

حذف نویز از سیگنال گفتار با استفاده از تبدیل موجک بر اساس تشخیص واک

بهزاد زمانی دهکردی	مهدی محمدی	شیما طبیبیان	بابک ناصرشریف	احمد اکبری
دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران
bzamani@iust.ac.ir	mh_mohammadi@iust.ac.ir	shimatabibian@comp.iust.ac.ir	nasser_s@iust.ac.ir	akbari@iust.ac.ir

است، حد آستانه را در باندهای بالائی افزایش داده و در صورتی که فریم بی‌واک باشد، حد آستانه را در باندهای بالائی کاهش می‌دهند. [۳] برای حل مشکل روشهای حد آستانه، فیلترهایی نظیر وینر در حوزه تبدیل موجک مطرح شدند که پارامترهای فیلتر وابسته به شرایط آماری ضرایب تبدیل موجک تنظیم می‌گردد [۵، ۶]. در این مقاله، با استفاده از اندازه آنتروپی در یکی از زیرباندهای تبدیل موجک سیگنال گفتار، روشی برای تشخیص فریم واکدار از بی‌واک ارائه شده است. و دو روش حذف نویز یکی تغییر تابع پایه موجک و دیگری تغییر میزان حذف نویز در فیلتر وینر بر اساس واکدار/بی‌واک بودن پیشنهاد شد.

در ادامه مقاله، در بخش دوم روشهای پیشنهادی کاهش نویز معرفی شده و نحوه محاسبه فیلتر وینر در حوزه موجک شرح داده شده است. بخش سوم حاوی روش تشخیص فریم واکدار/بی‌واک است. بخش چهارم شامل نتایج ارزیابی روشهای پیشنهادی است. در انتها نیز نتیجه‌گیری نهایی آورده شده است.

۲- سیستمهای پیشنهادی کاهش نویز در حوزه موجک

۱-۲ بلوک دیگرگرام سیستمهای پیشنهادی کاهش نویز

روشهای حذف نویز در حوزه موجک را می‌توان به دو دسته تقسیم نمود، روشهای بر مبنای آستانه‌گذاری و روشهای فیلترینگ. در سیگنال گفتار، غالباً قسمت عمده انرژی در تعداد کمی از ضرایب تبدیل موجک متمرکز است. این ضرایب نسبت به ضرایب سیگنال نویز که انرژی آن بر روی تعداد زیادی از ضرایب گسترش می‌یابند، بزرگتر هستند. بنابراین می‌توان با حذف ضرایب کوچکتر و مقایسه آنها با یک حد آستانه، علاوه بر حفظ اطلاعات مهم سیگنال اصلی، نویز را نیز محدود کرد. روشهای آستانه‌گذاری عموماً دارای کیفیت بد شنیداری بدلیل تخمین کلی برای آستانه می‌باشند [۱، ۳]. همچنین به دلیل تخمین حد آستانه در یک بازه زمانی وسیع، سیستم تأخیر زمانی دارد که برای کاهش این تأخیر زمانی می‌توان از فریم‌بندی سیگنال استفاده کرد. برای بالا بردن کیفیت سیگنال بهبود یافته می‌توان از روشهای فیلترینگ استفاده نمود. [۲، ۹]

چکیده: سیستمهای بهبود گفتار مبتنی بر موجک موفقیت خود را در عمل نشان داده‌اند، که سیگنال صوتی می‌تواند در هر دو حوزه زمان و فرکانس با رزولوشن دلخواه تجزیه گردد. و با آستانه‌گذاری یا فیلتر کردن اجزا سیگنال بهبودیافته بازسازی می‌گردد. در مقاله حاضر، ابتدا یک سیستم تشخیص فریمهای واکدار/بی‌واک مبتنی بر آنتروپی زیرباندهای موجک معرفی می‌گردد. سپس، دو روش ارائه می‌گردد. در روش اول، برحسب واکدار/بی‌واک بودن فریمها تابع پایه موجک تغییر می‌کند و سپس با استفاده از فیلتر وینر عمل حذف نویز صورت می‌گیرد. در روش دوم، با تنظیم میزان حذف نویز در فیلتر وینر برحسب واکدار/بی‌واک بودن فریمها، عمل کاهش نویز در زیرباندهای موجک صورت می‌گیرد. نتایج ارزیابی‌های عملی و نظری بیانگر کارایی بهتر انتخاب پایه و تنظیم میزان حذف نویز بر حسب واکدار/بی‌واک بودن فریمها در حذف نویز و بهبود کیفیت گفتار می‌باشد.

