخی مقادیر بزرگی طیف حاصل از نگاشت صورت گرفته توسط GP است که موجب پیدایش نویز موزیکال می‌شوند؛ برای حل این مشکل، مقادیر منفی به صفر تبدیل شد. مشکل دیگر در فایل‌های بهسازی شده، ایجاد مقادیری بزرگ برای بزرگی طیف در بعضی فرکانس‌ها در قاب‌های متوالی است که در بیشتر قاب‌ها موجود است. نمونه‌ای از طیف‌نگاره فایل بهسازی شده قبل از حذف این اثر در شکل (۵) نشان داده شده است. نوارهای پررنگ درون شکل به مشکل فوق مربوط هستند. برای حل این مشکل، میانگین هر یک از نمونه‌های بزرگی طیف ظاهر شده در خروجی GP در قاب‌های متوالی محاسبه شده و در صورتی که در حداقل ۸۰٪ مواقع دامنه یک نمونه از طیف در قاب‌های متوالی از میانگین آن نمونه بزرگتر باشد یا در نمونه‌های دو سوم انتهایی فریم‌ها، اگر مجموع اختلاف نمونه‌ها از میانگین به دست آمده، مثبت و مقدار نمونه بزرگتر از نصف

۶-۱- جهش

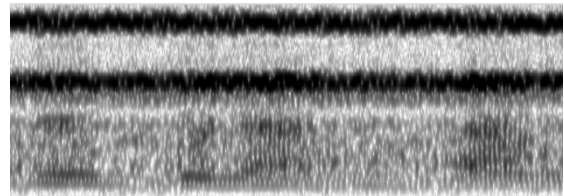
در این مقاله، از چند نوع جهش استفاده شده که به تفصیل بررسی می‌شوند. در جهش، نخست عددی تصادفی تولید می‌شود که انجام یا عدم انجام جهش را با توجه به مقایسه با نرخ جهش تعیین می‌کند. سپس، در صورت انجام جهش، محل جهش به طور تصادفی انتخاب می‌شود.

جدول ۱ مقادیر ویژگی‌های تولید جمعیت

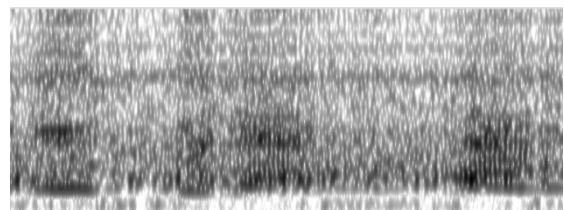
ویژگی GP	مقدار ویژگی
جمعیت اولیه	۱۰۰۰
تعداد نسل	۱۵۰
مجموعه توابع	$\{+, -, *, /, \sqrt{\cdot}, \log, \log_{10}, \sin, \cos, \tan\}$
توابع مورد استفاده در برگ	π
محدوده عدد مورد استفاده در پایانه	۱۰-۰
روش بازترکیبی	یک نقطه‌ای یکنواخت
روش انتخاب	مسابقات دوره‌ای
اندازه تور در مسابقات دوره‌ای	۳

جهش مقدار ساده^۲: اگر محل انتخاب شده برای جهش، از برگ‌ها و پایانه‌هایی باشد که مقدار عددی دارد، می‌تواند با مقدار عددی دیگری که به طور تصادفی انتخاب می‌شود، جایگزین شود. انتخاب مقادیر عددی ثابت، یکی از مهمترین مسائل در آموزش مدل GP است که برای شایسته‌تر کردن مدل‌ها در هر مسئله، به کار می‌روند. برای یافتن اعداد جایگزین و به‌منظور پیدا کردن مقادیر بهتر، از راهبرد تکاملی یا ES استفاده شده است. به‌دلیل زمانبر بودن راهبرد تکاملی، به‌عنوان جهش در نظر گرفته شده تا ضمن استفاده از بهبود ایجاد شده توسط آن، از افزایش بسیار زیاد زمان آموزش مدل‌ها جلوگیری شود.

روش مسابقات دوره‌ای^۱ با اندازه تور برابر سه صورت می‌گیرد که به معنای انتخاب تصادفی سه فرد و برگزیدن یکی از آنها است.



