CEO Clinical and Experimental Otorhinolaryngology Vol. 5, Suppl 1: S65-S68, April 2012 http://dx.doi.org/10.3342/ceo.2012.5.S1.S65

Original Article

Noise Reduction Using Wavelet Thresholding of Multitaper Estimators and Geometric Approach to Spectral Subtraction for Speech Coding Strategy

Kai Chuan Chu, MSc[1,2] · Charles T. M. Choi, PhD[1,3]

[1]Department of Computer Science and Institute of Biomedical Engineering, National Chiao Tung University, Taiwan;
[2]Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia;
[3]Department of Electrical Engineering, National Chiao Tung University, Taiwan

Objectives. Noise reduction using wavelet thresholding of multitaper estimators (WTME) and geometric approach to spectral subtraction (GASS) can improve speech quality of noisy sound for speech coding strategy. This study used Perceptual Evaluation of Speech Quality (PESQ) to assess the performance of the WTME and GASS for speech coding strategy.

Methods. This study included 25 Mandarin sentences as test materials. Environmental noises including the air-conditioner, cafeteria and multi-talker were artificially added to test materials at signal to noise ratio (SNR) of -5, 0, 5, and 10 dB. HiRes 120 vocoder WTME and GASS noise reduction process were used in this study to generate sound outputs. The sound outputs were measured by the PESQ to evaluate sound quality.

Results. Two figures and three tables were used to assess the speech quality of the sound output of the WTME and GASS.

Conclusion. There is no significant difference between the overall performance of sound quality in both methods, but the geometric approach to spectral subtraction method is slightly better than the wavelet thresholding of multitaper estimators.

Key Words. *Cochlear implant, Speech production measurement*

INTRODUCTION

One of the main challenges in developing an efficient cochlear implant lies in speech coding strategy that can elicit neural sensations that correspond to those generated by the normal hearing mechanism. Currently, the speech coding strategy of cochlear implants has improved significantly over the past few decades as a result of advancements in technology. In a quiet environment, some of the cochlear implant users achieved sentence intelligibility scores of 80% to 90%. However, the ability of most implant users to understand speech in noisy environments, understand music and understand tone languages remain a challenge to improve.

Even though most speech enhancement algorithms are able to improve speech quality for speech coding strategy they suffer from an annoying artifact called "musical noise" (1-3). Musical noise is caused by randomly spaced spectral peaks that come and go in each frame, and occur at random frequencies. The randomly spaced peaks are due to the inaccurate and large-variance estimates of the spectra of noise and noisy signals, typically computed using periodogram-type methods (3).

Two noise reduction methods for speech coding strategy that can reduce musical noise are discussed in this study: the wavelet thresholding of multitaper estimators (WTME) (3) and the geo-

• Received December 1, 2011
Revision January 12, 2012
Accepted February 1, 2012

• Corresponding author: **Charles T. M. Choi, PhD**
Department of Computer Science and Institute of Biomedical Engineering, National Chiao Tung University, 1001, Ta Hsueh Road Hsinchu 30010, Taiwan
Tel: +886-3-573-1978, Fax: +886-3-572-1490
E-mail: c.t.choi@ieee.org

metric approach to spectral subtraction (GASS) (4, 5). In order to evaluate the sound quality of the discussed algorithms, the Perceptual Evaluation of Speech Quality (PESQ) (6, 7) objective measurement is used.

MATERIALS AND METHODS

This study included 25 sentences of test materials. The sentences were recorded in Mandarin and produced by a male speaker. The sentences were originally sampled in 44.1 kHz and down sampled to 16 kHz. Noise from different environments was artificially added to test materials, including air-conditioner, cafeteria and multi-talker, at SNRs of -5, 0, 5, and 10 dB. A total of 300 combination signals were generated (25 sentences×3 noises×4 difference signal to noise ratio [SNR]).

This study used a HiRes 120 strategy (8) together with a noise reduction process (Fig. 1) as the speech coding strategy or vocoder. All the test materials were processed as the input sound of the vocoder and total of 600 sound outputs were generated (300 input sounds×2 noise reduction methods). The sound outputs were then measured by the PESQ (Figs. 1-3) to evaluate the sound quality.