واژه های کلیدی: آنتروپی، تبدیل موجک، فیلتر وینر، درخت موجک بسته‌ای، واکدار و بی‌واک.

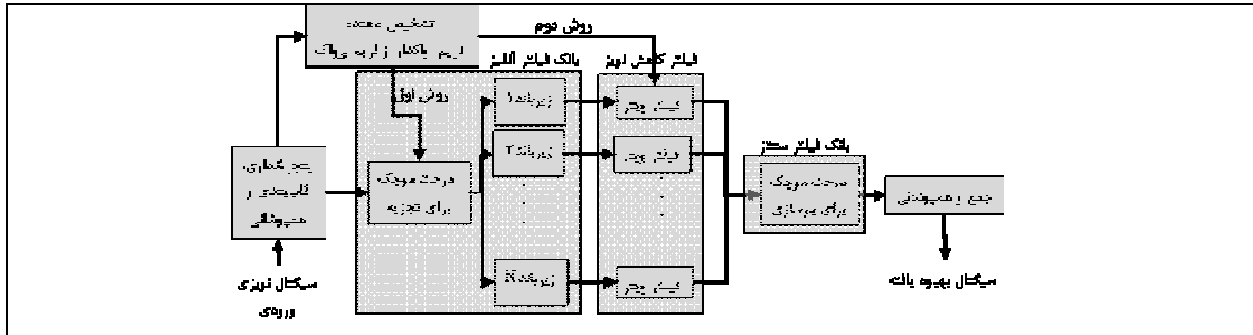
۱- مقدمه

حذف نویز از سیگنال گفتار در بسیاری از کاربردها از جمله تلفنهای بدون گوشی که صدا در یک محیط نویزی دریافت می‌گردد، یک نیاز ضروری است. حذف نویز در حوزه موجک یکی از روشهای مهم در کیفیت بهبود گفتار می‌باشد.

در روشهای مبتنی بر تبدیل موجک، از پنجره‌های زمانی با طول متفاوت برای باندهای فرکانسی مختلف استفاده می‌گردد. برای حذف نویز استفاده از آستانه‌گذاری ضرایب تبدیل موجک، متداولترین روش است. [۱، ۲، ۳] به علت عملکرد ضعیف روشهای آستانه‌گذاری در نویزهای رنگی، روش آستانه‌گذاری در باندهای بحرانی درخت تجزیه موجک ارائه گردید [۴]، که اندازه آستانه بطور اتوماتیک در باندهای بالا در قطعه واکدار و در باندهای پائین در قطعه بی‌واک افزایش می‌یابد [۴]. آقایان شیخ زاده و ابوطالبی پس از فریم‌بندی سیگنال نویزی ورودی، فریم را به واکدار/بی‌واک تقسیم کرده و با توجه به اینکه فریم واکدار

جمع بدست می‌آید. در این مقاله دو الگوریتم حذف نویز ارائه شده که بر مبنای یک تشخیص دهنده فریم واکدار از بی‌واک عمل می‌کند. این تشخیص دهنده با استفاده از آنتروپی زیرباند اول در درخت موجک بسته‌ای ۴ تایی، واکدار یا بی‌واک بودن فریم را مشخص می‌کند. سپس در روش اول حذف نویز تابع پایه موجک را با توجه به واکدار یا بی‌واک بودن فریم تغییر داده که شکل (۱) بلوک دیاگرام آن می‌باشد.

در این مقاله از روش حذف نویز با فیلتر وینر استفاده شد. [۵] پس از فریم‌بندی و اعمال پنجره، فریم حاصل توسط موجک بسته‌ای به تعدادی زیرباند تبدیل می‌شود. از طرفی فیلتر وینر برای هر یک از زیرباندها محاسبه و در آن زیرباند اعمال می‌شود. پس از حذف نویز از سیگنال ورودی طی مرحله سنتز که بر اساس درخت موجک بسته‌ای معکوس کار می‌کند سیگنال بهبود یافته بر اساس روش همپوشانی و



شکل (۱): بلوک دیاگرام سیستم‌های پیشنهادی کاهش نویز بر مبنای موجک بر اساس واکدار/بی‌واک بودن فریم

باشد که از رابطه (۵) بدست می‌آید و β ضریبی است که میزان سطح نویز باقیمانده را کنترل می‌کند.

