

بررسی اثر باندهای بحرانی و پارامترهای بکارگیری تبدیل ویولت در بهبود گفتار

بابک ناصرشریف
دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران
nasser_s@iust.ac.ir

احمد اکبری
دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران
Akbari@iust.ac.ir

بهزاد زمانی دهکردی
دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران
bzamani@comp.iust.ac.ir

اطلاعات مهم سیگنال اصلی، نویز را نیز محدود کرد. در این راستا، برای تعیین مقدار حد آستانه حذف ضرایب، Donoho و Johnston یک روش عمومی آستانه‌گذاری برای حذف نویز سفید گوسی پیشنهاد داده‌اند که برای نویزهای رنگی به خوبی عمل نمی‌کند. [2، 3]. شیخ زاده و ابوطالبی در سال ۲۰۰۱ روشی بهبود یافته پیشنهاد داده‌اند که در آن سیگنال به فریمهای واگذار و بی‌واک تقسیم‌بندی می‌شود و حد آستانه در باندهای بالائی برای فریم واگذار افزایش و برای فریم بی‌واک کاهش می‌یابد [4]. به علت عملکرد ضعیف روشهای آستانه در نویزهای رنگی، Wang و Lu روش آستانه‌گیری در باندهای بحرانی درخت تجزیه ویولت را ارائه دادند. ابتدا آستانه ماسک نویز برای یک فریم تخمین زده شده و سپس این آستانه برای تنظیم آستانه نهائی بکار گرفته می‌شود که اندازه آستانه بطور اتوماتیک در باندهای بالا در طول یک سگمنت واگذار و در باندهای پائین در طول یک سگمنت بی‌واک افزایش می‌یابد [5].

بعضاً برای حل مشکل روشهای حد آستانه، استفاده از فیلترهایی نظیر وینر در حوزه تبدیل ویولت پیشنهاد شده است که پارامترهای فیلتر وابسته به شرایط آماری ضرایب تبدیل ویولت تنظیم می‌گردد [6]. در کار حاضر با این رویکرد و همچنین نزدیکی به عملکرد گوش، فیلتر وینر در درخت هرس شده تبدیل ویولت بسته‌ای بکار گرفته شده است. از سوی دیگر، در روشهای حذف نویز مبتنی بر تبدیل ویولت، انتخاب مقادیر پارامترهای تبدیل ویولت به عنوان یک بانک فیلتر از جمله مسائل مهم در ارتقاء کیفیت سیگنال نهائی می‌باشد. به همین دلیل در این مقاله تعیین مقادیر بهینه برای پارامترهای بانک فیلتر ویولت نیز مد نظر قرار گرفته است.

در ادامه بخش دوم به معرفی تبدیل ویولت و انواع درخت تجزیه ویولت اختصاص دارد. بخش سوم به نحوه محاسبه فیلتر وینر در حوزه ویولت می‌پردازد. در بخش چهارم به ارزیابی پارامترهای بانک فیلتر در درخت ویولت بسته‌ای پرداخته شده است. سپس در بخش پنجم ارزیابی نتایج حاصل از اثر درخت پیشنهادی Mel در حذف نویز گزارش شده است. در انتها نیز نتیجه‌گیری کلی ارائه گردیده است.

۲- تبدیل ویولت

تبدیل ویولت توسط Grossman و Morlet پایه‌گذاری شده است [7]. دسته‌ای از توابع پایه ویولت را می‌توان با انتقال و تغییر مقیاس ویولت بدست آورد:

