

A MODIFIED BARK SPECTRAL DISTORTION MEASURE WHICH USES NOISE MASKING THRESHOLD

Wonho Yang, Myron Dixon and Robert Yantorno

Speech Processing Lab

Electrical & Computer Engineering, Temple University, Philadelphia, PA 19122-6077
wonho@astro.ocis.temple.edu, mdixon10@erols.com, ryantorn@nimbus.temple.edu

ABSTRACT

A new Bark Spectral Distortion (BSD) for an objective speech quality measure is being proposed in this paper. This new BSD measure (called MBSD) takes into account the noise masking threshold for the calculation of BSD. Simulation results have shown improvement of the modified BSD over the conventional BSD.

1. INTRODUCTION

Development of an objective speech quality measure that correlates well with subjective speech quality measure has been considered important because subjective tests are expensive and time consuming. Objective measures are also easier to implement and less time consuming. In addition, a good objective speech quality measure can be used to improve speech quality by providing a selection criterion of excitation in CELP type coders [1].

Since BSD has been developed [2], it has become a good candidate for highly correlated objective quality measure according to the several researchers [3][4][5]. 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. 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.

Even though the conventional BSD measure showed a relatively high correlation with MOS score, there are possible areas for improvement. Motivated by the transform coding of audio signals which uses the noise masking threshold [6], we have incorporated the 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 threshold replaces the empirically derived distortion threshold value used in the conventional BSD. The concept of a noise masking threshold was also used to improve speech quality [7]. It was shown that coding gain could

be obtained with no loss of speech quality without transmitting spectral samples below 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 BSD measure because these components are considered inaudible.

2. MBSD MEASURE

Fig. 1 shows the block diagram of MBSD method. The noise masking threshold estimation is added to the conventional BSD. The noise masking threshold is estimated by critical band analysis, spreading function application and absolute threshold consideration [6]. 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.

Indicator of perceptible distortion is denoted by $M(i)$ where, i is the i -th critical band. When the distortion is perceptible, $M(i)$ is 1, otherwise $M(i)$ is 0. The calculation of MBSD is given by equation (1). Imperceptible distortion is excluded in the MBSD calculation by multiplying $M(i)$ because $M(i)$ is zero when the distortion is not perceptible. So, MBSD value can be defined as the average difference of estimated loudness which is only perceptible.

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

where,

N : number of frames processed

K : number of critical bands

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

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

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

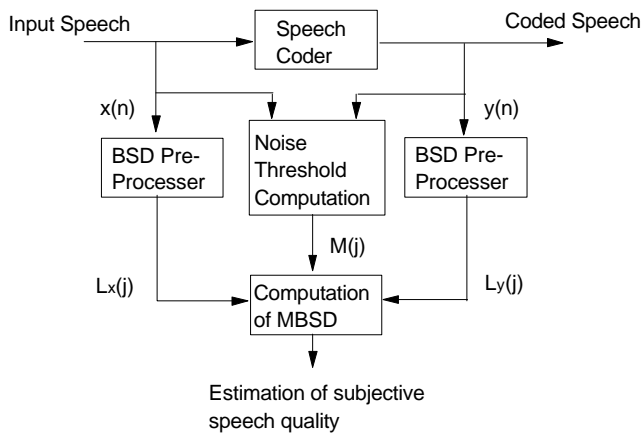


Figure1. Block diagram of MBSD method

3. RESULTS AND CONCLUSION

We computed the BSD measure frame by frame, with the frame length of 160 samples. Each frame was weighted by a Hanning window. We processed voiced frames only because [8] has shown that degradation in quality of LPC speech is not due to coding the unvoiced portion of speech. This suggests measuring the speech quality for unvoiced speech is not necessary. Since the BSD measure is a comparison measure of two speech utterances, we estimated the difference of MOS's of the original speech and the coded speech with a second-order regression analysis. In our experiment, 64Kbps PCM was regarded as the original speech.

Table I shows the correlation coefficients for MNRU of 5 - 25 dB. Note that the MBSD measure showed a slightly better correlation than the conventional BSD. Table II shows the correlation coefficients for 4.8Kbps - 32Kbps coders. Both methods for coders showed lower correlation than for MNRU. When MNRU and coders are combined, the MBSD showed higher correlation. This indicates that MBSD measure is more consistent at predicting MOS score.

Some areas of possible future investigation are to compare both methods for different length of frames because the conventional BSD has the best correlation with the frame length of 80 samples.

Table I. Correlation coefficients for MNRU of 5 - 25dB

	MIXED	FEMALE	MALE
BSD	0.9685	0.9836	0.9539
MBSD	0.9737	0.9841	0.9644

Table II. Correlation coefficients for 4.8Kbps - 32Kbps coders

	MIXED	FEMALE	MALE
BSD	0.7531	0.8551	0.8289
MBSD	0.7304	0.8128	0.7221

Table III. Correlation coefficients for MNRU and coders

	MIXED	FEMALE	MALE
BSD	0.8976	0.9377	0.8537
MBSD	0.9562	0.9686	0.9446

Acknowledgments :

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

4. REFERENCES

- [1] D. Sen and W. H. Holmes, "Perceptual enhancement of CELP speech coders," ICASSP, vol. 2, pp. 105-108, 1994
- [2] S. Wang, A. Sekey and A. Gersho, "An objective measure for predicting subjective quality of speech coders," IEEE J. on Select. Areas in Commun., vol. SAC-10, pp. 819-829, 1992
- [3] 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
- [4] 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
- [5] S. Voran and C. Sholl, "Perception-based objective estimators of speech quality," IEEE Speech Coding Workshop, pp. 13-14, Annapolis 1995
- [6] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. on Select. Areas in Commun., vol. SAC-6, pp.314-323, 1988
- [7] 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
- [8] G. Kubin, B. S. Atal and W. B. Kleijn, "Performance of noise excitation for unvoiced speech," IEEE Speech Coding Workshop, pp. 35-36, 1993