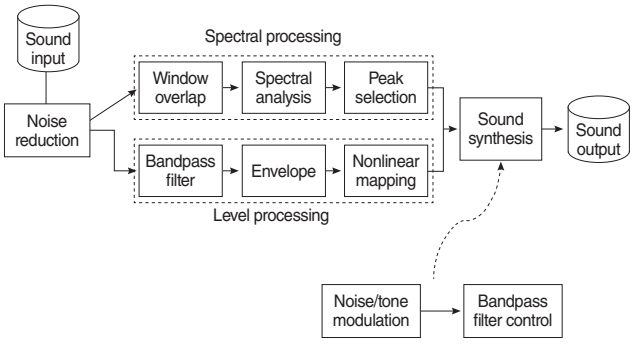

Fig. 1. HiRes 120 vocoder with noise reduction process.

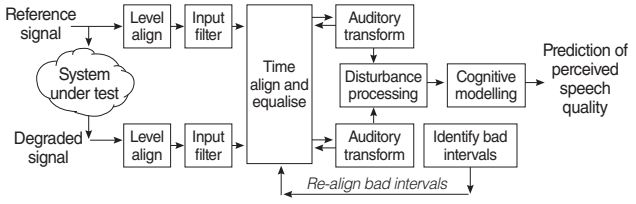

Fig. 2. Structure of perceptual evaluation of speech quality model (6).

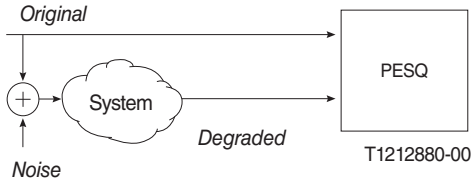

Fig. 3. Perceptual Evaluation of Speech Quality (PESQ) evaluation method for testing quality with environmental noise (7).

The PESQ evaluation (6) begins by level aligning both signals to a standard listening level. They are filtered (using a fast Fourier transform [FFT]) with an input filter to model a standard telephone handset. The signals are aligned in time and then processed through an auditory transform similar to that of perceptual speech quality measure (PSQM). The transformation also involves equalising for linear filtering in the system and for gain variation. Two distortion parameters are extracted from the disturbance (the difference between the transforms of the signals), and are aggregated in frequency and time and mapped to a prediction of subjective mean opinion score (MOS). The final PESQ score is a linear combination of the average disturbance value and the average asymmetrical disturbance value. The range of the PESQ score is -0.5 to 4.5, although for most cases the output range will be a listening quality MOS-like score between 1.0 and 4.5, the normal range of MOS values found in an absolute category rating (ACR) experiment (7). The bigger the value of PESQ is, the better the sound quality will be.

Two noise reduction methods were implemented in the vocoder. The first noise reduction method, WTME (3, 9, 10) can be implemented in four steps. For each speech frame:

1. Calculate the logarithm of the multitaper estimate.
2. Apply a standard, periodic, partial discrete wavelet transform (DWT) out to decomposition level (q_0) to the log periodogram ordinates and get the empirical DWT. For implementation, $q_0=5$.
3. Apply thresholding to the wavelet coefficients.
4. Invert the partial DWT to the thresholded wavelet coefficients and produce the smoothed spectrum estimate.

The second noise reduction method, GASS (4, 5) can be implemented in five steps. For each speech frame:

1. Using the FFT magnitude spectrum of the noisy signal.
2. Using a noise estimation algorithm (5), update the power spectrum of the noise signal.
3. Compute the instantaneous estimate and use it to compute the smoothed estimate. Then estimate the gain function (4).
4. Obtain the enhanced magnitude spectrum by the product of noisy signal and gain.
5. Compute the inverse FFT with enhanced magnitude spectrum to obtain the enhanced speech signal.