چکیده: تبدیل ویولت رزولوشن زمان-فرکانسی مناسبی برای مدل کردن سیگنالهای نالیستانی همچون سیگنال گفتار فراهم می‌آورد. برای مدل کردن دقیقتر سیگنال گفتار می‌توان از نزدیک کردن عملکرد تبدیل ویولت به عملکرد گوش بهره جست. به همین منظور در مقاله حاضر استفاده از درخت تجزیه ویولت مبتنی بر بانک فیلتر Mel و نیز فیلتر وینر پیشنهاد شده است. به علاوه، ارزیابی پارامترهای مؤثر تبدیل ویولت در بهبود گفتار نظیر طول فریم، تعداد زیرباندها، همپوشانی و تابع پایه ویولت نیز مورد نظر قرار گرفته است تا بتوان به کیفیت بالاتری در گفتار بهبود یافته دست یافت. نتایج حاصل از آزمونهای نظری و شنوایی نشان داده است که روش پیشنهادی ترکیب درخت Mel و فیلتر وینر نسبت به روشهای قراردادی نظیر درخت ویولت بسته‌ای، در بهبود گفتار بهتر عمل می‌کند.

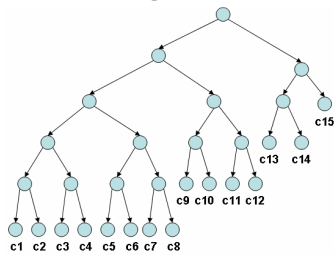
واژه های کلیدی: تبدیل ویولت، فیلتر وینر، درخت ویولت بسته‌ای

۱- مقدمه

استفاده از بهبود گفتار، در بسیاری از کاربردها از جمله تلفن‌های بدون گوشی که صدا در یک محیط نویزی دریافت می‌گردد، ضروری است. برای حذف نویز از سیگنال گفتار، می‌توان طیف فرکانسی سیگنال نویزی را به گونه ای اصلاح نمود که پس از بازیابی سیگنال در حوزه زمان، صداهای پس زمینه گفتار کاهش یابد. از جمله شناخته ترین روشهای این زمینه می‌توان به روش تفاضل طیف توان و استفاده از فیلترهای طیفی نظیر فیلتر وینر اشاره نمود [1]. از جمله اثرات نامطلوب این شیوه ایجاد اعوجاج در سیگنال گفتار و ایجاد نویز موسیقی می‌باشد. همچنین، بنا به ماهیت تبدیل فوری و نالیستانی بودن سیگنال گفتار، عملیات باید بر روی طیف دوره کوتاه سیگنال صورت گیرد که محدودیتهایی در تنظیم رزولوشن زمانی و فرکانسی ایجاد می‌نماید.

از دیگر روشهای بهبود گفتار می‌توان روشهای مبتنی بر ویولت را نام برد. ویژگی اصلی تبدیل ویولت بهره جستن از پنجره های زمانی با طول متفاوت برای باندهای فرکانسی مختلف است. روش آستانه‌گذاری ضرایب تبدیل ویولت متداولترین روش حذف نویز مبتنی بر ویولت است. در سیگنال گفتار، انرژی سیگنال غالباً در تعداد کمی از ضرایب تبدیل ویولت متمرکز است. این ضرایب نسبت به ضرایب سیگنال نویز که انرژی آن بر روی تعداد زیادی از ضرایب گسترش می‌یابد، بزرگتر هستند. بنابراین می‌توان با حذف ضرایب کوچکتر، علاوه بر حفظ

نشان داده شده است. جدول (۱) پهنای فرکانسی هر باند را بر حسب هرتز نشان می‌دهد. چون در گسترش درخت ویولت از تجزیه دوتائی استفاده شده، لذا درخت حاصل تقریبی از فیلتربانک Mel می‌باشد.



شکل ۵- درخت ویولت بر اساس فیلتر بانک Mel (۱۵ زیرباند)

جدول ۱- گستره باندهای بحرانی درخت ویولت مبتنی بر فیلتر بانک Mel

باند بحرانی	محدوده فرکانسی	باند بحرانی	محدوده فرکانسی
C1	0-125	C9	1000-1250
C2	125-250	C10	1250-1500
C3	250-375	C11	1500-1750
C4	375-500	C12	1750-2000
C5	500-625	C13	2000-2500
C6	625-750	C14	2500-3000
C7	750-875	C15	3000-4000
C8	875-1000		