شکل ۵ طیف‌نگاره داده بهسازی‌شده قبل از حذف اثر مؤلفه‌های طیفی با مقادیر بزرگ



شکل ۶ طیف‌نگاره داده بهسازی‌شده با استفاده از روش GP پس از حذف اثر مؤلفه‌های طیفی با مقادیر بزرگ

نخبگان و بازماندگان افرادی هستند که بدون شرکت در بازترکیبی یا جهش، به نسل بعدی منتقل می‌شوند. با توجه به انتخاب یک فرد به‌عنوان بهترین فرد نهایی، تعیین ۲۰ نخبه که بهترین‌های نسل هستند، مناسب است. بازماندگان، لزوماً بهترین‌های نسل نیستند و از بین افراد انتخاب شده به اندازه تور، انتخاب می‌شوند. جهش قوی، سه بار اعمال جهش است. مجموع نرخ جهش و جهش قوی زیاد در نظر گرفته شده تا کاهش جمعیت را جبران کند و افراد با حالت‌های متنوع‌تری دیده شوند. به‌خلاف برنامه‌نویسی ژنتیک پایه، تعدادی افراد جدید در هر نسل تولید می‌شوند تا افراد نامناسب‌تر هر نسل حذف شده و افرادی ایجاد شوند که شاید بتوانند جزو بهترین‌ها باشند.

جهش جایگزینی^۱: در این روش، اگر درخت به حد اشباع نرسیده باشد، می‌توان گره را با درختی تصادفی جایگزین کرد. گره انتخاب شده برای جایگزینی، همان محل جهش است که به‌طور تصادفی انتخاب شده. زیردرخت برای جایگزینی نیز به‌طور تصادفی تولید می‌شود. این جهش به‌منظور تکامل افراد انجام می‌شود. جهش جایگشت^۲: در این روش، عملوندهای گره که تابع است، جایگزین یکدیگر می‌شوند. این روش لزوماً بر روی تابعی انجام می‌شود که بیش از یک عملوند داشته باشد. در تابعی که بیش از دو عملوند دارند، جابه‌جایی عملوندها به‌طور تصادفی تعیین می‌شود. از این جهش به‌منظور بهینه‌سازی مدل‌ها استفاده می‌شود.

جهش ویرایش: این جهش، به کل ژنوم اعمال شده و اطلاعات اضافی و افزونه ژنوم را هرس می‌کند. از میان چند روشی که برای ویرایش به کار رفته، یک نمونه با توجه به بررسی گره و نیاز آن انتخاب می‌شود. اگر گره پایانه باشد، عملی انجام نمی‌شود و اگر تابع با عملوندهای عدد صحیح یا اعشاری باشد، تابع به عملوندها اعمال شده و مقدار نتیجه، جایگزین زیردرخت موردنظر می‌شود. اگر گره تابع ضرب یا تقسیم با عملوند صفر باشد که پاسخ را صفر می‌کند، زیردرخت با گره پایانی آن زیردرخت جایگزین می‌شود. اگر تابع جمع یا تفریق با عملوند صفر باشد که تأثیری در نتیجه ندارد، زیردرخت با ریشه عملوند غیرصفر، جایگزین درخت با ریشه گره تابع موردنظر خواهد شد. مشکل چاقی یا رشد بیش از حد درخت‌ها، یکی از مشکلاتی است که برنامه‌نویسی ژنتیک با آن درگیر است و یکی از روشهای پیشنهادی برای کنترل آن این است که ابتدا درخت هرس شده و شاخ و

برگ‌های زاید آن حذف شود. در نتیجه، این جهش به قصد رفع چاقی بوده - و به دلیل اینکه زمان آموزش را افزایش داده و نیز ممکن است در بازترکیبی‌های مختلف، بخشهایی که با مدل‌های دیگر جابه‌جا می‌شوند مفید باشند - به‌صورت جهش در نظر گرفته شده و برای تمامی افراد اعمال نمی‌شود.

جهش معاوضه^۳: در این روش، هر گره با گرهی با تعداد ویژگی مشابه تعویض می‌شود؛ چنانچه گره پایانه باشد، با پایانه‌ای که می‌تواند عدد صحیح، اعشاری، توابع بدون عملوند مانند عدد π یا ویژگی باشد، تعویض می‌شود؛ چنانچه گره یکی از توابع باشد، تابعی با تعداد عملوند مشابه، به‌طور تصادفی تولید شده و با گره غیرپایانه‌ای محل جهش، جایگزین می‌شود. این جهش نیز برای بهبود دقت مدل آموزشی استفاده می‌شود.

