



Question(s):

STUDY GROUP 12 – CONTRIBUTION <no.>

SOURCE*: SPEECH PROCESSING LAB, TEMPLE UNIVERSITY

TITLE: MODIFIED BARK SPECTRAL DISTORTION (MBSD)

ABSTRACT

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 [1] [2]. In this paper, the MBSD is improved by scaling noise masking threshold and its performance is compared to that of ITU-T Recommendation P.861 [3] and MNB [4] ITU-T Recommendation P.861A.

1. INTRODUCTION

Among the various different objective speech quality measures, we have been interested in the perceptual distortion measures such as Bark Spectral Distortion (BSD) [5] and Perceptual Speech Quality Measure (PSQM) [6]. PSQM has been recommended as an objective quality measurement of telephone-band speech codecs by ITU. Since the development of the BSD, it has become a good candidate for a highly correlated objective quality measure, according to several researchers [7][8][9]. 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 is the average squared Euclidean distance of estimated loudness of the original and the coded utterances.

Even though the conventional BSD measure showed a relatively high correlation with MOS, there are areas for possible improvement. Motivated by the transform coding of audio signals, which uses the noise masking threshold [10], 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 for the calculation of distortion. This new addition of the noise masking threshold replaces the empirically derived distortion threshold value used in the conventional BSD [5]. Since the MBSD compares the distorted speech to the original speech, its

* **Contact:** Wonho Yang or
Robert Yantorno
Speech Processing Lab
Temple University

Tel:1-215-204-6984 Fax:1-215-204-5960
Tel:1-215-204-3381 Fax:1-215-204-5960
e-mail: wonho@astro.ocis.temple.edu
e-mail: ryantorn@nimbus.temple.edu

performance would be sensitive to the temporal misalignment. So, a synchronization algorithm based on loudness domain is applied prior to performing the MBSD [11].

In this paper, we describe the MBSD measure. The effect of noise masking threshold on the estimation of distortion is discussed. The performance of the MBSD is improved by scaling noise masking threshold and compared to that of ITU-T Recommendation P.861 and ITU-T Recommendation P.861A.

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 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. The procedure of loudness transformation is same as that of the BSD [5]. However, there are two differences between the conventional BSD and the MBSD. First, the MBSD uses the noise masking threshold for the determination of audible distortion, while the BSD uses an empirically determined power threshold. Second, the computation of distortion in the BSD is different from that of the MBSD. The BSD defines the distortion as the average squared Euclidean distance of estimated loudness, while the MBSD defines the distortion as the average difference of estimated loudnesses. We suggest that it is more important to use the DMOS scores for the evaluation of objective measures because the current objective measures are comparison measures [11]. This will be discussed further (below).

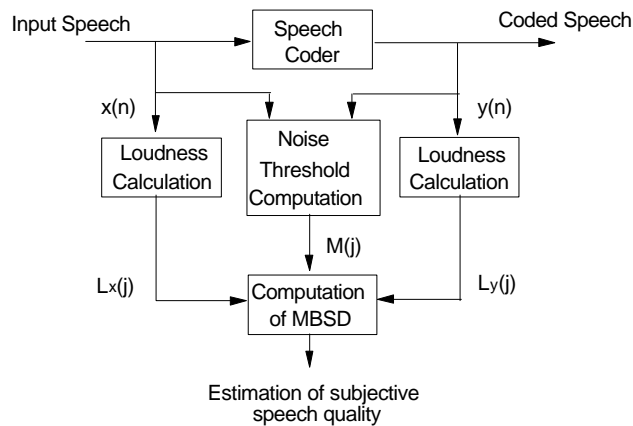


Figure 1. Block diagram of MBSD method

The noise masking threshold is estimated by critical band analysis, spreading function application and absolute threshold consideration [10]. This noise masking threshold estimation considers tone-masking noise and noise-masking tone. 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.