RESULTS

The performance of the two methods in measuring speech quality is showed in Tables 1-3 and Figs. 4, 5. Firstly, Table 1 showed the PESQ scores of 25 sentences from difference background environments (air-conditioner, cafeteria and multi-talker) at SNRs of -5, 0, 5, and 10 dB that processed by the WTME noise reduction method in the HiRes 120 vocoder. The PESQ scores of 25 sentences from difference background environments (air-conditioner, cafeteria and multi-talker) at SNRs of -5, 0, 5, and 10 dB

Table 1. The perceptual evaluation of speech quality scores of 25 sentences from difference background environments at signal to noise ratios (SNRs) of -5, 0, 5, and 10 dB*

Noise	SNR	Sentences																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Air-conditioner	-5	0.87	0.84	0.68	0.89	0.58	1.29	0.78	0.74	0.77	1.00	0.84	1.03	0.69	0.59	1.07	0.76	0.62	0.63	0.79	0.63	0.85	0.79	0.90	0.74	0.70
	0	1.21	1.25	1.16	1.28	1.35	1.47	1.11	1.21	1.40	1.26	1.28	1.23	1.13	1.20	1.38	1.12	0.96	1.16	1.46	1.18	1.26	1.30	1.36	0.87	1.24
	5	1.36	1.53	1.51	1.50	1.29	1.58	1.42	1.51	1.58	1.49	1.43	1.44	1.30	1.29	1.67	1.20	1.35	1.43	1.49	1.40	1.38	1.55	1.55	1.21	1.45
	10	1.29	1.24	1.48	1.53	1.63	1.62	1.53	1.56	1.56	1.25	1.37	1.45	1.36	1.50	1.67	1.19	1.46	1.40	1.71	1.52	1.54	1.79	1.59	1.31	1.42
Cafeteria	-5	1.02	0.98	1.12	0.93	1.07	1.19	1.00	0.87	0.85	0.97	1.06	0.95	1.02	0.75	1.17	0.84	0.84	0.98	0.98	0.84	0.85	0.88	1.02	0.78	0.88
	0	1.14	1.35	1.36	1.53	1.50	1.51	1.40	1.47	1.38	1.11	1.36	1.36	1.27	1.21	1.50	1.29	1.13	1.31	1.44	1.27	1.33	1.35	1.38	1.02	1.32
	5	1.51	1.37	1.41	1.49	1.58	1.83	1.50	1.39	1.70	1.42	1.63	1.50	1.37	1.44	1.60	1.32	1.54	1.42	1.42	1.44	1.50	1.43	1.29	1.31	1.40
	10	1.60	1.69	1.52	1.50	1.74	1.53	1.70	1.70	1.47	1.43	1.49	1.71	1.46	1.55	1.72	1.33	1.50	1.55	1.82	1.51	1.67	1.56	1.56	1.30	1.61
Multi-talker	-5	0.50	0.72	0.33	0.20	0.99	1.02	0.39	0.54	0.70	0.56	0.33	0.27	0.47	0.68	0.35	0.75	0.49	0.51	0.42	0.47	0.60	0.21	0.73	0.37	0.11
	0	0.93	0.59	0.90	1.03	1.44	1.23	1.02	1.01	0.99	0.99	0.91	1.14	1.12	0.96	1.25	1.11	0.90	0.98	1.28	0.91	1.00	1.00	1.16	0.54	0.85
	5	1.24	1.26	1.18	1.13	1.56	1.38	1.26	1.31	1.48	1.19	1.15	1.38	1.30	1.22	1.39	1.28	1.10	1.18	1.48	1.32	1.32	1.36	1.43	1.13	1.36
	10	1.51	1.43	1.30	1.36	1.65	1.75	1.54	1.54	1.62	1.49	1.50	1.67	1.63	1.48	1.40	1.43	1.47	1.32	1.56	1.55	1.41	1.47	1.66	1.42	1.51

*Processed by the wavelet thresholding of multitaper estimators noise reduction method in the HiRes 120 vocoder.