$$E_{est}\{N_j(i)\} = a.E_{est}\{N_j(i-1)\} + (1-a).E\{N_j(i,m)\} \quad (4)$$

که $E\{N_j(i,m)\}$ انرژی زیرباند m فریم i نویز تخمین زده شده، $E_{est}\{N_j(i-1)\}$ تخمین انرژی نویز در زیرباند m فریم $(i-1)$ و a و α فاکتور فراموشی می‌باشد که بر اساس روش واریانس بالانس شده بدست می‌آید [۱۰]. در این روش اندازه ضریب فراموشی به طول فریم بستگی دارد به این صورت که α برای فریم‌های بزرگ به یک نزدیک شده و برای فریم‌های کوچک به صفر نزدیک می‌شود.

۳- تشخیص دهنده فریم واکدار و بی‌واک

فونم‌های گفتار به دو دسته واکدار و بی‌واک تقسیم می‌شوند. جدول (۱) فونم‌های واکدار و بی‌واک را در زبان انگلیسی نشان می‌دهد. فونم‌های واکدار پر انرژی بوده و فونم‌های بی‌واک شبه نویز می‌باشند.

جدول (۱) - فونم‌های واکدار و بی‌واک

voiced	ây-oy-ey-ow-ay-i-e-a-u
unvoiced	o-â-b-d-g-q-v-z-ž-m-n
	p-t-k-f-s-š-x-č

در تشخیص دهنده فریم واکدار از بی‌واک، ابتدا تبدیل موجک بسته‌ای فریم ورودی تا سطح ۲ بدست آمده، سپس آنتروپی زیرباند اول محاسبه می‌گردد. با مقایسه اندازه آنتروپی با حد آستانه، واکدار یا بی‌واک بودن فریم مشخص می‌شود. طول فریم ۲۵۶ نمونه، تابع پایه db10 و حد آستانه ۱۰.۲ می‌باشد، حد آستانه از روی ۹۴۸ فریم بدست آمده که ۶۹۳ فریم آن واکدار و ۱۵۸ فریم آن بی‌واک می‌باشند، مابقی فریمها شامل گفتار نمی‌باشند که جزء فریم‌های واکدار در نظر گرفته می‌شوند. فریمی که اندازه آنتروپی آن از حد آستانه بیشتر باشد،

در روش پیشنهادی دوم، میزان سطح نویز باقیمانده بر اساس واکدار یا بی‌واک بودن فریم، تغییر می‌یابد که بلوک دیاگرام آن در شکل (۱) به عنوان روش دوم آورده شده است. که هدف کاهش تخریب سیگنال گفتار بهبود یافته می‌باشد. به این صورت که در فریم‌های واکدار نسبت به فریم‌های بی‌واک نویز بیشتری حذف می‌شود. در ادامه نحوه محاسبه فیلتر حذف کننده نویز یعنی فیلتر وینر آورده شده است.

۲-۲ فیلتر وینر

اگر سیگنال گفتار تمیز s با نویز جمع شونده n آورده شده باشد. می‌توان سیگنال نویزی x در حوزه زمان را با رابطه (۱) نشان داد.

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

که در آن m اندیس زمان، L طول فریم و i اندیس فریم می‌باشد. رابطه (۲) معادل حوزه موجک رابطه (۱) را نشان می‌دهد.

$$X_j(i,m) = S_j(i,m) + N_j(i,m), \quad m = 0,1,\dots,L_j-1 \quad (2)$$

که $X_j(i,m)$ ، $S_j(i,m)$ و $N_j(i,m)$ به ترتیب نمونه m ام از زیرباند j را برای تبدیل موجک سیگنال‌های گفتار نویزی، گفتار تمیز و نویز در فریم i ام نشان می‌دهد. و L_j طول زیرباند j ام می‌باشد. فیلتر حذف نویز در حوزه موجک فیلتر حذف نویز $H_j(i)$ به صورت زیر تعریف می‌گردد:

$$H_j(i) = \frac{E\{S_j(i,m)\} + b.E_{est}\{N_j(i)\}}{E\{S_j(i,m)\} + E_{est}\{N_j(i)\}} \quad (3)$$

که $E\{S_j(i,m)\}$ انرژی زیرباند j ام در فریم i ام سیگنال تمیز، $E_{est}\{N_j(i)\}$ تخمین انرژی نویز در زیرباند j ام مربوط به فریم i ام می‌-

