IMPROVEMENT OF MBSD BY SCALING NOISE MASKING THRESHOLD AND CORRELATION ANALYSIS WITH MOS DIFFERENCE INSTEAD OF MOS

Wonho Yang and Robert Yantorno

Speech Processing Lab
Electrical & Computer Engineering, Temple University, Philadelphia, PA 19122-6077
wonho@astro.temple.edu, ryantorn@nimbus.temple.edu
http://nimbus.temple.edu/~ryantorn/speech/

Abstract

The Modified Bark Spectral Distortion (MBSD), used for an objective speech quality measure, was presented previously [1][2]. The MBSD measure estimates speech distortion in the loudness domain taking into account the noise masking threshold in order to include only audible distortions in the calculation of the distortion measure. Preliminary simulation results have shown improvement of the MBSD over the conventional BSD. In this paper, the performance of the MBSD is improved by scaling noise masking threshold and comparing it to ITU-T Recommendation P.861 [3] and MNB [4] measures. Correlation analysis with MOS difference instead of MOS has been examined in order to evaluate objective speech quality measures.

1. Introduction

Development of an objective speech quality measure that correlates well with subjective speech quality measures has been considered important because subjective tests are expensive and time-consuming. Even though none of current objective speech quality measures can replace subjective quality measures, a good objective speech quality measure would be a valuable assessment tool for speech coder development, speech codec deployment on communication systems, and even for speech codec selection. In fact, various types of objective speech quality measures have been used to improve speech quality in Analysis-By-Synthesis (ABS) speech coders [5].

Among the various different objective speech quality measures, we have been interested in the perceptual distortion measures such as Bark Spectral Distortion (BSD) [6] and Perceptual Speech Quality Measure (PSQM) [7]. These measures transform the speech signal into a perceptually relevant domain incorporating psychoacoustic responses. PSQM has been recommended as an objective quality measurement of telephone-band speech codecs by ITU [3]. Since the development of the BSD, it has become a good candidate for a highly correlated objective quality measure, according to several researchers [8][9][10]. The BSD measure is based on the assumption that speech quality is directly related to speech loudness, which is a psychoacoustical term, defined as the magnitude of auditory sensation. The BSD measure calculates the average squared Euclidean distance of estimated loudness of the original and the coded utterances to estimate the distortion in the coded speech. In order to calculate loudness, the speech signal is processed using results of psychoacoustic measurements, which include critical band analysis, equal-loudness preemphasis and intensity-loudness power law [6].

Even though the conventional BSD measure showed a relatively high correlation with Mean Opinion Score (MOS) – the most popular subjective speech quality measure - there are areas for possible improvement. Motivated by the transform coding of audio signals, which uses the noise masking threshold [11], the MBSD measure has incorporated this concept of a noise masking threshold into the conventional BSD measure, where any distortion below the noise masking threshold is not included in the BSD measure. This new addition of the noise masking threshold replaces the empirically derived distortion threshold value used in the conventional BSD [6]. The concept of a noise masking threshold was also used to improve speech quality in coder development [12]. It was shown that coding gain could be obtained with no loss of speech quality, by transmitting only spectral samples above the noise masking threshold. This implies that the noise below the noise masking threshold is not perceptible. Therefore, the noise spectral components below the noise masking threshold are excluded in the calculation of the MBSD measure because these components are considered inaudible.

Precisely speaking, the use of the psychoacoustically derived noise masking threshold has not been validated for speech. The psychoacoustic results are based on steady-state signals such as sinusoids rather than speech signals which contain a series of tones. Consequently, noise masking threshold taken directly from the psychoacoustics literature may not be appropriate for estimating distortion in speech signals. As a first step, we have examined the performance of the MBSD by scaling the noise masking threshold.

In this paper, we describe the MBSD measure and show the effect of noise masking threshold. The performance of the MBSD is improved by scaling noise masking threshold and compared to other measures such as ITU-T Recommendation P.861 and MNB. The correlation analysis with the MOS difference is discussed for the evaluation of objective quality measures.

2. MBSD Measure

The block diagram of the MBSD measure is shown in Fig. 1. There are three major processing steps: loudness calculation, noise masking threshold computation, and computation of MBSD. The loudness calculation transforms speech signal into the loudness domain. In order to transform speech into the loudness domain, the speech signal is processed in several steps: critical band analysis, equal-loudness preemphasis and intensity-loudness power law.