جهش تلنگری^۴: در این روش، دو گره مجاور که تعداد ویژگی‌های مشابهی دارند، تعویض می‌شوند. چنانچه گره‌های مجاور، پایانه‌ای (که می‌توانند عدد صحیح، اعشاری، توابع بدون عملوند مانند عدد π یا ویژگی) باشند، با یکدیگر جابه‌جا می‌شوند. چنانچه گره‌های مجاور یکی از توابع باشند، در صورتی که تعداد عملوندهای یکسانی داشته باشند، با هم جایگزین شده و در اصل ویژگی‌های دو تابع با هم تعویض می‌شوند. از این جهش به‌منظور انتخاب بهینه‌تر ویژگی و توابع استفاده می‌شود.

۶-۲- دادگان

دادگان مورد استفاده، از دادگان گفتاری TPersianDat استخراج شده است. این دادگان در آزمایشگاه سیستم‌های

3. Swap
4. Flip

1. Substitution
2. Permutation

هوشمند صوتی-گفتاری دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر، در طی دو سال ضبط شده و حاوی گفتار بیش از ۱۲۰ گوینده است که هر گوینده، در پنج جلسه با سیستم تلفنی ضبط صدا تماس گرفته و در جلسات ضبط شرکت کرده است. فرکانس نمونه برداری ۸KHz است و نمونه ها ۱۶ بیتی است، که از فایل های PCM لگاریتمی با استاندارد A-Law به دست آمده.

سیستم بهسازی گفتار توسط GP، برای ۱۰ گوینده مرد آموزش داده شده است. به منظور آموزش درخت های بهسازی گفتار، از دو جلسه از گفتار هر یک از گویندگان، عبارت «روح بلند پرواز» استفاده شده است. استفاده از جملات ثابت به این دلیل است که هدف نهایی بهسازی، استفاده از گفتار بهسازی شده در تصدیق هویت گوینده وابسته به متن بوده است. در تصدیق هویت وابسته به متن، از جملات یا عبارتهایی ثابت برای انجام تصدیق هویت استفاده می شود. در مرحله آزمایش از سه جلسه دیگر ۱۰ گوینده با بیان همان عبارت ها استفاده شده است. قاب های استخراج شده طول ۶۴ دارند. در مرحله آموزش، قاب ها همپوشانی نداشته و دلیل این کار این است که زمان آموزش کوتاه تر شده و GP نیز هر یک از نمونه ها را یک بار دیده باشد. اما قاب های مرحله آزمایش، به اندازه ۰/۷ طول فریم، همپوشانی دارند.

نویز شبیه سازی ها، از پایگاه داده نویز NoiseX انتخاب شده و از میان نویز های موجود، سه نوع نویز سفید، کارخانه و نویز مهمه، به صورت جمع شونده با سیگنال گفتار ورودی، ترکیب شده است.

۷- نتایج و تحلیل

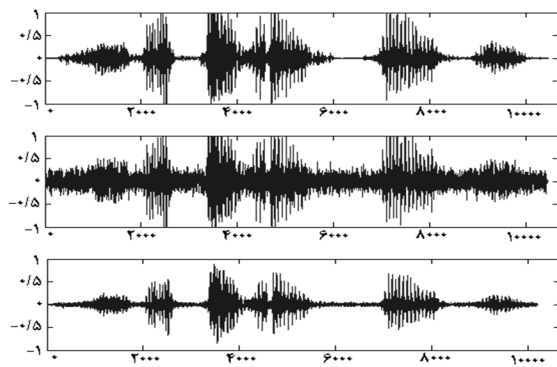
برای هر یک از ۱۰ گوینده، ۶۴ درخت GP به طور جداگانه آموزش داده می شود. داده آموزشی در سمت

ورودی درخت GP، به صورت داده آغشته به نویز سفید با SNR های برابر صفر، ۱۰ و ۲۰ بوده و در سمت خروجی، داده تمیز متناظر با هر یک از داده های ورودی قرار داده شده است. دلیل انتخاب نویز سفید برای آموزش آن است که طیف نویز سفید در تمامی فرکانس ها یکنواخت بوده و طیف نویز های دیگر زیر مجموعه هایی از پوشش این طیف را دارند و در ضمن بدترین نتایج در آزمایش های انجام شده از نتایج نویز سفید حاصل شده است.

