THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.23, NO.4E 2004. 12 pp. 125~133

Critical Banded Wavelet Packet-Based Spectral Subtractions for Speech Enhancement

Sungwook Chang*, Sung-il Yang*
*School of Electrical and Computer Engineering, Hanyang University
(Received July 29 2004; revised November 29 2004; accepted December 1 2004)

Abstract

In this paper, we propose a critical banded wavelet packet-based spectral subtraction for speech enhancement. Critical banded wavelet packet, which reflects the human auditory system, may lead to minimization of intelligibility loss and quality improvement of the enhanced speech in the spectral domain, when combined with an appropriate spectral subtraction gain function. The proposed method shows better performance than the conventional one in comparative assessments. We also show that, for effective evaluation of enhanced speech, it is essential to consider the characteristics of speech quality measures.

***Keywords:** Speech Enhancement, Critical Banded Wavelet Packet, Spectral Subtraction, Speech Quality Measure*

I. Introduction

In speech enhancement, it is essential to minimize the intelligibility loss of speech while its quality is enhanced. One of the methods to enhance the speech quality is a signal subspace approach proposed by Ephraim and Van Trees[1]. Based on their approach, perceptual signal subspace approaches, which would enhance the perceptual performance, were proposed[2-4]. There are also two kinds of adaptive wavelet packet-based spectral subtraction (AWPSS) methods proposed by our group[5-7]. These focus on the perceptual performance through adaptive spectral subtraction weight estimation procedure for each frame. However, AWPSS methods have a drawback that the performance of adaptive basis may be deteriorated in low SNR environment.

In this paper, we propose new kinds of critical banded wavelet packet-based spectral subtraction (CB-WPSS) methods to avoid poorly-adapted basis problem in low SNR environment. The critical banded wavelet packet, which reflects the human auditory system, may lead to both minimization of intelligibility loss and quality improvement of the speech, when it would be combined with an appropriate spectral gain function.

In speech enhancement, another important issue is to have an effective assessment method. It frequently happens that evaluation by objective measures (SNR, segmental SNR, etc.) and one by subjective ones (listening test, spectrogram, mean opinion score, etc.) do not coincide. The cases when this problem occurs are known as follows; (1) a case of low correlation between objective measure and subjective one (2) a case where characteristics of the designed speech enhancement method and the quality measure are not properly considered. To avoid the problem, various works[12-20] on effective speech assessment have been reported. Essentially, they consider an objective measure which has relatively high correlation with subjective one.

In this paper, we use objective measures considering the correlation with subjective ones, for effective comparative assessment between the conventional speech enhancement method[2] and the proposed CB-WPSS ones. And we

Corresponding author: Sungwook Chang (schang@ihanyang.ac.kr)
Robust Speech Processing Laboratory, Hanyang University, 17, Haengdang-dong, Seongdong-gu, Seoul, 133-791, Korea

evaluate new speech enhancement methods and show the dependence of performance on selected speech quality measures.

This paper consists of the following sections. We describe the proposed CB-WPSS methods in Section II. In Section III, we present speech quality measures in detail. We also discuss dependency of speech quality measure on speech enhancement method. In Section IV, we present results of comparative assessments between the proposed CB-WPSS methods and the conventional methods. Finally, in Section V, we provide conclusions.

II. Critical Banded Wavelet Packet-Based Spectral Subtractions

Recently, two kinds of AWPSS methods[5-7] have been proposed by our group. These methods extract an adaptive spectral subtraction weight using adaptive wavelet packet and apply the weight to spectral subtractions. They provide a proper spectral subtraction gain for non-uniform frequency resolution like adaptive wavelet packet structure.

Though AWPSS methods show good performance in various noisy environments, they have a shortcoming that the performance of adaptive basis would be deteriorated in low SNR environment. For that reason, we propose two new kinds of CB-WPSS methods to avoid the drawback.

In this section, we introduce the designed critical banded wavelet packet decomposition (CB-WPD) and two kinds of CB-WPSS methods.

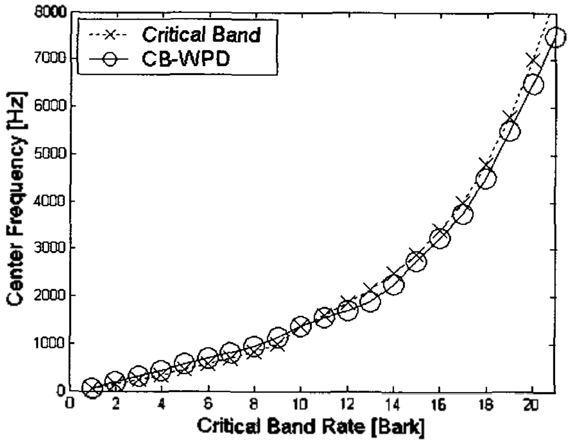


Figure 1. Approximation of the Critical Band by CB-WPD.