جدول (۲) مقادیر SNR بهبودیافته برای توابع پایه مختلف در فریم‌های واکدار و بی‌واک در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB نشان می‌دهد. در این جدول سطرها، تابع پایه برای فریم‌های بی‌واک و ستونها، تابع پایه را برای فریم واکدار نشان می‌دهند. مطابق با جدول تغییرات SNR بهبودیافته در یک سطر بیشتر از تغییرات در یک ستون می‌باشد. بنابراین می‌توان گفت که انتخاب تابع پایه برای فریم واکدار از اهمیت بالاتری نسبت به فریم بی‌واک برخوردار است. چراکه تعداد فریم‌های واکدار در طول یک جمله نسبت به تعداد فریم‌های بی‌واک بیشتر بوده، همچنین این فریم‌ها دارای انرژی بیشتری هستند و در شنیدن نیز گوش نسبت به آنها حساسیت بیشتری نشان می‌دهد.

جدول (۲) - مقادیر SNRI برای توابع پایه مختلف در فریم‌های واکدار و بی‌واک در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB

		voiced				
		bior3.1	bior3.7	bior4.4	db20	db5
unvoiced	bior3.1	18.32	23.36	23.72	23.32	23.08
	bior3.7	18.27	23.72	24.21	23.13	23.41
	bior4.4	17.94	23.51	23.57	22.93	23.12
	db20	18.04	23.71	24.06	23.12	23.31
	db5	17.92	23.56	23.7	22.95	23.14

جدول (۳) - مقادیر PESQ برای توابع پایه مختلف در فریم‌های واکدار و بی‌واک در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB

		voiced				
		bior3.1	bior3.7	bior4.4	db20	db5
unvoiced	bior3.1	2.972	3.106	2.951	2.968	3.012
	bior3.7	2.814	3.156	3.119	3.305	3.191
	bior4.4	2.737	3.055	3.17	3.308	3.219
	db20	2.659	3.027	3.223	3.345	3.249
	db5	2.765	3.064	3.21	3.343	3.267

جدول (۳) مقادیر PESQ را برای حالت‌های مختلف بکارگیری پایه موجک نشان می‌دهد. مطابق با جدول تغییرات PESQ در یک سطر بمراتب بیشتر از تغییرات در یک ستون است که گواه اهمیت بیشتر فریم واکدار نسبت به فریم بی‌واک در تست شنوایی است.

توابع پایه با درجه پایین‌تر پیچیدگی محاسبات کمتری دارند که در کارهای زمان واقعی بهتر است این امر را در نظر گرفته شود. می‌توان از bior3.1 که به تنهایی دارای نتایج ضعیفی بود نیز در فریم‌های بی‌واک استفاده کرد و نتایج قابل قبولی بدست آورد.

۴-۲ اثر تغییر میزان نویز باقیمانده بر اساس فریم

واکدار/بی‌واک

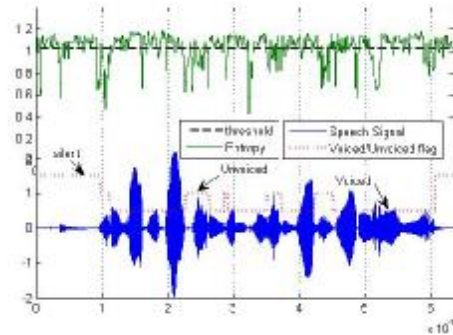
برای بررسی اثر تغییر میزان نویز باقیمانده (ضریب β) با توجه به فریم واکدار/بی‌واک از رابطه (۳)، فیلتر وینر استفاده می‌کنیم، که β عددی بین صفر و یک می‌باشد. مقدار صفر به معنی حذف کامل نویز و مقدار یک به معنی عدم حذف نویز می‌باشد.