برای ارزیابی کارایی روش ترکیبی پیشنهادی، نتایج به دست آمده با نتایج حاصل از دو روش پایه یعنی تفاضل طیفی و برنامه نویسی ژنتیک مقایسه شده است. در ضمن نتایج به دست آمده با نتایج روش شبکه عصبی نیز مقایسه شده است. علت انتخاب شبکه عصبی آن است که عملکردی مشابه روش برنامه نویسی ژنتیک داشته و منحنی را آموزش می دهد که داده های ورودی را به داده های خروجی نگاشت می کند و کارایی آنها با یکدیگر قابل مقایسه است. شبکه عصبی مورد استفاده، نوعی شبکه عصبی سه لایه پرسپترون است که تعداد نرون های لایه میانی پنج برابر نرون های لایه ورودی هستند. شبکه ۱۰۰۰ مرحله به روز می شود تا بهترین نتایج حاصل شود.

آزمایش بر روی داده های نویزی با ۴ نسبت سیگنال به نویز متفاوت ۰، ۵، ۱۰ و ۲۰ برای نویز سفید، مهمه و کارخانه انجام شده است. ارزیابی نتایج بهسازی گفتار، با معیارهای ارزیابی SNR و PESQ انجام گردید.

نخست اثر حذف مؤلفه های طیفی با مقادیر بزرگ بر میزان بهبود SNR و PESQ بررسی شد. نتایج بررسی در شکل (۷) و شکل (۸) نشان داده شده است. این شکل ها به ترتیب میزان بهبود SNR و PESQ را قبل و پس از حذف اثر مؤلفه های طیفی با مقادیر بزرگ در روش GP نشان می دهند.

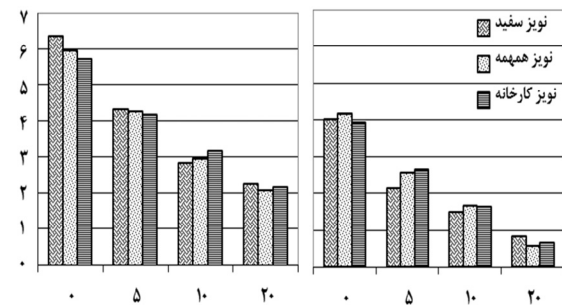


شکل ۹ از بالا به پایین: گفتار تمیز، نویزی و بهسازی شده توسط روش ترکیبی

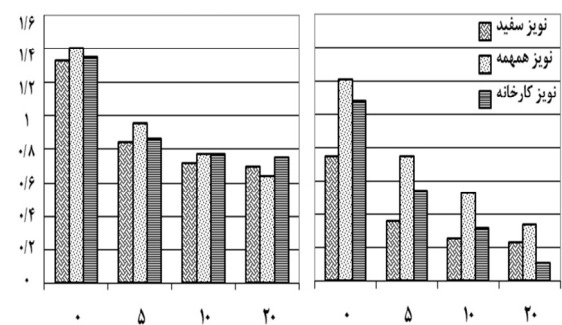
به منظور ارزیابی کارایی روش پیشنهادی ترکیبی، بهسازی گفتار با استفاده از روشهای شبکه عصبی و تفاضل طیفی، برنامه نویسی ژنتیک و روش ترکیبی انجام و نتایج آنها در جدول (۲) و جدول (۳) با یکدیگر مقایسه شده است. نتایج نشان می دهد که با وجود موفقیت روشهای برنامه نویسی ژنتیک و تفاضل طیفی، روش شبکه عصبی نه تنها موفقیت چندانی به دست نیاورده بلکه در بسیاری از موارد موجب تخریب بیشتر سیگنال شده است. مشاهده می شود که نتایج بهبود SNR توسط شبکه عصبی بسیار نامطلوب است. نتایج نشان می دهند که برنامه نویسی ژنتیک در نویزهای همهمه و کارخانه مطلوبتر عمل کرده است اما روش پیشنهادی ترکیبی، در تمامی نویزها بسیار بهتر از روشهای دیگر عمل کرده و نتایج مطلوبتری را ارائه می کند.

برای مقایسه کیفیت سیگنال بهسازی شده از نظر شنیداری، نتایج PESQ بهبود یافته نیز بررسی شده است. دیده می شود که برنامه نویسی ژنتیک به تنهایی نیز برتر از روشهای شبکه عصبی و تفاضل طیفی عمل کرده است. اما روش ترکیبی، نتایجی بسیار مطلوبتری را ارائه کرده که نشان دهنده کارایی و عملکرد این روش پیشنهادی است.