2.1. Critical Banded Wavelet Packet Decom-position

Let $\{\psi_n(t) : n \in \mathbf{Z}_+\}$ denote a dyadic wavelet packet family, and let $T \subset \{(l, n) : 0 \le l < L, 0 \le n < 2^l\}$ represent the terminal nodes of a wavelet packet tree. Here, l, n, and L denote level, node, and the deepest level index on the tree, respectively. And dyadic interval $\{I_{l,n} \subset \quad : (l, n) \in T\}$ of each terminal node is defined by $I_{l,n} \equiv [2^{-l}n, 2^{-l}(n+1)[$. A terminal node $(l, n) \in T$ is associated with a subband whose center frequency and bandwidth are roughly given by

$$f_{l,n} = 2^{-l}(n+0.5)\cdot(f_s/2) \tag{1}$$

$$\Delta f_{l,n} = 2^{-l}\cdot(f_s/2) \tag{2}$$

where f_s is the sampling frequency[8-9].

To construct a CB-WPD, we carry out a full 6 level (L=6) wavelet packet decomposition. Then we select nodes whose distance between their center frequencies is approximately 1 Bark. Due to the nature of dyadic wavelet packet, we use the approximated critical bands instead of exact ones as shown in Fig. 1. The constructed CB-WPD provides 21 frequency bands for 16kHz sampled speech. Center frequencies and bandwidths of CB-WPD are given in Table 1.

2.2. Critical Banded Wavelet Packet-Based

Table 1. Center Frequencies fc and Bandwidths Δf_G for Critical Band and CB-WPD.

Bark	Critical Band		CB-WPD	
	f_c(Hz)	Δf_G(Hz)	f_c(Hz)	Δf_G(Hz)
1	50.0	100	62.5	125
2	150.0	100	187.5	125
3	250.0	100	312.5	125
4	350.0	100	437.5	125
5	450.0	110	562.5	125
6	570.0	120	687.5	125
7	700.0	140	812.5	125
8	840.0	150	937.5	125
9	1000.0	160	1125.0	250
10	1170.0	190	1375.0	250
11	1370.0	210	1562.5	125
12	1600.0	240	1687.5	125
13	1850.0	280	1875.0	250
14	2150.0	320	2250.0	500
15	2500.0	380	2750.0	500
16	2900.0	450	3250.0	500
17	3400.0	550	3750.0	500
18	4000.0	700	4500.0	1000
19	4800.0	900	5500.0	1000
20	5800.0	1100	6500.0	1000
21	7000.0	1300	7500.0	1000

Spectral Subtraction: Type I

An adaptive spectral subtraction weight, adapted to background noise, is estimated with CB-WPD. The CB-WPD has a structure of non-uniform frequency resolution. Although the structure of CB-WPD is helpful to enhance the intelligibility, it may have an undesired effect introduced by some nodes which have minimum or maximum value on CB-WPD tree. To resolve the problem, we use a log scaled geometric mean of average magnitude spectrum of noisy speech (*NSGM*) and one of an estimated noise magnitude spectrum (*NGM*) in each *k*th frame,

$$NSGM(k) = \log\left(\prod_{n=0}^{N_{node}-1} |X_k(n)|\right)^{\frac{1}{N_{node}}} \quad (3)$$

$$NGM(k) = \log\left(\prod_{n=0}^{N_{node}-1} |\hat{N}_k(n)|\right)^{\frac{1}{N_{node}}} \quad (4)$$

where N_{node} is the number of node on CB-WPD tree. And N_{node} is the number of critical band, in our work, 21-level as shown in Table 1. $X_k(n)$ is average magnitude spectrum of noisy speech at *n*th node on CB-WPD tree. And $\hat{N}_k(n)$ is average magnitude spectrum of noise estimated by Hirsch's method[10].

Geometric noisy signal to noise ratio (*GNNR*) is calculated as follows;

$$GNNR = \frac{\sum_{k=0}^{N_{frame}-1} NSGM(k)}{\sum_{k=0}^{N_{frame}-1} NGM(k)} \quad (5)$$

where *k* is the frame index and N_{frame} the number of frame. Now we estimate an adaptive spectral subtraction weight in each *k*th frame,

$$\xi(k) = \frac{GNNR}{\tau} \cdot \left(\frac{NSGM_{max} - NSGM(k)}{NSGM_{max} - NSGM_{min}} \right) \quad (6)$$

where $NSGM_{max}$ and $NSGM_{min}$ denote the maximum and minimum *NSGM* over all frames, respectively. And τ is an auxiliary parameter with $2.0 \le \tau \le 3.0$.