۳- فیلتر وینر

اگر سیگنال گفتار تمیز s با نویز جمع شوند n آلوده شده باشد آنگاه سیگنال نویزی x در حوزه زمان مطابق رابطه (۳) بدست می‌آید:

$$x(i, m) = s(i, m) + n(i, m), \quad m = 0, 1, \dots, L-1 \quad (3)$$

که در آن m اندیس زمان، i اندیس فریم و L طول فریم می‌باشد. رابطه (۴) معادل حوزه ویولت معادله (۳) را نشان می‌دهد.

$$X_j(i, m) = S_j(i, m) + N_j(i, m), \quad m = 0, 1, \dots, L_j - 1 \quad (4)$$

که $X_j(i, m)$ نمونه m ام از تبدیل ویولت سیگنال گفتار نویزی در زیرباند j ام از فریم i ام، $S_j(i, m)$ نمونه m ام از تبدیل ویولت سیگنال تمیز در زیرباند j ام از فریم i ام و $N_j(i, m)$ نمونه m ام از تبدیل ویولت سیگنال نویز در زیرباند j ام از فریم i ام بوده و L_j طول زیرباند j ام می‌باشد. فیلتر وینر به صورت رابطه (۵) تعریف می‌شود:

$$H_j(i) = \frac{E\{S_j(i, m)\}}{E\{S_j(i, m)\} + est\{N_j(i)\}} \quad (5)$$

در رابطه فوق $H_j(i)$ اندازه فیلتر برای زیرباند j ام فریم i ام می‌باشد. $E\{S_j(i, m)\}$ انرژی زیرباند j ام فریم i ام سیگنال تمیز و $est\{N_j(i)\}$ تخمین انرژی نویز در زیرباند j ام فریم i ام می‌باشد که از رابطه (۶) بدست می‌آید.

$$est\{N_j(i)\} = \alpha \cdot est\{N_j(i-1)\} + (1-\alpha) \cdot E\{N_j(i, m)\} \quad (6)$$

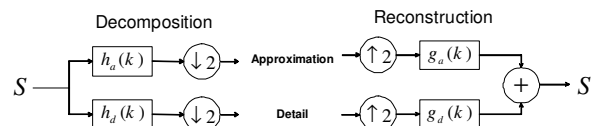
در این رابطه $E\{N_j(i, m)\}$ انرژی زیرباند j ام فریم i ام نویز تخمین زده شده، $est\{N_j(i-1)\}$ تخمین انرژی نویز در زیرباند j ام فریم $(i-1)$ ام و

$$\Psi_{a^i, b}(t) = \frac{1}{\sqrt{a^i}} \Psi\left(\frac{k-b}{a^i}\right) \quad (1)$$

که متغیر a مقیاس تابع پایه و b میزان انتقال را در طول محور t نشان می‌دهد و ضریب $1/\sqrt{a}$ جهت نرمالیزاسیون توابع پایه می‌باشد. تبدیل ویولت گسسته را می‌توان با استفاده از یک دسته از توابع پایه ارتونرمال محاسبه کرد که با DWT نشان داده می‌شود.

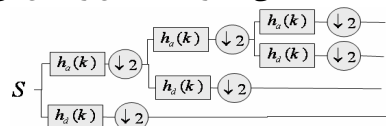
$$DWT(a^i, b) = \frac{1}{\sqrt{a^i}} \sum s(k) \Psi^*\left(\frac{k-b}{a^i}\right) \quad (2)$$

شکل ۱ ساختار ساده تبدیل ویولت یک سطحی را نشان می‌دهد که سیگنال ورودی به دو زیرباند بالا و پائین تفکیک شده و سپس بازسازی می‌شود. پاسخ فرکانسی $h_a(k)$ ، فیلتر پائین‌گذر و $h_d(k)$ ، فیلتر بالاگذر وابسته به نوع تابع پایه ویولت می‌باشد. چون فرکانس سیگنال خروجی از فیلترهای فوق نصف فرکانس سیگنال ورودی است، بنابراین تئوری نایکوئیست می‌تواند از نمونه‌ها را دور ریخت.