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

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. RESULTS AND DISCUSSION

In order to examine the performance of the MBSD, we performed several different types of experiments. Some results regarding the performance of the MBSD, as compared with P.861 and P.861A, have been previously reported [12] [13]. In this paper, we show that the noise masking threshold plays an important role in estimating perceptual speech quality. The performance of the MBSD is compared to that of ITU-T Recommendation P.861 and P.861A.

For the experiments, we computed the MBSD measures frame by frame, with the frame length of 320 samples overlapping by a half frame. Each frame was weighted by a Hanning window. We processed only non-silence frames. 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. Since an objective quality measure is a comparison measure of two speech utterances and MOS is an absolute measure, 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. In our experiment, 64Kbps PCM was regarded as original speech.

3.1. 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 the noise masking threshold on the performance of the MBSD. In order to examine the effect of the noise masking threshold, we compare the performance of the MBSD with and without the noise masking threshold. The estimated distortion for the MBSD without the noise masking threshold has been computed by setting $M(i)$, indicator of perceptible distortion to 1. Figure 2 shows the performance of the MBSD without the noise masking threshold. According to Figure 2, the MBSD without the noise masking threshold overestimates some distortions because it simply calculates the loudness difference without considering perceptual distortion. Figure 3 shows the performance of the MBSD with the noise masking threshold over the same speech data set. It clearly shows that the overestimated distortion has been decreased and the MBSD with the noise masking threshold gives a higher correlation with subjective quality measure. Therefore, the noise masking threshold plays an important role in estimating perceptually relevant distortion of objective speech quality measure in the MBSD. It should be noted that P.861 and

P.861A do not take into account the noise masking threshold in estimating perceptual distortion [4][6].

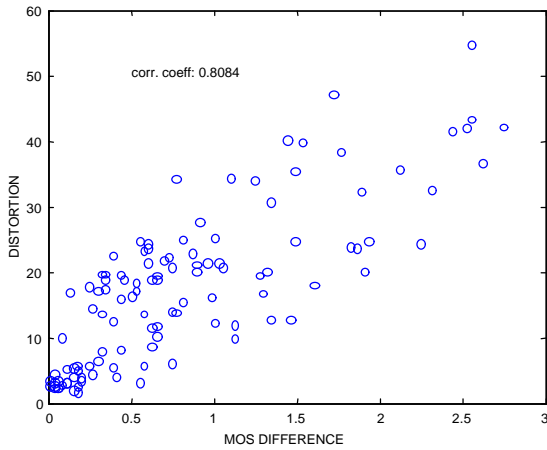


Figure 2. Plot of MBSD distortion without noise masking threshold and MOS difference.

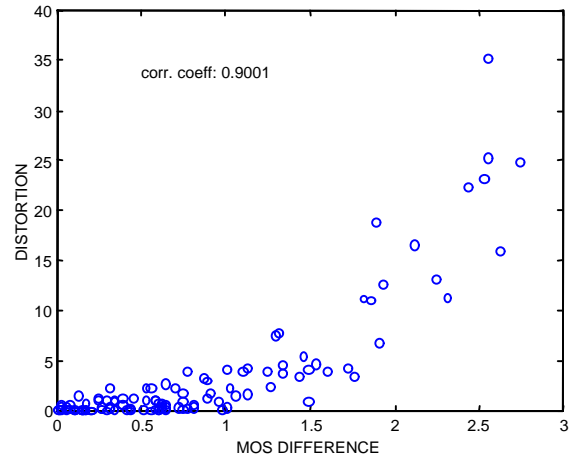


Figure 3. Plot of MBSD distortion with noise masking threshold and MOS difference.

3.2. Comparison of MBSD with P.861 and P.861A

Correlation coefficients with MOS scores have been the traditional evaluation tool for the performance of objective speech quality measures. Yang *et.al.* have suggested that it is more appropriate to use correlation coefficients with DMOS rather than MOS for evaluation of the performance of objective speech quality measures [12]. 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 the reference speech, objective speech quality measures are based on the distortion using a reference. In our speech database, since the associated DMOS scores were not available, we used the MOS difference instead of DMOS scores.