Table 2. The perceptual evaluation of speech quality scores of 25 sentences from difference background environments at signal to noise ratios (SNRs) of -5, 0, 5, and 10 dB*

Noise	SNR	Sentences																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Air-conditioner	-5	1.06	1.25	0.51	2.29	0.91	0.84	1.08	0.85	1.09	1.19	1.26	1.16	1.12	1.00	1.19	0.67	0.91	1.83	1.12	0.98	0.54	0.46	1.08	0.83	0.23
	0	1.33	1.07	0.52	1.05	1.06	1.31	0.99	0.85	1.28	1.24	1.34	0.94	0.80	0.81	1.28	0.89	0.84	0.98	1.25	1.21	0.77	0.69	1.21	0.57	0.65
	5	1.60	1.27	1.03	1.27	1.54	1.63	1.34	1.24	1.54	1.41	1.59	1.33	1.16	1.15	1.43	1.38	1.12	1.19	1.48	1.41	1.27	1.32	1.49	1.08	1.14
	10	1.75	1.50	1.31	1.52	1.65	1.86	1.62	1.60	1.84	1.52	1.82	1.58	1.47	1.42	1.55	1.38	1.41	1.58	1.62	1.63	1.54	1.61	1.61	1.25	1.26
Cafeteria	-5	1.31	1.53	1.30	1.26	1.07	1.07	0.67	1.39	1.13	0.82	0.91	1.13	0.73	0.86	0.67	0.92	0.82	1.55	0.84	1.15	0.97	0.70	1.31	1.04	0.76
	0	1.43	1.27	0.86	1.46	1.40	1.36	1.29	1.32	1.46	1.25	1.69	1.20	0.87	1.16	0.94	1.15	1.12	1.29	1.37	1.29	1.69	0.92	1.46	0.71	0.40
	5	1.58	1.41	0.88	1.50	1.66	1.69	1.50	1.46	1.69	1.44	1.70	1.39	1.36	1.29	1.65	1.43	1.33	1.42	1.46	1.56	1.35	1.22	1.55	0.90	1.23
	10	1.64	1.52	1.41	1.54	1.74	1.85	1.67	1.56	1.85	1.56	1.78	1.56	1.54	1.48	1.61	1.58	1.40	1.61	1.70	1.66	1.51	1.50	1.60	1.13	1.39
Multi-talker	-5	1.09	1.11	0.53	0.28	1.10	1.14	0.97	1.31	1.13	0.61	1.42	0.43	0.39	0.94	1.27	0.64	0.90	1.88	0.85	1.03	0.61	0.87	1.23	0.71	0.73
	0	1.16	1.11	1.01	1.01	1.40	1.22	1.06	1.35	1.61	1.13	1.17	0.93	0.94	1.04	1.36	1.30	1.06	1.33	1.08	1.21	0.74	0.94	1.47	0.62	0.50
	5	1.52	1.38	1.13	1.42	1.57	1.61	1.34	1.24	1.56	1.35	1.62	1.51	1.10	1.14	1.47	1.45	1.16	1.39	1.40	1.37	0.56	1.27	1.57	1.03	1.01
	10	1.68	1.52	1.32	1.45	1.72	1.76	1.56	1.57	1.81	1.51	1.77	1.62	1.46	1.48	1.57	1.59	1.31	1.48	1.53	1.60	1.54	1.62	1.80	1.30	1.28

*Processed by the geometric approach to spectral subtraction noise reduction method in the HiRes 120 vocoder.

Table 3. The perceptual evaluation of speech quality mean scores of sentences from difference background environments at difference signal to noise ratio (SNR)*

SNR	WTME				GASS			
	Air-conditioner	Cafeteria	Multi-talker	Overall	Air-conditioner	Cafeteria	Multi-talker	Overall
-5	0.80	0.95	0.51	0.76	1.02	1.04	0.93	0.99
0	1.23	1.33	1.01	1.19	1.00	1.21	1.11	1.11
5	1.44	1.47	1.30	1.40	1.34	1.43	1.33	1.36
10	1.48	1.57	1.51	1.52	1.56	1.58	1.55	1.56

Values are presented as mean.
*Processed by the wavelet thresholding of multitaper estimators (WTME) and geometric approach to spectral subtraction (GASS) noise reduction methods in the HiRes 120 vocoder.