Finally spectral subtraction gain $G_k(n)$ is estimated at *n*th node with $\xi(k)$ in each *k*th frame.

$$G_k(n) = \begin{cases} \text{if } |X_k(n)| \ge (1+\xi(k)) \cdot |\hat{N}_k(n)| & \sqrt{1 - \frac{(1+\xi(k)) \cdot |\hat{N}_k(n)|}{|X_k(n)|}} \\ \text{else} & \eta \cdot \sqrt{\frac{|\hat{N}_k(n)|}{|X_k(n)|}} \end{cases} \quad (7)$$

where $0 \le \eta \le 0.1$.

2.3. Critical Banded Wavelet Packet-Based Spectral Subtraction: Type II

We consider average of the 1st order differential power spectrums at the *n*th node on CB-WPD tree as another approach to estimate an adaptive spectral subtraction gain in the structure of non-uniform frequency resolution as follows;

$$D(n) = \frac{1}{N_{nodesize} - 1} \sum_{m=0}^{N_{nodesize}-2} \left| |c_n(m)|^2 - |c_n(m+1)|^2 \right| \quad (8)$$

where $N_{nodesize}$ is the width of node and $c_n(m)$ the *m*th coefficient in *n*th node of CB-WPD tree. Then we estimate noise intensity (*NI*) for each frame.

Estimation of Noise Intensity

[Step 1]: Initialize *NI* and node index on CB-WPD tree, $NI = 0$, $n = 0$.

[Step 2]: If *n* is larger than the largest node index, then exit.

[Step 3]: If $\left(1 \le \sqrt{|\hat{N}(n)|/D(n)}\right)$, then $NI = NI + 1$, $n = n + 1$, and go to [Step 2].

where $\hat{N}(n)$: average magnitude spectrum of noise estimated by Hirsch's method[10].

Next speech dominant indicator (*SDI*) is estimated as follows;

$$SDI = 10\log_{10}\left(\left| \left(\prod_{n=1}^{N_{node}} D(n) \right)^{\frac{1}{N_{node}}} \right| \right) \cdot \lambda^{\gamma} \quad (9)$$

$$\lambda = \log_{10}\left(\frac{N_{node}}{NI} \right) \quad (10)$$

where N_{node} : the number of node on CB-WPD tree

λ^{γ} : weight depending on NI

γ : experimental exponent parameter, an experimentally determined value with $1 \le \gamma \le 3$.

The estimated SDI is used to estimate an adaptive spectral subtraction weight as shown in Eq. 11,

$$\beta_k = \left(\frac{SDI_{\max} - SDI_{\min}}{(SDI_k - SDI_{\max}) \cdot \mu + SDI_{\max} - SDI_{\min}} \right)^2 \qquad (11)$$

where k and μ denote frame index and auxiliary parameter with $0.1 \le \mu \le 0.3$, respectively.

Finally adaptive spectral subtraction gain is estimated at the nth node in the kth frame, with the estimated adaptive spectral subtraction weight. Then we can apply it to spectral subtraction for speech enhancement.

$$G_k(n) = \begin{cases} \text{if } \beta_k \cdot \sqrt{\hat{N}_k(n)} < \sqrt{D_k(n)} \\ \quad \sqrt{1 - \beta_k \left(\frac{\hat{N}_k(n)}{D_k(n)} \right)^{0.5}} \\ \text{else} \\ \quad \eta \cdot \sqrt{\left(\frac{\hat{N}_k(n)}{D_k(n)} \right)^{0.5}} \end{cases} \qquad (12)$$

where $0 \le \eta \le 0.01$.

The reference [6] and [7] supply more detailed account about the two kinds of estimation methods used in this paper for adaptive spectral subtraction gain.

III. Speech Quality Measures

Although subjective measures are useful to evaluate the performance of speech enhancement, there is some restriction. That is, if listener group could find the difference in speech quality, by the subjective measure, it can be used. Furthermore, subjective testing requires significant time and personnel resources. We therefore need to consider the class of objective measures that are reliable [16]. As shown in Table 2, the correlation of objective measures with a subjective one is presented by some researchers.

Table 2. Comparison of the Average Correlation Coefficient $|\rho|$ Between Objective and Subjective Quality [12,13]

Objective Speech Quality Measure	$\lvert\rho\rvert$
Time domain measure	
SNR	0.24
Segmental SNR	0.77
LPC-based measure	
Log Area Ratio	0.62
Log Likelihood Ratio	0.50
Itakura Distance	0.59
Frequency Variant Log Spectral Distance	
LPC-based	0.68
Filter bank	0.72
Weighted Slope Spectral Distance	0.74

When we choose an objective measure, the correlation with subjective one, the type of residual noise present in the enhanced speech, and the characteristics of noise and speech must be considered. For example, measures such as those based on SNR, which is a distortion measure based on sample-by-sample differences in original and processed time waveforms, do not provide a meaningful measure of performance when two waveforms differ in their phase spectra. And if a speech enhancement method is applied to speech or speaker recognition system, LPC-based measure or frequency variant log spectral distance can be important measure even though it is less correlated with subjective one.