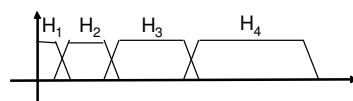


شکل ۱- ساختار تبدیل و عکس تبدیل ویولت

می‌توان تبدیل ویولت را به صورت دو درخت ویولت گسسته دوتائی و ویولت بسته ای در نظر گرفت. در درخت ویولت گسسته دوتائی که در شکل ۲ نشان داده شده است، گسترش درخت از طرف زیرباندهای پائین انجام می‌گیرد که تحلیل فرکانسی دقیقتری در این زیرباندها داریم. شکل ۳ مشخصه طیفی فیلتربانک حاصل را نشان می‌دهد.

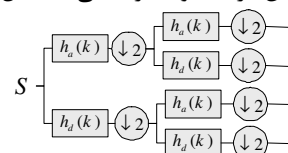


شکل ۲- درخت ویولت گسسته دوتائی با ۴ زیر باند نامساوی



شکل ۳- مشخصه طیفی فیلتربانک ویولت نشان داده شده در شکل (۲)

درخت ویولت بسته‌ای که در شکل ۴ نشان داده شده است دارای تحلیل فرکانسی یکسان در محدوده فرکانسی سیگنال خواهد بود.



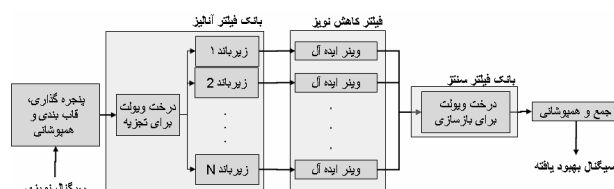
شکل ۴- درخت ویولت بسته ای با ۴ زیرباند یکنواخت

مطابق جدول (۱) درخت از ریشه شروع به گسترش می‌کند تا درخت ویولت مبتنی بر بانک فیلتر Mel ساخته شود. درخت مبتنی بر بانک فیلتر Mel پیشنهادی دارای ۱۵ زیرباند می‌باشد که درخت ویولت آن حداکثر تا عمق ۵ گسترش یافته است. ساختار این درخت در شکل ۵

α فاکتور فراموشی می‌باشد که بر اساس روش واریانس بالانس شده مطرح شده در [8] بدست می‌آید.

۴- ارزیابی اثر پارامترهای بانک فیلتر ویولت در درخت ویولت بسته ای

به منظور مقایسه اثر پارامترهای بانک فیلتر ویولت در سیستم حذف نویز از فیلتر وینر ایده‌آل استفاده شد تا در نهایت در سیگنال بهبود یافته فقط اثر بانک فیلتر را داشته باشیم. بلوک دیگرام سیستم کاهش نویز مورد استفاده در این ارزیابی در شکل ۶ نشان داده شده است. مطابق شکل ابتدا سیگنال ورودی به فریمهایی با طول مشخص با در نظر گرفتن همپوشانی تقسیم می‌شود. سپس روی هر فریم پنجره‌ای اعمال شده و در ادامه فریم توسط فیلتربانک ویولت به N زیرباند تقسیم می‌گردد و روی هر زیرباند یک فیلتر وینر ایده‌آل اعمال می‌گردد. پس از آن با استفاده از عکس تبدیل ویولت، فریم بهبود یافته بدست می‌آید و در نهایت با در نظر گرفتن همپوشانی، فریمها در کنار هم قرار می‌گیرند تا سیگنال بهبود یافته نهائی بدست آید.