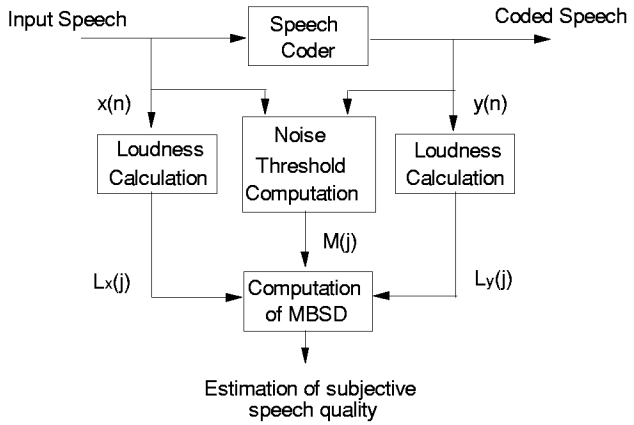

Figure 1. Block diagram of MBSD method

The noise masking threshold is estimated by critical band analysis, spreading function application and absolute threshold consideration [11]. The loudness of the noise masking threshold is compared to the loudness difference of the original and the coded speech to determine if the distortion is perceptible. When the loudness difference is below the loudness of the noise masking threshold, this loudness difference is imperceptible. Therefore, it is not included in the calculation of the MBSD.

In order to formally define the distortion for the MBSD, an indicator of perceptible distortion *M(i)* is introduced, where *i* is the *i-th* critical band. When the distortion is perceptible, *M(i)* is 1, otherwise *M(i)* is 0. The indicator of perceptible distortion is obtained by comparing the loudness to the noise masking threshold. The calculation of the MBSD is given by equation (1). Imperceptible distortion is excluded in the MBSD calculation when *M(i)* is zero. The MBSD is then defined as the average difference of estimated loudness which is perceptible.

$$MBSD = \frac{1}{N}\sum_{j=1}^{N}\left[\sum_{i=1}^{K} M(i)\left|L_x^{(j)}(i) - L_y^{(j)}(i)\right|\right] \qquad (1)$$

where,

N : number of frames processed

K : number of critical bands

$M(i)$: Indicator of perceptible distortion at *i-th* critical band

$L_x^{(j)}(i)$: Bark spectrum of *j-th* frame of original speech

$L_y^{(j)}(i)$: Bark spectrum of *j-th* frame of coded speech

3. Improvement of MBSD by Scaling Noise Masking Threshold

It has been found that there is an improvement of the performance of the MBSD by using noise masking threshold. However, since the noise masking threshold calculation is based on the psychoacoustics in which single tones and narrow band noises are usually used, noise masking threshold may not be very accurate if it is directly applied to signals such as speech, which contain a serires of tones. So, we examined the performance of the MBSD by scaling the noise masking threshold. In other words, *M(i)*, the indicator of perceptible distortion, is determined by comparing the loudness difference to the scaled noise masking threshold. Figure 2 shows the relationship between the performance of the MBSD and scaling factor. A scaling factor of 0.7 gives the highest correlation coefficient per coder. The MBSD which uses a scaling factor of 0.7 has been labeled MBSD II.