In this paper, we take segmental SNR, LP model-based log area ratio, weighted spectral slope and ITU-T P.862 PESQ (perceptual evaluation of speech quality) to evaluate the speech quality in various aspects.

Now let us describe the measures used in our evaluation. The signal-to-noise ratio (SNR) is one of the most common objective measures to assess the speech quality. SNR may not be a good objective measure because speech is natively non-stationary signal and noise has different perceptual values depending on the ambient signal level. To overcome the shortcoming of SNR, segmental SNR (SegSNR) was proposed by Noll [11] as follows;

$$d_{SegSNR} = \frac{10}{M} \sum_{m=0}^{M-1} \log \frac{\sum_{n=Nm}^{Nm+N-1} s_{clean}^2(n)}{\sum_{n=Nm}^{Nm+N-1} \left[s_{enhanced}(n) - s_{clean}(n) \right]^2} \qquad (13)$$

where N is the frame size and M is the number of frames. SegSNR seems to correspond to the auditory experience. Actually, it was reported that SegSNR is relatively highly correlated with subjective quality as shown in Table 2. The value $|\rho| = 1$ would indicate that the measure predicts mean opinion score (MOS) perfectly, while $|\rho| = 0$ means that it guesses MOS randomly.

We note that the SNR type measures (SNR, SegSNR) are meaningful only for the cases where the residual noise in the enhanced signal is seems to be additive. This can be seen in the denominator of Eq. (13) which requires the additivity of residual noise[14, 16]. Thus, it is required to consider the characteristic of residual noise which would depend on speech enhancement method when SNR type measure is used.

When SegSNR is used as an objective speech quality measure, limitation of upper and lower bound of SegSNR at each frame needs to be considered. That is, frames with SNRs above 35dB need to be considered to have 35dB because it is hard to distinguish perceptually the frames with SNR above 35dB and those with just 35dB SNRs. Likewise, during periods of silence, SNR values can become very negative since signal energies are small. These frames do not reflect the perceptual contributions to the signal. Therefore, a lower threshold is often set to provide a bound on frame-based SNR. In this paper, we select -10dB as recommended in [15].

As we have mentioned, SNR type measure do not provide a meaningful measure of evaluating an enhancement of speech when two waveforms differ in their phase spectra. Rather LP model-based measures might be the good ones. These measure the dissimilarity between sets of LP parameters from clean and enhanced speeches. In particular, it has been shown that among all LP model-based measures, log area ratio (LAR) has the highest correlation with subjective quality[12, 13, 15-19]. If $k(l)$ is the l^{th} reflection coefficient of P^{th} order LP model, LAR can be defined as follows;

$$d_{LAR} = \sqrt{\frac{1}{P}\sum_{l=1}^{P}\left(\log\frac{1+k_{clean}(l)}{1-k_{clean}(l)} - \log\frac{1+k_{enhanced}(l)}{1-k_{enhanced}(l)}\right)^2} \qquad (14)$$

Objective measures on auditory models[20] are also used in the field of speech application. Weighted spectral slope measure (WSSM) is an objective measure based on auditory model (i.e. critical band filter analysis). This measure finds a weighted difference between the spectral slopes in each band. The magnitude of each weight reflects whether the band is near a spectral peak or valley, and whether the peak is the largest in the spectrum. The WSSM for m^{th} frame in decibels is defined by

$$d_{wss}(m) = K_{spl}(K - \hat{K}) + \sum_{k=1}^{N_{CB}} w_{k,m}\left[\left|S_{k,m}\right| - \left|\hat{S}_{k,m}\right|\right]^2 \qquad (15)$$

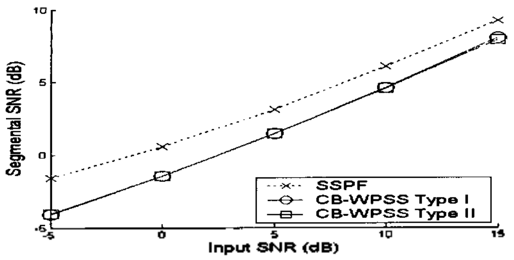
where K and $\hat{K}$ are related to overall sound pressure level of the original and enhanced speech, and K_{spl} is a parameter which can be varied to increase overall performance. Here $w_{k,m}$ is a weight and $S_{k,m}$ short term DFT. N_{CB} is the number of critical band. In this paper, 21 critical bands are used as shown in Table 1. Since WSSM is based on auditory model, it is more closely related to listener intelligibility than the other measures such as SNR based ones. Actually, WSSM shows high correlation value, 0.74, as shown in Table 2[15, 16].

Recently, ITU-T Recommendation P.862, PESQ[21] was developed as another human auditory model based on objective speech quality measure. In their own benchmark, it is reported that the correlations between objective and subjective scores were around 0.935 for both known and unknown data.