Table 1. shows the correlation coefficients of the MBSD, P.861 and MNB. As can be noted, there are two correlation coefficients. One is the correlation coefficient with each speech file and the other is with each condition or each coder. Since objective measures are used for the coder evaluation, the correlation coefficient with each coder is usually used. However, it should be noted that the correlation coefficients with each coder could increase correlation coefficient by compensating two oppositely correlated components. So, we report both correlation coefficients. The correlation coefficient of the MBSD per speech is slightly better than P.861. However, the performance of the MBSD per coder is not as good as P.861 and MNB II.

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

3.3. Improvement of MBSD by Scaling Noise Masking Threshold

We have shown here that there is an improvement of the performance of the MBSD by using the 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, the noise masking threshold might not be very accurate if it is directly applied to nonstationary signals such as speech. 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 4. shows the relationship between the performance of the MBSD and scaling factor, and a scaling factor of 0.7 gives the highest correlation coefficient per coder. The MBSD with a scaling factor of 0.7 is identified as MBSD II. In Table 1., the correlation coefficients of MBSD II and other measures have been shown. The performance of the MBSD II per coder is as good as P.861 and MNB II and the performance of the MBSD II per speech is clearly better than P.861 and MNB II.

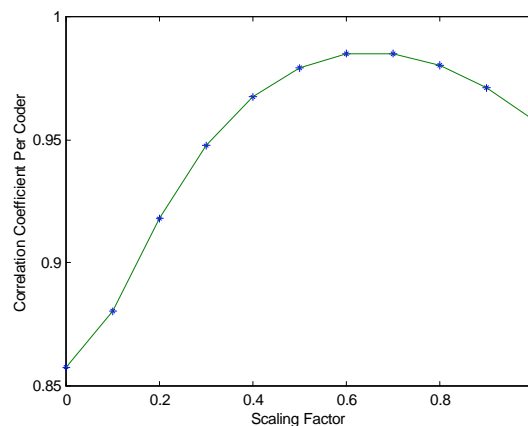


Figure 4. Relationship between the performance of the MBSD and scaling factor

4. CONCLUSION

The MBSD is a modified conventional BSD, which incorporates the noise masking threshold. The noise masking threshold plays an important role in estimating perceptual distortion in the MBSD. The performance of the MBSD is improved by adopting a scaling factor of 0.7. The performance of MBSD II is as good as ITU-T Recommendation P.861 and MNB II for per-coder while MBSD II is better than P.861 and MNB II for per-speech according to the coding distortions available in our speech database. In this paper, we suggest the use of the DMOS scores for the evaluation of objective speech quality measures. Currently, the performance of the MBSD measures is being examined with other speech databases.

ACKNOWLEDGEMENT

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

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] 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
- [6] 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
- [7] 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
- [8] M. M. Mekey and T. N. Saadawi, "A perceptually-based objective measure for speech coders using abductive network," ICASSP, vol. 1, pp. 479-482, 1996
- [9] S. Voran and C. Sholl, "Perception-based objective estimators of speech quality," IEEE Speech Coding Workshop, pp. 13-14, Annapolis 1995
- [10] 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
- [11] M. Benbouchta, "A waveform synchronization algorithm in the context of objective measures of speech quality", Master Thesis, Electrical and Computer Engineering Department, Temple University, Philadelphia, PA, 1998
- [12] W. Yang, and R. Yantorno: "Improvement of MBSD by scaling noise masking threshold and correlation analysis with MOS difference instead of MOS". Submitted to ICASSP 1999
- [13] W. Yang, and R. Yantorno: "Comparison of two objective speech quality measures: MBSD and ITU-T recommendation P.861". Second Annual IEEE Signal Processing Multimedia Conference, 1998.