نتایج نشان می دهد که این مؤلفه های بزرگ اثر نامطلوبی بر نتایج می گذارند و حذف آنها نقش به سزایی در بهبود نتایج دارد.



شکل ۷ میزان بهبود SNR داده نویزی با نویزهای مختلف به ترتیب از راست به چپ: قبل و پس از حذف اثر مؤلفه های طیفی با مقادیر بزرگ در روش GP



شکل ۸ میزان بهبود PESQ داده نویزی با نویزهای مختلف به ترتیب از راست به چپ: قبل و پس از حذف اثر مؤلفه های طیفی با مقادیر بزرگ در روش GP

شکل (۹) گفتار تمیز، گفتار نویزی و گفتار بهسازی شده به روش ترکیبی تفاضل طیفی - برنامه نویسی ژنتیک را نشان می دهد. دیده می شود که داده بهسازی شده کاملاً شبیه به داده تمیز بوده و به خوبی نویز از داده نویزی حذف شده و فقط اثر اندکی از نویز بر داده بهسازی شده باقی مانده است.

یکی از برتری‌های روش تفاضل طیفی، محاسبات کم مورد نیاز در آن است. در این مقاله علاوه بر بررسی میزان بهبود در بهسازی گفتار توسط روشهای مورد بررسی، زمان اجرا نیز به عنوان معیاری برای انتخاب روش بررسی شده است. زمان اجرا روی کامپیوتر شخصی با میکروپروسسور پنتیم دو هسته‌ای با فرکانس ۲ گیگاهرتز، حافظه RAM ۱ گیگابایت و کاشه ۴ مگابایت برای بهسازی گفتار «روح بلند پرواز» که در حدود ۲ ثانیه به طول می‌انجامد در روشهای

مختلف در جدول (۴) ارائه شده است. در اینجا با توجه به اینکه طول هر قاب گفتار شامل ۶۴ نمونه در نظر گرفته شده و قاب‌ها به اندازه دو - سوم همپوشانی دارند و جابه‌جایی فریم‌ها به اندازه یک - سوم طول فریم است، لذا با توجه به فرکانس نمونه‌برداری ۸۰۰۰ هرتز، تعداد قاب‌ها در ۲ ثانیه گفتار در حدود ۷۵۰ قاب خواهد بود. لذا مقادیر ارائه شده در جدول (۴) زمان بهسازی برای ۷۵۰ قاب هر یک به طول ۶۴ نمونه است.

جدول ۲ میزان بهبود SNR و PESQ داده نویزی با نویزهای مختلف پس از بهسازی توسط شبکه عصبی و تفاضل طیفی

شبکه عصبی												SNR
نویز سفید						نویز کارخانه		نویز همهمه		نویز سفید		
نویز کارخانه		نویز همهمه		نویز سفید		نویز کارخانه		نویز همهمه		نویز سفید		
بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	
PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	
۰	-۰/۹۳	۰/۵۰	-۱/۳۷	۰/۱۸	-۱/۷۰	۰/۷۳	۴/۷۰	۰/۴۴	۴/۰۲	۰/۵۵	۳/۴۸	۰/۵۲
۵	-۲/۶۷	۰/۳۲	-۳/۲۸	۰/۰۷	-۳/۲۱	۰/۳۷	۲/۸۵	۰/۴۴	۲/۶۱	۰/۵۴	۲/۴۸	۰/۵۱
۱۰	-۴/۰۵	۰/۱۸	-۴/۲۲	۰/۰۷	-۴/۰۸	۰/۳۱	۱/۶۳	۰/۴۷	۱/۵۵	۰/۳۷	۱/۵۹	۰/۴۱
۲۰	-۴/۹۱	۰/۰۷	-۴/۸۱	۰/۰۳	-۴/۹۲	۰/۱۰	۰/۶۵	۰/۲۵	۰/۶۶	۰/۱۳	۰/۶۶	۰/۱۸

جدول ۳ میزان بهبود SNR و PESQ داده نویزی با نویزهای مختلف پس از بهسازی توسط برنامه نویسی ژنتیک و روش ترکیبی