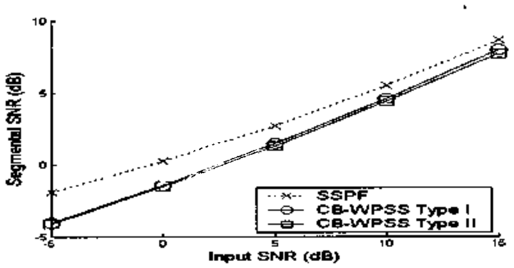
IV. Evaluation

To evaluate the performance of proposed CB-WPSS methods, comparative assessments were performed with signal subspace speech enhancement with perceptual post-filtering (SSPF)[2, 22] which is a kind of the perceptual signal subspace approaches. SSPF has a perceptual post-filter to smooth the enhanced speech. To design the filter, a noise correlation matrix and the masking threshold of speech is estimated.

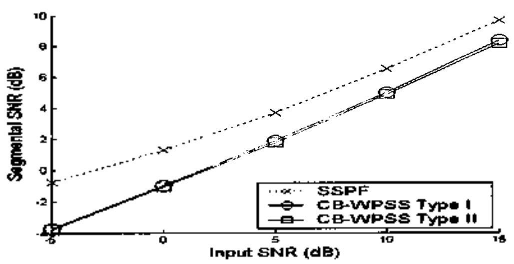
Various noise types, from Noisex-92 database, were used in our assessment: F-16 cockpit, factory, pink, Volvo car interior and white noises. The assessment results were averaged out using 20 different speeches taken from the TIMIT database. Half of the speeches were taken from male speakers, and the others from female speakers.



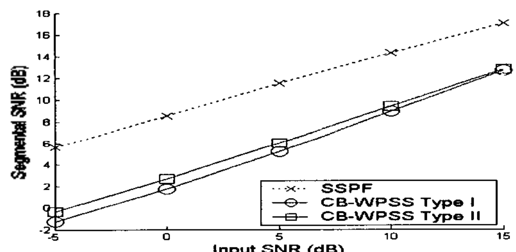
(a) F16 cockpit Noisy Environment



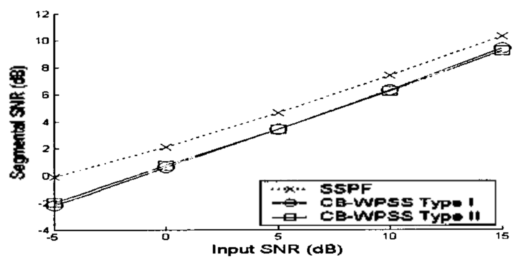
(b) Factory1 Noisy Environment



(c) Pink Noisy Environment

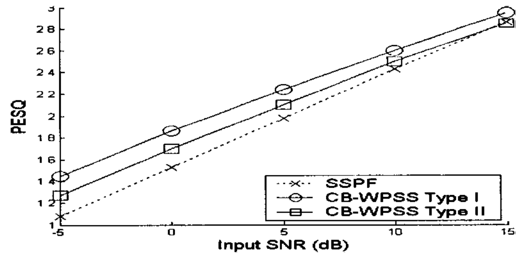


(d) Volvo Noisy Environment

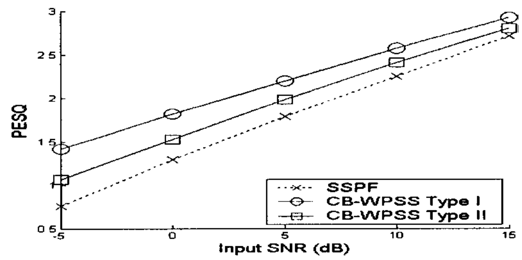


(e) White Noisy Environment

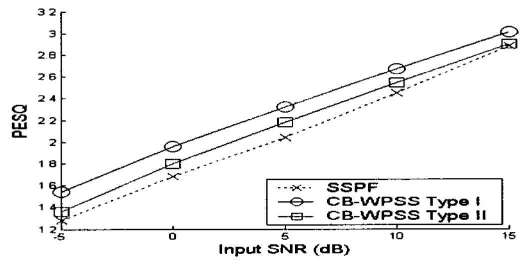
Figure 2. Average Segmental SNR for 20 speeches