واکدار و فریمی که کمتر باشد، بی‌واک می‌باشد. دقت سیستم تشخیص‌دهنده در حدود ۸۲.۲۹٪ می‌باشد. شکل (۲) نمونه سیگنال گفتار را به همراه آنتروپی فریم‌های متوالی نشان می‌دهد. پس از بدست آمدن ضرائب موجک زیرباند اول، هیستوگرام آن بدست آمده، آنتروپی زیرباند با استفاده از رابطه (۵) بدست می‌آید. لازم به ذکر است که در بدست آوردن هیستوگرام، تعداد میله ها، B ، مهم است.

$$E_1(i) = -\sum_{k=1}^B P_k \log(p_k) \quad (5)$$

که $E_1(i)$ اندازه آنتروپی زیرباند یکم از فریم i ام می‌باشد. همچنین مقدار P_k نیز از رابطه (۶) بدست می‌آید:

$$P_k = \frac{\text{No. of wavelet packet coefficients in bin } k}{\text{Node size in wavelet packet tree}} \quad (6)$$



شکل (۲): سیگنال گفتار به همراه آنتروپی فریم‌های متوالی

۴-۴ نتایج پیاده سازی

در آزمایشات از دادگان TIMIT (که دنباله فونم‌ها، شروع و خاتمه هر فونم مشخص می‌باشد) استفاده شد. و برای ارزیابی نتایج از دو معیار نرخ سیگنال به نویز بهبودیافته و معیار PESQ استفاده شده است. روش PESQ یک تست نظری از جانب ITU برای ارزیابی کیفیت سیگنال گفتار می‌باشد که به فایده بهبودیافته نمره‌ای بین صفر و ۴.۵ می‌دهد. نمره صفر به سیگنال نویزی و تخریب شده و نمره ۴.۵ به سیگنال تمیز داده می‌شود. در این روش هر چه عدد بزرگتر باشد به معنی کیفیت بالاتر می‌باشد. [7, 8]

۴-۱ اثر تغییر نوع تابع پایه بر اساس فریم واکدار و

بی‌واک

در این ارزیابی با توجه به واکدار یا بی‌واک بودن آن فریم، تابع پایه موجک تغییر می‌یابد. توابع bior3.1, bior3.7, bior4.4, coif3, sym2, dmey, db5, db20 به عنوان تابع پایه موجک ارزیابی شدند. در این قسمت در رابطه فیلتر وینر ضریب β صفر در نظر گرفته می‌شود بنابراین تمامی نویز حذف می‌گردد. بدلیل حجم زیاد آزمایشات بخشی از نتایج در مقاله آورده شد.

تشخیص‌دهنده ارائه شده در حدود ۸۲.۲۹٪ می‌باشد. از آنجایی که تعداد فریم‌های واکدار بیشتر از تعداد فریم‌های بی‌واک می‌باشد و همچنین این فریم‌ها پر انرژی هستند انتظار می‌رود که رفتار متفاوت برای هر فریم متناسب با ماهیتش می‌تواند مؤثر باشد. از اینرو بر اساس نتیجه تشخیص‌دهنده به دو نحو عمل می‌شود.

در رویکرد اول، تابع پایه بر اساس واکدار/بی‌واک بودن فریم تغییر داده می‌شود. در این بخش نتایج عملی و نظری قابل قبول با احتساب میزان محاسبات، در بکارگیری تابع پایه با درجه بالا در فریمهای واکدار و تابع پایه با درجه پائین در فریمهای بی‌واک بدست آمد. هر چه درجه توابع پایه موجک بیشتر شود، عملکرد بهتری دارد ولی از طرفی طول پایه بزرگتر شده و موجب بیشتر شدن تعداد عمل ضرب و جمع در شکل گیری درخت موجک می‌گردد. روش فوق با بکارگیری تابع پایه موجک ضعیف با طول کم در فریم‌های بی‌واک موجب کاهش محاسبات بدون از دست دادن کیفیت نهایی می‌شود.

در رویکرد دوم، میزان حذف نویز در فیلتر وینر در فریمهای واکدار/بی‌واک بطور متفاوت تنظیم می‌شود که میزان حذف نویز در این روش در فریمهای واکدار بیش از حذف نویز در فریمهای بی‌واک است. ارزیابی عملی و نظری نشان داد که تخریب سیگنال گفتار با این روش کاهش یافته است.