برنامه نویسی ژنتیک												SNR
نویز سفید						نویز کارخانه		نویز همهمه		نویز سفید		
بهبود در مقدار		بهبود در ویژگی		بهبود در مقدار		بهبود در ویژگی		بهبود در مقدار		بهبود در ویژگی		
PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	PESQ	SNR	
بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	بهبود در مقدار	بهبود در ویژگی	
۴/۳۰	۰/۹۵	۴/۳۴	۰/۳۷	۴/۱۲	۰/۷۶	۶/۳۴	۱/۳۳	۵/۹۲	۱/۴۱	۵/۶۹	۱/۳۵	۰
۲/۶۲	۰/۸۵	۲/۷۰	۰/۲۷	۲/۰۸	۰/۶۳	۴/۳۱	۰/۸۵	۴/۲۶	۰/۹۶	۴/۱۵	۰/۸۷	۵
۱/۳۵	۰/۷۲	۱/۶۰	۰/۱۵	۱/۶۰	۰/۴۳	۲/۸۲	۰/۷۲	۲/۹۳	۰/۷۷	۳/۱۶	۰/۷۷	۱۰
۰/۷۴	۰/۳۵	۰/۹۱	۰/۰۱	۱/۰۷	۰/۱۶	۲/۲۲	۰/۷۰	۲/۰۶	۰/۶۴	۲/۱۴	۰/۷۵	۲۰

پیشنهاد شده است. برای این منظور، درخت‌هایی در روش برنامه‌نویسی ژنتیک GP آموزش داده می‌شود که داده‌های نویزی را به داده‌های تمیز نگاشت می‌کند. در این روش پس از آموزش درخت‌های GP، هر ضریب طیف فوریه قاب سیگنال به یک درخت GP داده می‌شود تا آن درخت ضریب فوریه بهسازی شده معادل را تخمین بزند. مجموعه ضرایب فوریه بهسازی شده با تبدیل معکوس فوریه به قاب گفتاری تبدیل و با جمع همپوشان قاب‌های متوالی، سیگنال گفتار بازسازی می‌شود.

در گام بعدی و به‌منظور بهبود بیشتر سیگنال بهسازی‌شده، از ترکیب دو روش تفاضل طیفی و برنامه‌نویسی ژنتیک استفاده شد. سیگنال نویزی در ابتدا توسط الگوریتم تفاضل طیفی بهسازی می‌شود و سپس خروجی به‌دست آمده به درخت‌های برنامه‌نویسی ژنتیک ارائه شده و بهسازی نهایی بر روی آن صورت می‌گیرد.

روش برنامه‌نویسی ژنتیک در مقایسه با شبکه عصبی و روش تفاضل طیفی، کارایی بهتری داشته و حدود $4/3 - 0/7$ دسی‌بل بهبود را در SNR نشان می‌دهد. بهبود نسبت سیگنال به نویز به روش ترکیبی، از ۲ تا $6/5$ دسی‌بل و بسیار قابل ملاحظه است. مقایسه سه روش برنامه‌نویسی ژنتیک، شبکه عصبی و تفاضل طیفی با روش ترکیبی در بهسازی گفتار نشان داد که در مجموع روش تفاضل طیفی - برنامه‌نویسی ژنتیک نسبت به سایر روشها در کیفیت سیگنال بهسازی‌شده نتایج بهتری ارائه می‌کند.

در ادامه این تحقیق پیشنهاد می‌شود که روشهایی برای برطرف ساختن خطاهای ایجاد شده توسط برنامه‌نویسی ژنتیک بررسی شود. در ضمن، می‌توان ترکیب سایر روشها با برنامه‌نویسی ژنتیک را نیز بررسی کرده و ترکیب‌های بهتری به‌دست آورد.

در روش برنامه‌نویسی ژنتیک به کار رفته در بهسازی گفتار، به ازای هر قاب، ۶۴ مدل که هر یک در حقیقت یک معادله ساده ریاضی است، ارزیابی شده و فریم بهسازی شده به‌دست می‌آید. در صورتی که در روش تفاضل طیفی، به ازای هر قاب گفتار، باید طیف توان نویز محاسبه شده و از طیف توان سیگنال نویزی کم شود. با توجه به محاسبات مورد نیاز و زمان اجرای به‌دست آمده در جدول (۴) زمان مورد نیاز در روش بهسازی با برنامه‌نویسی ژنتیک قابل مقایسه و حتی کمتر از روش تفاضل طیفی بوده و در روش ترکیبی نیز بار محاسباتی افزایش قابل توجهی ندارد و بهسازی در زمانی کمتر از زمان فایل گفتاری که هدف بهسازی آن است صورت می‌گیرد.