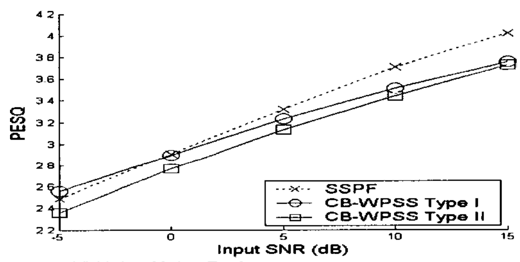
(a) F16 cockpit Noisy Environment



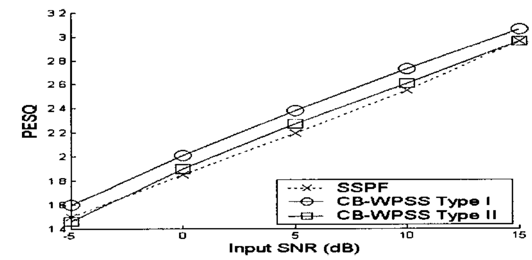
(b) Factory1 Noisy Environment



(c) Pink Noisy Environment



(d) Volvo Noisy Environment



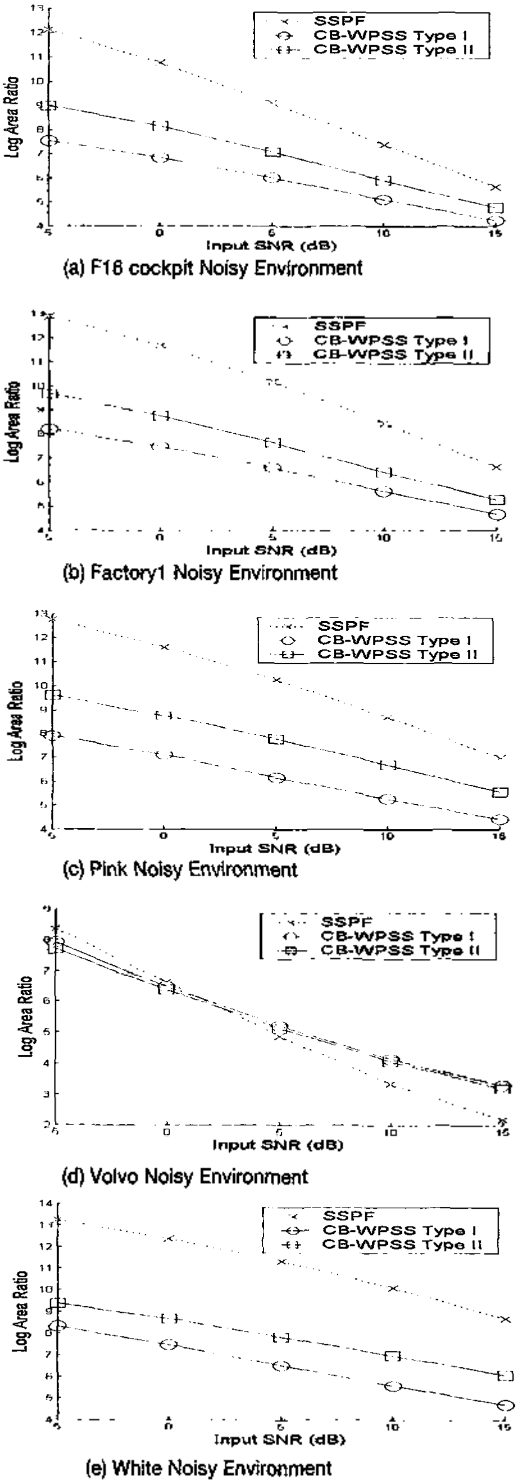
(e) White Noisy Environment

Figure 3. Average PESQ for 20 speeches

SegSNR, PESQ, LAR, and WSSM, which are highly correlated with the subjective measure, were employed to perform comparative assessments in various aspects. And speech quality assessment MATLAB toolbox[23] was used except PESQ.

The result of SegSNR assessment in various noisy environments is shown in Fig. 2. Unfortunately, SSPF shows better performance than the proposed CB-WPSS methods in all noisy environments. Especially, SSPF shows relatively good SegSNR performance in Volvo noisy environment where most of noise components are concentrated in low frequency area.

However, another comparative assessment with PESQ shows very interesting results in Fig. 3. Except for Volvo nosy environments, SegSNR and PESQ show the opposite result for the same enhanced speeches. The proposed CB-WPSS methods use only magnitude or power spectrum. There is no consideration for phase because human auditory system is insensitive to phase distortion [16, 24]. Furthermore, there is no perfectly symmetric wavelet basis because no wavelet system can be simultaneously