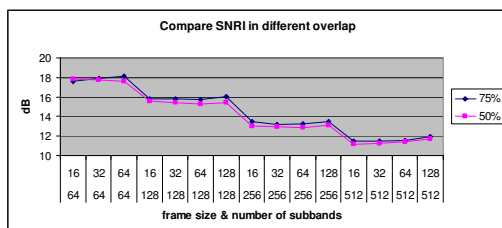


شکل ۶ - بلوک دیگرام سیستم کاهش نویز

۱-۴ ارزیابی اثر طول فریم، تعداد زیرباندها و همپوشانی

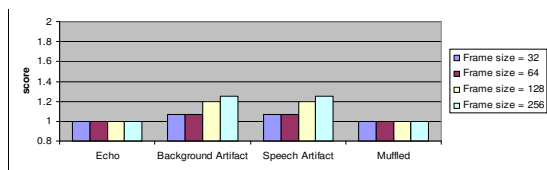
تخمین مؤلفه های فیلتر وینر ایده آل از روی دو سیگنال نویز و سیگنال گفتار انجام می‌گیرد. در این ارزیابی از ویولت بسته‌ای برای تجزیه سیگنال به زیرباندها استفاده شد. همچنین طول فریم از ۶۴ تا ۲۵۶، تعداد زیرباندها از ۱۶ تا ۱۲۸ و اندازه همپوشانی ۵۰٪ و ۷۵٪ در نظر گرفته شده است. آزمایشها روی ۵ فایل شامل صدای زن و مرد و در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB انجام گرفته است. شکل ۷ مقادیر SNRI را برای مقادیر مختلف پارامترها نشان می‌دهد. چنانچه در شکل دیده می‌شود با افزایش طول فریم مقدار SNRI کاهش می‌یابد، چرا که با افزایش طول فریم تخمین دقیقی از نویز وجود ندارد. با افزایش تعداد زیرباندها نیز مقدار SNRI افزایش می‌یابد، به این دلیل که با افزایش تعداد زیر باندها تحلیل فرکانسی، دقیقتر خواهد بود. همچنین مطابق با شکل اندازه SNRI برای همپوشانی ۷۵٪ بیشتر از همپوشانی ۵۰٪ می‌باشد. دلیل این امر آن است که فریمهای متوالی دارای وابستگی زیادی می‌باشند، بنابراین در همپوشانی بزرگتر این وابستگی بیشتر لحاظ شده است. همانطور که

در شکل دیده می‌شود، تأثیر پارامتر طول فریم بیشتر از پارامترهای دیگر می‌باشد. چنانچه با افزایش طول فریم از ۱۲۸ به ۲۵۶ حدود 5dB کاهش در اندازه SNRI مشاهده می‌شود، ولی تفاوت بکارگیری همپوشانی ۵۰٪ و ۷۵٪ حدود 1dB می‌باشد که قابل نظر کردن می‌باشد. این مطلب برای تعداد زیرباندها نیز دیده می‌شود، به این معنی که با افزایش تعداد زیر باندها تغییرات SNRI حداکثر 1dB می‌باشد.

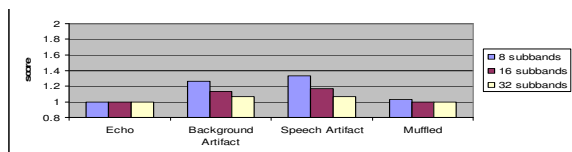


شکل ۷- مقادیر SNRI برای طول فریمها، تعداد زیرباندها و اندازه همپوشانی مختلف

تست شنوائی به این صورت است که شنوندگان به سیگنال بهبودیافته گوش داده و به ۴ جنبه وجود اکو، خفگی (Muffled)، نویز زمینه (Background Artifact) و تخریب صدا (Speech Artifact) امتیازی بین ۱ تا ۳ داده‌اند. در این تست به سیگنالهای بهبود یافته ای که کیفیت خوبی دارند امتیاز یک، به سیگنالهای بهبودیافته‌ای که دو و به سیگنالهایی که کیفیت بدی دارند و از دید شنونده قابل تحمل نیستند، امتیاز سه داده می‌شود. بر اساس نتایج این تست، با افزایش طول فریم اندازه تخریب سیگنال گفتار و نویز زمینه افزایش می‌یابد. نتایج این تست در شکل ۸ آورده شده است. در این تست تعداد زیرباندها برابر با ۳۲ است. نتایج برای طول فریمهای دیگر نیز به همین ترتیب است.