۶- مراجع

- [1] Bont F., Groenewegen M., and Oomen W., "A High Quality Audio-Coding System at 128 kb/s", Audio Engineering Society Conference, Feb 1995.
- [2] Donoho, D.L., "De-noising by soft-thresholding", IEEE Transaction Informmation Theory 41(3), pp. 613-627, 1995.
- [3] Ing Y.S., Soo N.K., Chai K.Y., "Wavelet for speech denoising", IEEE. TENCON'97 vol.2, pp. 479 - 482, 1997.
- [4] Lu C.T., Wang H.C., "Enhancement of single channel speech based on masking property and wavelet transform", Speech Communication, Volume 41, Issues 2-3, pp. 409-427, 2003.
- [5] بهزاد زمانی، احمد اکبری، بابک ناصر شریف، بررسی اثر باندهای بحرانی و پارامترهای بکارگیری تبدیل موجک در بهبود گفتار، یازدهمین کنفرانس بین المللی کامپیوتر ایران، ۱۳۸۴.
- [6] Huang A., Acero A., Hon H., "Spoken language processing", Prentice Hall, 2001.
- [7] "Perceptual Evaluation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrow band telephone network and speech codec", ITU-T Recommend, pp. 862, Feb 2001.
- [8] Rix A.W., Beerends J.G., Hollier M.P., Hekstra A.P., "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs", in Proc. IEEE ICASSP'2001, vol. 2, pp. 749-752, May 2001
- [9] Mallat S., "A Wavelet Tour of Signal Processing", Academic Press, 1999.

جدول (۶) نتایج SNR بهبودیافته را برای β های مختلف در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB نشان می‌دهد. در این جدول با توجه به اینکه فریم واکدار یا بی‌واک است β های مختلف بکار رفته است. بعنوان مثال، سطر $\beta=0.6$ و ستون $\beta=0.2$ به این معنی است که در این آزمایش برای فریم‌های واکدار β را 0.6 و برای فریم‌های بی‌واک β را 0.2 قرار داده شده است.

چنانچه در جدول دیده می‌شود تغییرات SNRI در یک ستون کمتر از تغییرات SNR بهبودیافته در یک سطر می‌باشد و یعنی میزان نویز باقیمانده در فریم‌های واکدار از اهمیت بیشتری برخوردار است، پس بهتر است که در فریم‌های واکدار نویز بیشتری حذف و در فریم‌های بی‌واک نویز کمتری حذف شود.

چراکه فریم‌های بی‌واک شبه نویز می‌باشند بنابراین بهتر است که در این فریمها نویز کمتری حذف گردد تا تخریب گفتار کمتر ایجاد شود.

جدول (۶) - مقادیر SNR بهبودیافته برای β های مختلف در حضور نویز

سفید گوسی با نرخ سیگنال به نویز ورودی 10dB

		voiced					
		0	0.2	0.4	0.6	0.8	1
unvoiced	0	23.24	17.85	13.6	10.81	8.81	7.29
	0.2	21.44	17.2	13.35	10.69	8.74	7.25
	0.4	18.72	15.95	12.82	10.41	8.58	7.15
	0.6	16.34	14.55	12.12	10.03	8.34	6.99
	0.8	14.41	13.2	11.34	9.56	8.05	6.79
	1	12.84	11.98	10.54	9.05	7.71	6.56

جدول (۷) - مقادیر PESQ برای β های مختلف در حضور نویز سفید

گوسی با نرخ سیگنال به نویز ورودی 10dB

		voiced					
		0	0.2	0.4	0.6	0.8	1
unvoiced	0	3.251	3.128	2.899	2.723	2.588	2.471
	0.2	3.255	3.125	2.896	2.735	2.602	2.488
	0.4	3.197	3.077	2.858	2.708	2.579	2.464
	0.6	3.147	3.046	2.842	2.706	2.573	2.449
	0.8	3.077	2.997	2.814	2.684	2.551	2.434
	1	3.02	2.969	2.776	2.649	2.509	2.396

جدول (۷) مقادیر PESQ برای β های مختلف در حضور نویز سفید گوسی با نرخ سیگنال به نویز ورودی 10dB نشان می‌دهد. مطابق با نتایج جدول واضح است که تغییرات PESQ در یک ستون کمتر از تغییرات در یک سطر است.

مطابق با نتایج بدست آمده می‌توان گفت که با تشخیص واکدار یا بی‌واک بودن فریم، می‌توان بار محاسباتی را بدون از دست دادن کیفیت نهایی کاهش داد. همچنین می‌توان میزان نویز باقیمانده را نیز بگونه‌ای تنظیم کرد که کمترین تخریب را در گفتار داشته باشیم.

۵- نتیجه گیری

در مقاله حاضر روشی برای تشخیص واکدار/بی‌واک بودن فریمها در حوزه موجک بر اساس آنتروپی زیرباند موجک ارائه گردید. دقت

Transaction on Speech and Audio Processing, Vol.9, No.5, July 2001.

[10] Martin R., "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistic", IEEE