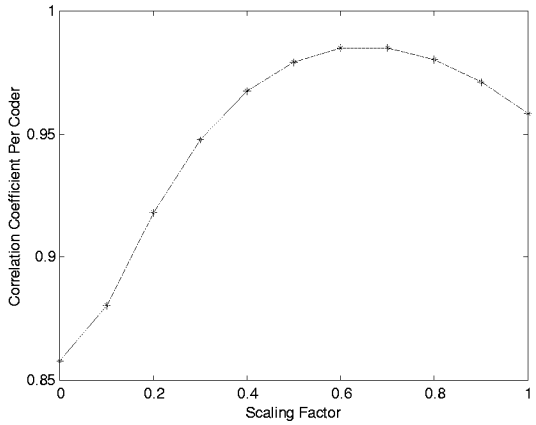

Figure 2. Performance of the MBSD versus the scaling factor

For the experiments, we used a speech data set which included 5 MNRU conditions and various different types of speech coders such as ADPCM, GSM, IS54, FS1016, LD-CELP and CELP. In our experiment, 64Kbps PCM was regarded as original speech. Table 1. shows the correlation coefficients of various measures. There are two different correlation coefficients that can be used. One is the correlation coefficients related to each speech utternace and is identified as Per Speech. The other is the correlation coefficent related to each coder and is identifed as Per Coder. Since objective measures are used for the coder evaluation, the correlation coefficient with each coder is usually used. However, the Per Coder correlation coefficient may increase correlation coefficient by compensating for two oppositely correlated components. Therefore, we report both correlation coefficients for the evaluation of objective quality measures. The performance of the MBSD II Per Coder is as good as P.861 and MNB II, as shown in table 1. The performance of the MBSD II Per Speech is clearly better than P.861 and MNB II. The MNB showed a relatively poor Per Speech performance because the performance of MNB was optimized with Per Coder.

Table 1. Correlation coefficients of MBSD II and other measures

	Per Speech	Per Coder
P.861	0.8933	0.9801
MNB I	0.8319	0.9658
MNB II	0.8478	0.9833
MBSD	0.9001	0.9582
MBSD II	0.9252	0.9851

4. Effect of Noise Masking Threshold

Since the MBSD uses the noise masking threshold which determines if the distortion is perceptible, it is worthwhile to examine the effect of noise masking threshold on the performance of the MBSD. We have compared the performance of the MBSD without noise masking threshold and with noise masking threshold. The estimated distortion for the MBSD without noise masking threshold has been computed by setting $M(i)$, indicator of perceptible distortion to 1. Figure 3 shows the performance of the MBSD without noise masking threshold. According to Figure 3, the MBSD without noise masking threshold overestimates some distortions because it simply calculates the loudness difference without considering perceptual distortion.

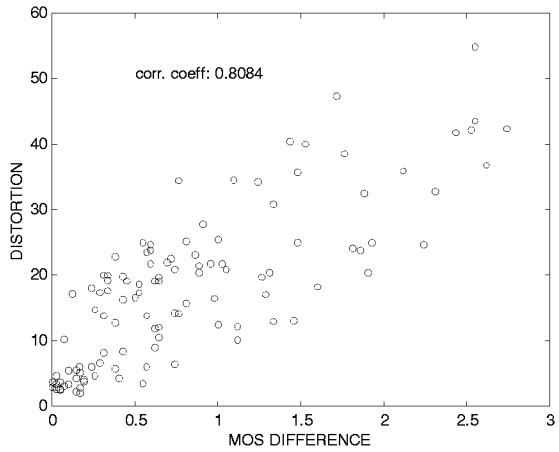

Figure 3. Scattered plot of MBSD without noise and masking threshold versus MOS difference

Figure 4 shows the performance of the MBSD with noise masking threshold over the same speech data set. It shows clearly that the overestimated distortion has been decreased and the MBSD with noise masking threshold gives a higher correlation with subjective quality measure. Therefore, for the MBSD, the noise masking threshold plays an important role in estimating perceptually relevant distortion of objective speech quality measure. Figure 4 shows the performance of the MBSD with noise masking threshold over the same speech data set. It shows clearly that the overestimated distortion has been decreased and the MBSD with noise masking threshold gives a higher correlation with subjective quality measure.

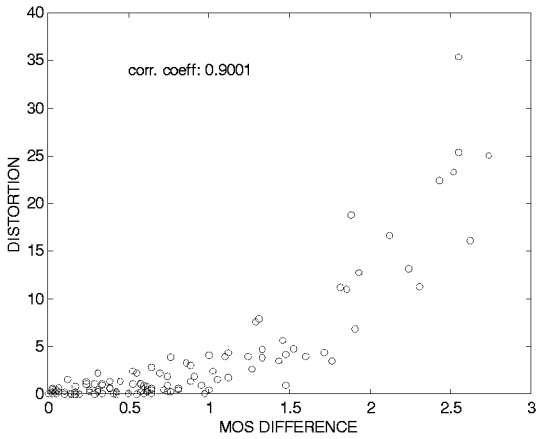

Figure 4. Scattered plot of MBSD with noise masking threshold versus MOS difference

5. Correlation Analysis with MOS Difference

Correlation coefficients with the MOS scores have been the traditional evaluation tool for the performance of objective speech quality measures. It has been suggested that it is more appropriate to use correlation coefficients with the DMOS (Distortion Mean Opinion Score) rather than the MOS for evaluation of the performance of objective speech quality measures [2]. One reason for this claim is based on the observation of the difference between the MOS test and objective speech quality measures. While the subjects in a MOS test determine the speech quality without hearing the original speech, objective speech quality measures estimate the distortion by comparing the distorted speech to the original speech. In the DMOS test, listeners hear both the original and the distorted speech and assign the degree of distortion with DMOS scores of 1 to 5. Consequently, the procedure of the DMOS test is very similar to that of the objective measures.