(a) F16 cockpit Noisy Environment

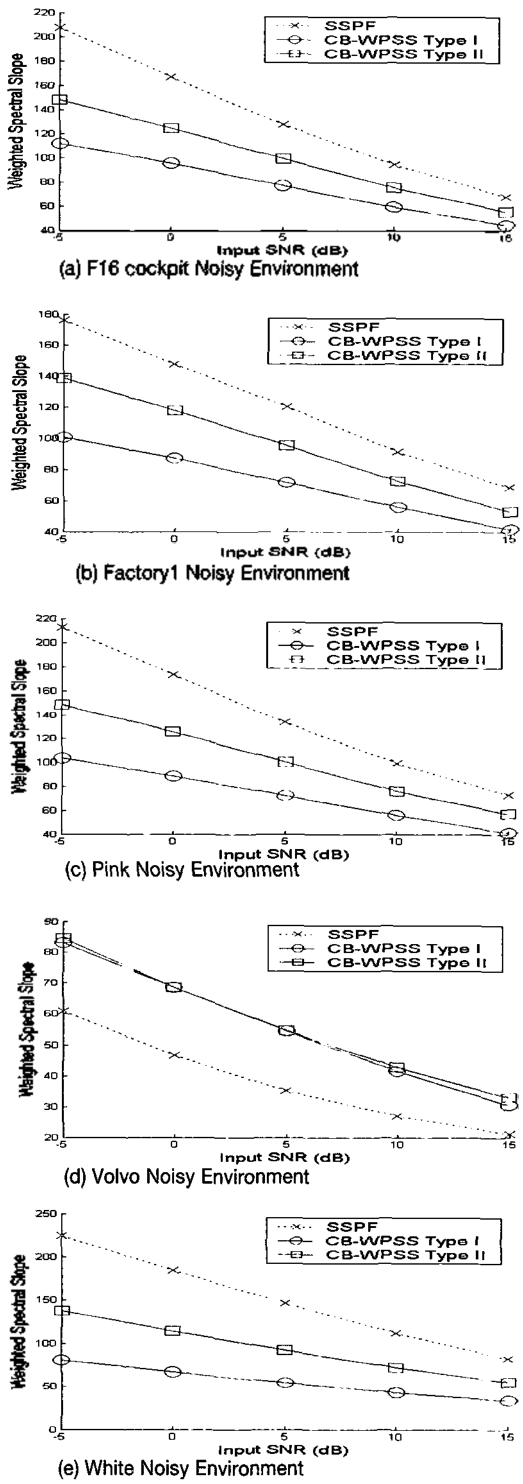
(b) Factory1 Noisy Environment

(c) Pink Noisy Environment

(d) Volvo Noisy Environment

(e) White Noisy Environment

Figure 5. Average Log Area Ratio for 20 speeches



(a) F16 cockpit Noisy Environment

(b) Factory1 Noisy Environment

(c) Pink Noisy Environment

(d) Volvo Noisy Environment

(e) White Noisy Environment

Figure 6. Average Weighted Spectral Slope for 20 speeches

compactly supported and symmetric, except for Haar wavelet system. Thus, CB-WPD cannot provide linear phase characteristic. For that reason, the residual noise of CB-WPSS methods may be shown as a speech-correlated one in SegSNR. Since SegSNR is meaningful only for the case where the residual noise seems to be additive, the result of SegSNR can be unreliable for the proposed methods.

For more evident comparative assessment, we evaluate the enhanced speech by LAR and WSSM which have relatively high correlation with subjective measure. In Fig. 4 and Fig. 5, LAR and WSSM show a relative merit of CB-WPSS methods more clearly.

In LAR assessment, SSPF has largest distortion but CB-WPSS Type-I has relatively smallest one. The result of WSSM also shows similar results with PESQ and LAR in Fig. 5.

Additionally, we need to note that SSPF shows good performance in narrow band noisy environment like Volvo noisy one but bad performance in the other environments. Actually, this phenomenon coincides with the reported result in [2]. Experimentally, it is concluded that the drawback is caused by poor noise estimation and pre-whitening filter in wide band noisy environment. Consequently, SSPF may be restricted in application environment. This is an undesired restriction when we consider that most of the speech applications may be operated in various noisy environments and the characteristic of noise is not known previously.

V. Conclusions

To assess the speech quality in various aspects, we take SegSNR, LP model based LAR, WSSM, and PESQ. In most of the noisy environments, the proposed CB-WPSS methods show better performance than SSPF. And the comparative assessments with LAR and PESQ show similar or better performance in low SNR (<5dB) environments even in narrow band noisy one.

Additionally, we note that an assessment without consideration of the characteristic of residual noise cannot show reliable result even though a measure which is highly correlated with a subjective measure is used. In other words, SegSNR, with an assumption that the residual noise seems to be additive, shows the opposite results to the other measures. From this result, we can see that there is a dependency of speech quality measure on speech enhancement method, because the characteristic of residual noise is decided by speech enhancement method. Finally, we can conclude that it is required to consider the characteristic of speech quality measure when we assess a speech enhancement method.