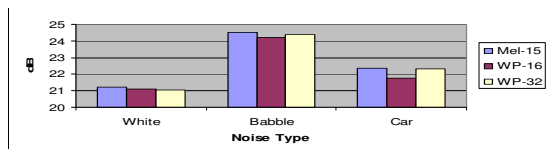
شکل ۸- نتایج تست شنوائی در تغییر طول فریم



شکل ۹- نتایج تست شنوائی در تغییر تعداد زیرباندها

چنانچه در شکل ۹ مشاهده می‌شود، با افزایش تعداد زیرباندها تخریب سیگنال گفتار و میزان نویز زمینه کاهش می‌یابد. در این تست اندازه طول فریم ۳۲ است که برای طول فریمهای دیگر نیز نتایج مشابهی بدست آمده است. در تمام سیگنالهای بهبود یافته با استفاده از ویولت دو جنبه اکو و خفگی صدا وجود نداشته‌اند. نتایج شنوائی برای ارزیابی

با ۳۲ باند از درخت ویولت بسته‌ای با ۱۶ باند بیشتر می‌باشد، چراکه با افزایش تعداد زیرباندها تحلیل فرکانسی دقیقتر خواهد بود. اندازه SNRI برای درخت پیشنهادی و درخت ویولت بسته‌ای با ۳۲ باند تقریباً یکی است و در اکثر موارد برای درخت پیشنهادی نیز بیشتر است از طرفی، چون درخت بر مبنای Mel دارای تعداد زیرباند کمتریست دارای پیچیدگی زمانی کمتری نسبت به درخت ویولت بسته‌ای می‌باشد.



شکل ۱۱- مقدار SNRI برای انواع درخت ویولت در حضور سه نوع نویز با نرخ سیگنال به نویز 10dB

بنابر نتایج تست شنوایی، درخت پیشنهادی تخریب کمتری در سیگنال گفتار نسبت به دو درخت ویولت بسته‌ای ایجاد می‌کند. بیشترین تخریب را درخت ویولت بسته‌ای با ۱۶ زیرباند داشته است. همچنین شنوندگان سیگنالهای بهبودیافته با درخت پیشنهادی را نسبت به دو درخت ویولت بسته‌ای با ۳۲ و ۱۶ باند ترجیح داده‌اند.

۶- نتیجه‌گیری

در بانک‌فیلتر ویولت، پارامتر طول فریم بیشترین تأثیر را داشته و تنظیم تعداد زیرباند متناسب با طول فریم، باعث بدست آوردن کیفیت قابل قبول برای سیگنال گفتار با حداقل محاسبات می‌گردد. درخت ویولت پیشنهادی مبتنی بر بانک فیلتر Mel نیز با بهره‌گیری از پارامترهای بهینه بانک‌فیلتر و فیلتر وینر، دارای میزان SNRI بیشتر و همچنین کیفیت شنیداری بهتری نسبت به درخت ویولت بسته‌ای است.