Since we didn't have the DMOS data, we used the MOS differences in the correlation analysis. The MOS difference between the original speech and the coded speech is used for the evaluation of objective speech quality measures with a second-order regression analysis. Table 2. shows the correlation coefficients with the MOS as well as with the MOS difference. These analyses have been done with each speech file.

Table 2. Correlation coefficients with MOS and MOS difference

	MOS	MOS difference
P.861	0.8731	0.8933
MNB I	0.7958	0.8319
MNB II	0.8140	0.8478
MBSD	0.8782	0.9001
MBSD II	0.9041	0.9252

According to Table 2, all of the measures showed higher correlation with the MOS difference. This result indicates that it would be more appropriate to use DMOS scores in order to evaluate objective measures, and as noted previously, objective

measures directly resemble the DMOS test rather than the MOS test.

6. CONCLUSION

The MBSD is a modified conventional BSD, which incorporates the noise masking threshold. Noise masking threshold plays an important role in estimating perceptual distortion in the MBSD. The MBSD II improves the performance of the MBSD by simply adopting a scaling factor of 0.7 to the noise masking threshold and its performance per coder is as good as ITU-T Recommendation P.861 and MNB II. The performance of MBSD II per speech is better than P.861 and MNB II. However, the performance of MBSD II is not known for other nonlinear distortions such as channel impairments. Also, the performance of the MBSD measures needs to be examined with other speech data bases. Since objective quality measures compares two different speech signals, it would be more appropriate to use DMOS scores instead of MOS scores in order to evaluate objective quality measures. Our preliminary experiments also indicates that objective measures showed higher correlation with the MOS difference than with the MOS.

Acknowledgment:

We wish to thank Peter Kroon of Lucent Technologies for supplying original and coded speech and associated MOS scores.

5. REFERENCES

[1] W. Yang, M. Dixon and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold," IEEE Speech Coding Workshop, pp. 55-56, Pocono Manor, 1997

[2] W. Yang, M. Benbouchta and R. Yantorno, "Performance of the modified bark spectral distortion as an objective speech quality measure," ICASSP, vol. 1, pp. 541-544, Seattle, 1998

[3] ITU-T Rec. P.861, "Objective quality measurement of telephone-band speech codecs," Geneva, 1996

[4] S. Voran, "Estimation of perceived speech quality using measuring normalizing blocks," IEEE Speech Coding Workshop, pp. 83-84, Pocono Manor 1997

[5] D. Sen and W. H. Holmes, "Perceptual enhancement of CELP speech coders," ICASSP, vol. 2, pp. 105-108, 1994

[6] S. Wang, A. Sekey and A. Gersho, "An objective measure for predicting subjective quality of speech coders," IEEE J. on Select. Areas in Comm., vol. SAC-10, pp. 819-829, 1992

[7] J. G. Beerends & J. A. Stemerdink, "A perceptual speech quality measure based on a psychoacoustic sound representation," J. Audio Eng. Soc. vol. 42, pp. 115-123, March, 1994

[8] K. Lam, O. Au, C. Chan, K. Hui, and S. Lau, "Objective speech quality measure for cellular phone," ICASSP, vol. 1, pp. 487-490, 1996

[9] M. M. Meky and T. N. Saadawi, "A perceptually-based objective measure for speech coders using abductive network," ICASSP, vol. 1, pp. 479-482, 1996

[10] S. Voran and C. Sholl, "Perception-based objective estimators of speech quality," IEEE Speech Coding Workshop, pp. 13-14, Annapolis 1995

[11] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. on Select. Areas in Comm., vol. SAC-6, pp. 314-323, 1988

[12] D. Sen, D. H. Irving and W. H. Holmes, "Use of an auditory model to improve speech coders," ICASSP, vol. 2, pp. 411-414, 1993