References

1. Y. Ephraim and H. L. Van Trees, "A Signal Subspace Approach for Speech Enhancement," IEEE Trans. on Speech and Audio Processing, 3(4), pp. 251-266, July, 1995.
2. M. Klein and P. Kabal, "Signal Subspace Speech Enhancement with Perceptual Post-filtering," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp. 537-540, May, 2002.
3. F. Jabloun and B. Champagne, "A Perceptual Signal Subspace Approach for Speech Enhancement in Colored Noise," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 1, pp. 569-572, 2002.
4. Y. Hu and P. Loizou, "Perceptual Weighting Motivated Subspace based Speech Enhancement Approach," Proc. of Int. Conf. on Spoken Language Processing, pp. 1797-1800, 2002.
5. S. Chang, S. Jung, Y. Kwon, and S. Yang, "Speech Enhancement using Wavelet Packet Transform," Proc. of Int. Conf. on Spoken Language Processing, pp. 1809-1812, 2002.
6. S. Chang, Y. Kwon, S. Jung, S. Yang and K. Lee, "Speech Enhancement using Level Adapted Wavelet Packet with Adaptive Noise Estimation," The Journal of the Acoustical Society of Korea, 22(2E), pp. 87-92, 2003.
7. S. Chang, Y. Kwon, S. Jung and S. Yang, "Adaptive Wavelet based Speech Enhancement with Robust VAD in Non-Stationary Noise Environment," The Journal of the Acoustical Society of Korea, 22(4E), 2003.
8. M. V. Wickerhauser, Adapted Wavelet Analysis from Theory to Software, AK Peters, 1994.
8. I. Cohen, "Enhancement of Speech using Bark-Scaled Wavelet Packet Decomposition," EUROSPEECH 2001, pp. 3-7, 2001.
9. H. G. Hirsch, "Estimation of Noise Spectrum and its Application to SNR Estimation and Speech Enhancement," Technical Report TR-93-012, International Computer Science Institute, Berkeley, USA, 1993.
10. P. Noll, "Adaptive Quantization in Speech Coding Systems," Proc. Int. Zurich Seminar on Digital Communications, pp. B3.1-B3.6, Oct., 1974.
11. S. R. Quackenbush, T. P. Barnwell, M. A. Clements, Objective Measures of Speech Quality, Prentice-Hall, NJ, 1988.
12. S. Wang and A. Gersho, "An Objective Measure for Predicting Subjective Quality of Speech Coders," IEEE Journal on Selected Areas in Communications, 10(5), June, 1992.

13. D. G. Jamieson, L. Deng, M. Price, V. Parsa and J. Till, "Interaction of Speech Disorders with Speech Coders: Effects on Speech Intelligibility," Proc. of Int. Conf. on Spoken Language Processing, pp. 737-740, 1996.
14. J. H. L. Hansen and B. L. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," Proc. of Int. Conf. on Spoken Language Processing, pp. 2819-2822, 1998.
15. J. Deller, J. Proakis, J. H. L. Hansen, Discrete-Time Processing of Speech Signals, McMillan Series for Prentice Hall, New York, NY, 1993.
16. T. P. Barnwell and W. D. Voiers, "An analysis of objective measures for user acceptance of voice communication systems," DCA Final Technical Report, No. DCA100-78-C-0003, Sept. 1979.
17. T. P. Barnwell, M. A. Clements, S. R. Quackenbush et. al, "Improved objective measures for speech quality testing," DCA Final Technical Report, No. DCA100-83-C-0027, Sept. 1984.
18. T. P. Barnwell, "Improved objective quality for low bit speech compression," National Science Foundation, Final Technical Report, ECS-8016712, 1985.
19. D. Klatt, "Prediction of Perceived Phonetic Distance from Critical Band Spectra: A First Step," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp. 1278-1281, 1982.
20. ITU-T Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ), International Telecommunication Union, Feb. 2001.
21. P. Kabal, "Demo and Matlab Toolbox for [2]," http://www.tsp.ece.mcgill.ca/Kabal/papers/2002/KleinC2002-demo/
22. B. Pellom, "Matlab toolbox for Objective Speech Quality Assessment," CSLU Robust Speech Processing Laboratory, http://cslr.colorado.edu/ rspl/rspl_software.html.
23. P. E. Papamichalis, Practical Approaches to Speech Coding, Prentice Hall, England Cliffs, NJ, 1987.

[Profile]

• Sungwook Chang

Sungwook Chang was born in Anyang, Kyunggi, Korea in 1971. He received his B.S. and M.S. degrees in Control and Instrumentation Engineering from Han-yang University, Korea, 1997 and 1999, respectively and his Ph.D. degree in Electronic, Electrical, Control and Instrumentation Engineering from Hanyang University, 2004. His current research interests include speech enhancement, wavelets, speech recognition, and bioinformatics. He is also a member of the Acoustical Society of Korea.

• Sung-il Yang

Sung-il Yang was born in Geosan, Chungbuk, Korea in 1956. He received his B.S. degree in Electronics Engineering with the greatest honors from Hanyang University, Seoul, Korea, 1984, and his M.S. and Ph.D. degrees in Electrical & Computer Engineering from the University of Texas, Austin, Texas, 1986 and 1989, respectively. Since 1990, he has been with Hanyang University and he is now a Professor at the School of Electrical & Computer Engineering. His current research interests include speech recognition, digital signal processing, and responsible technology. He is also a member IEEE, Korea Institute of Telematics and Electronics, and the Acoustical Society of Korea.