جدول ۴ زمان اجرا برای بهسازی گفتار در روشهای مختلف بر

حسب ثانیه

روش بهسازی				زمان اجرا
برنامه‌نویسی ژنتیک	تفاضل طیفی	شبکه عصبی	ترکیبی	
۰/۳۳	۰/۴۰	۰/۵۱	۰/۶۲	

این بررسی نشان می‌دهد که روش پیشنهادی، در کنار بهبود نتایج از نظر زمان و حجم محاسبات مورد نیاز نیز بر سایر روشها برتری دارد.

۸- نتیجه‌گیری و پیشنهادها

برنامه‌نویسی ژنتیک، یکی از قدرتمندترین الگوریتم‌های تکاملی و روشهای پردازش هوشمند است که توانایی زیادی نسبت به بیشتر روشهای یادگیری ماشین دارد و عملکرد آن با شبکه عصبی مقایسه می‌شود. در این مقاله فکر استفاده از برنامه‌نویسی ژنتیک برای بهسازی گفتار

۹- منابع

- [8] H. Ding, et al., "A spectral filtering method based on hybrid wiener filters for speech enhancement," Elsevier, Speech Communication, vol. 51, pp. 259-267, 2009.
- [9] M. Weiss, et al., "Study and the development of the INTEL technique for improving speech intelligibility," Nicolet Scientific Corporation, 1974.
- [10] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in IEEE International. Conf. on Acoustics, Speech, and Signal Processing, 2002.
- [11] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," Elsevier, Speech Communication, vol. 50, pp. 453-466, 2008.
- [12] N. Yoma, et al., "Improving performance of spectral subtraction in speech recognition using a model for additive noise," IEEE Trans. Speech Audio Processing, vol. 6, pp. 579-582, 1998.
- [13] N. Evans, et al., "An assessment on the fundamental limitations of spectral subtraction," in Proc. IEEE Internat. Conf. on Acoustics, Speech, Signal Processing, 2006, pp. 145-148.
- [14] E. Zavarehei, et al., "Noisy Speech Enhancement Using Harmonic-Noise
- [1] Y. Ephraim, "Statistical- Model-Based Speech Enhancement Systems," Proceedings of the IEEE, vol. 80, pp. 1526-1555, 1992.
- [2] L. Cheng-Yuan and J. Jang, "New Refinement Schemes For Voice Conversion," in Proc. ICME, 2003, pp. 725-728.
- [3] B. Gillett and S. King, "Transforming Voice Quality," in Eurospeech Conf., Geneva, 2003, pp. 1713-1716.
- [4] B. Rivet, et al., "Mixing Audiovisual Speech Processing And Blind Source Separation For The Extraction of Speech Signals From Convolutional Mixtures," IEEE Trans. on Audio, Speech, And Language Processing, 2006.
- [5] S. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans. on Acoust., Speech, And Signal Processing, vol. 27, pp. 113-120, 1979.
- [6] P. Loizou, *Speech Enhancement: Theory and Practice*, ed: CRC Press LLC, Boca Raton, FL, 2007.
- [7] L. Deng and D. O'shaughnessy, *Speech Processing*: Marcel Dekker Inc, 2003.

- [19] Y. Hu and P. C. Loizou, "Evaluation of Objective Quality Measures For Speech Enhancement," IEEE Trans. on Audio, Speech, And Language Processing, vol. 16, 2008.
- [20] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech and Audio Processing, vol. 9, pp. 504-512, 2001.
- [21] VOICEBOX is a MATLAB toolbox for speech processing. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- Model and Codebook-Based Post-Processing," IEEE Trans. on Audio, Speech, And Language Processing, vol. 15, pp. 1194-1203, 2007.
- [15] N. Cramer, "A Representation For The Adaptive Generation of Simple Sequential Programs," in In J. Grefenstette (Ed.), Proceedings of the 1st Internat. Conference on Genetic Algorithms, 1985, pp. 183-187.
- [16] J. Koza, *Genetic Programming: on The Programming of Computers By Means of Natural Selection* MIT Press, 1992.
- [17] G. Liai, et al., "A New Codebook Design Method Based on Genetic Programming," IEEE, ICEMI, vol. 3, pp. 250-253, 2007.
- [18] W. R. Wu and P. C. Chen, "Subband Kalman Filtering for Speech Enhancement," IEEE Trans. on Circuit and Systems, vol. 45, pp. 1072-1083, 1998.