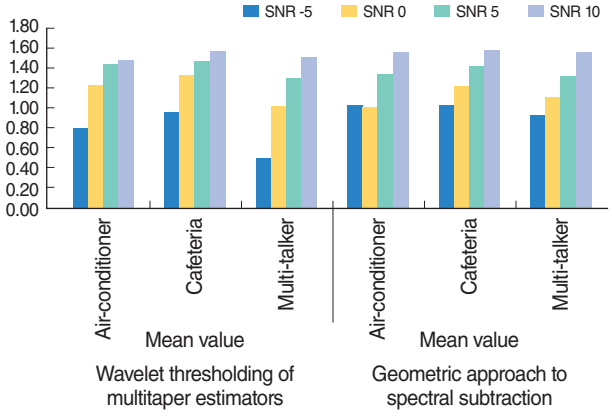

Fig. 4. The mean value of each signal to noise ratio (SNR) based on difference background environment.

that processed by the GASS noise reduction method in the HiRes 120 vocoder is showed in Table 2. Table 3 showed the PESQ mean scores of sentences from difference background environments (air-conditioner, cafeteria and multi-talker) at difference SNR that processed by the WTME and GASS noise reduction methods in the HiRes 120 vocoder. Fig. 4 shows the mean value of each SNR based on difference background environment whereas Fig. 5 shows the overall PESQ mean value of each SNR.

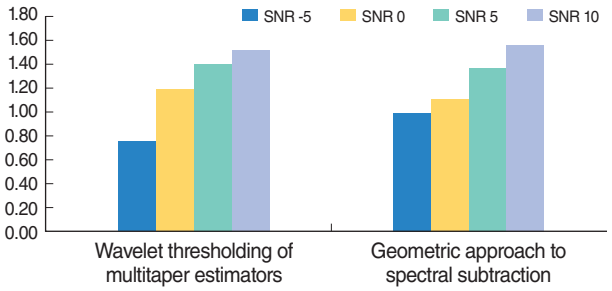

Fig. 5. The overall mean value of each signal to noise ratio (SNR).

DISCUSSION

Based on the results shown in Fig. 4, the PESQ mean of each difference background environments mostly score between 1.0 to 1.6 and only few PESQ mean scores are below 1.0, which are poor in sound quality. The WTME fail to improve the sound quality when the SNR is -5 dB under air-conditioning, cafeteria and multi-talker background environments condition. The GASS shows poor sound quality when the SNR is -5 dB under the multi-talker background environment condition. Both noise reduction methods have poor sound quality in multi-talker environment when the SNR is -5 dB because the noise estimator in both methods unable to estimate the correct speech or noise activity when the noise energy is bigger than speech signal energy, in which the noise might appear to be similar to "speech" signal.

Fig. 4 shows that both noise reduction methods are able to perform better sound quality when the SNR values increase. There is no significant difference between the sound quality performances in both methods, except for WTME in the multi-talker background environment. The overall mean value of each SNR (Fig. 5) indicated that there is no significant difference between the performances of both methods, except for WTME when SNR is -5dB.

To conclude, there is no significant difference between the overall performance of sound quality by both methods, but the GASS method is slightly better than the WTME.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Boll S. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust. 1979 Apr;27(2):113-20.
2. Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 1979 Apr 2-4; Washington, DC, USA. p. 208-11.
3. Hu Y, Loizou PC. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Trans Speech Audio Process. 2004 Jan;12(1):59-67.
4. Lu Y, Loizou PC. A geometric approach to spectral subtraction. Speech Commun. 2008 Jan;50(6):453-66.
5. Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans Speech Audio Process. 2001 Jul;9(5):504-12.
6. Rix AW, Beerends JG, Hollier MP, Hekstra AP. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 2001 May 7-11; Salt Lake City, UT, USA. p. 749-52.
7. International Telecommunication Union. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [Internet]. Geneva, Switzerland: International Telecommunication Union; 2001 [cited 2011 Nov 30]. Available from: http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.862-200102-I!!SOFT-ZST-E&type=items.
8. Choi CT, Hsu CH, Tsai WY, Lee YH. A vocoder for a novel cochlear implant stimulating strategy based on virtual channel technology. In: 13th International Conference on Biomedical Engineering; 2008 Dec 3-6; Singapore. p. 310-3.
9. Riedel KS, Sidorenko A. Minimum bias multiple taper spectral estimation. IEEE Trans Signal Process. 1995 Jan;43(1):188-95.
10. Walden AT, Percival DB, McCoy EJ. Spectrum estimation by wavelet thresholding of multitaper estimators. IEEE Trans Signal Process. 1998 Dec;46(12):3153-65.