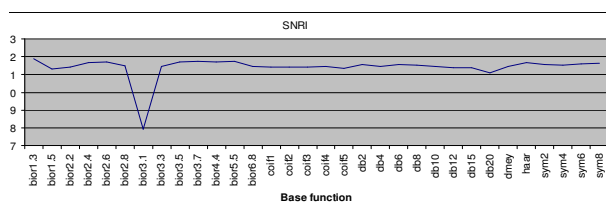
۷- مراجع

- [1] Huang A., Acero A., Hon H., *Spoken language processing*. Prentice Hall, 2001.
- [2] Donoho D.L., Johnston I.M., "Ideal spatial adaptation by wavelet shrinkage", *Biometrika* 81, 425-455, 1994.
- [3] Donoho D.L., "De-noising by soft-thresholding" *IEEE Trans. On Information Theory*, 613-627, 1995.
- [4] Sheikhzadeh H., Abutalebi H.R., "An improved wavelet based speech enhancement system", *ECSCCT2001*.
- [5] Lu C.T., Wang H.C., "Enhancement of single channel speech based on masking property and wavelet transform", *Speech Communication*, vol.41, 409-427, 2003.
- [6] Soon I.Y., Kah S.N., Yeo C.K., "Wavelet for speech denoising", *IEEE TENCON*, vol.2, 479-482, 1997.
- [7] Grossman A., Martinet R., Morlet J., *Reading and understanding continuous wavelet transform*. Springer-Verlag, 2-20, 1989.
- [8] Martin R., "Noise Power Spectral Density Estimation Based on Optimal smoothing and Minimum Statistic", *IEEE Trans. on Speech and Audio Processing*, Vol.9, No.5, July 2001.
- [9] Rufiner H.L., Goddard J., "A Method of Wavelet Selection in Phone Recognition", *IEEE International Conference on Circuits and Systems*, vol.2, 889-891, 1997.

پارامتر همپوشانی نشان داد که همپوشانی ۷۵٪ بهتر از همپوشانی ۵۰٪ عمل می‌کند ولی این بهبود ناچیز است.

۴-۲ ارزیابی اثر انتخاب نوع تابع پایه ویولت

پارامتر مهم دیگر ویولت، تابع پایه ویولت می‌باشد. Kah, Soon و Kiat با تعریف معیار خطائی که از تفاضل سیگنال ورودی با سیگنال تخمین ویولت داشت به ارزیابی انواع پایه‌های مختلف پرداخته‌اند [6]. در سال ۱۹۹۷ Goddard و Rufiner اثر پایه‌های ویولت در تشخیص سیگنال گفتار را بررسی کردند [9]. شکل ۱۰ نتایج SNRI را برای پایه‌های مختلف ویولت نشان می‌دهد. مطابق شکل تغییرات SNRI برای پایه‌های مختلف ویولت بجز پایه bior3.1 در حد 1dB است. چراکه فیلتر پائین‌گذر و بالاگذر پایه bior3.1 دارای همپوشانی زیادی در حوزه فرکانس هستند و در تفکیک سیگنال به زیرباندهای فرکانس بالا و پائین خوب عمل نمی‌کند. این اثر در [6,10] گزارش شده است.



شکل ۱۰- نتایج SNRI در توابع پایه ویولت مختلف

تست شنوایی نشان داد که پایه‌های ویولت Haar و bior3.1 دارای تخریب سیگنال گفتار بیشتری نسبت به پایه‌های دیگر می‌باشند. از طرفی پایه‌های Symlet, Coiflet, Daubechies با درجه بالاتر نسبت و نیز پایه Discrete Meyer دارای کارائی خوبی در حذف نویز می‌باشند.

۵- پیاده‌سازی و ارزیابی درخت Mel پیشنهادی

برای ارزیابی اثر درخت پیشنهادی ویولت با در نظر گرفتن مقادیر بدست آمده برای پارامترهای بانک‌فیلتر ویولت و فیلتر وینر ارائه شده در قسمت ۴، ارزیابی روی فایل‌های صوتی که نویز به آنها اضافه شده است، انجام می‌گیرد. فایل‌های ورودی با فرکانس 8kHz نمونه‌برداری می‌شود و هر نمونه 16 بیت می‌باشد. نویز گوسی سفید، car و همهمه با نرخهای سیگنال به نویز متفاوت 0dB، 10dB و 20dB به صداهای زن و مرد اضافه شده و مورد استفاده قرار گرفته‌اند. در این آزمایش برای تخمین نویز از فایل نویز استفاده شده است.

شکل ۱۱ مقادیر SNRI را برای سه نوع درخت ویولت با طول فریم ۱۲۸ در حضور سه نویز گوسی سفید، car و همهمه با سطح سیگنال به نویز ورودی 10dB نشان می‌دهد. مقدار SNRI برای درخت ویولتی که بر مبنای Mel طراحی شده است بالاتر از دو درختی است که زیرباندهای آنها یکنواخت می‌باشند. اندازه SNRI درخت ویولت بسته‌ای