PERCEPTUAL ANALYSIS OF HIGHER-ORDER STATISTICS IN ESTIMATING REVERBERATION

Thiago de M. Prego[†], *Amaro A. de Lima*[†‡] *and Sergio L. Netto*[†]
[†]PEE/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
[‡]Federal Center for Technological Education Celso Suckow da Fonseca (CEFET-RJ), Nova Iguaçu, RJ, Brazil.
{*thprego,amaro,sergioln*}*@lps.ufrj.br*

ABSTRACT

This paper presents a study of the capacity of four speech signal features to assess speech perceptual quality and their use in a typical two-stage algorithm for reverberant speech enhancement. This algorithm is divided into two blocks: one that deals with the coloration effect, due to the early reflections, and the other for reducing the long-term reverberation. The proposed features are skewness, two types of kurtosis and Shannon's entropy. This assessment capacity is evaluated by two perceptual-quality measure specific for the speech-reverberation context. Experimental results for a 204-signal database show that the proposed features can achieve a correlation coefficient of -75% (for entropy) which indicates the potential use for entropy in speech enhancement algorithms.

Index Terms— Kurtosis, Skewness, Entropy, Perceptual quality assessment, Inverse filtering.

1. INTRODUCTION

Speech intelligibility and quality are affected by several kinds of impairments during signal generation, processing or transmission. Such impairments, for example, may include speech coding distortions, packet loss, time clipping, background noise, echo and reverberation. Although most of these impairments is considered by most people to better be absent, the reverberation in a small amount turns the speech more pleasant [1] for normal listeners. However reverberation can drastically affect the performance of current automatic speech/speaker recognition or hearing-aid systems, requiring an appropriate speech enhancement technique to reduce its effects.

The main objective of this work is to present a study of possible signal features that can be used by a typical two-stage speech dereverberation algorithm [2, 3]. The features studied are skewness as given by [4, 5, 6], two definitions of kurtosis [2, 7] and Shannon's entropy as given by [8].

This paper is organized as follows: in Section 2, a typical two-stage speech dereverberation algorithm is described. Section 3 defines the 4 signal features (skewness, 2 definitions of kurtosis and entropy) which are going to be studied and how they are related to reverberation. Section 4 describes the speech database NBP (New Brazilian Portuguese database) used to study the features, a subjective score and an objective score associated to perceptual quality of reverberant speech signals. In Section 5, the experimental results for the capacity of each signal feature to assess speech perceptual quality of the NBP database and the discussion about these results are presented. Finally, a conclusion concerning the use of this approach in the dereverberation scenario is included in Section 6.

2. DEREVERBERATION ALGORITHM

A typical dereverberation algorithm is called two-stage algorithm, composed by two isolated signal processing blocks, inverse filtering and spectral subtraction, as shown in Figure 1. $y(n)$, $z(n)$ and $x(n)$ are the reverberant speech, inverse-filtered speech and dereverberated speech (or spectral-subtracted), respectively.

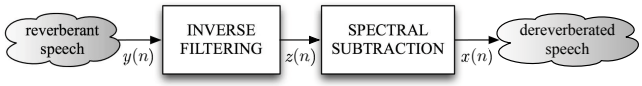

Fig. 1. Diagram of the two-stage dereverberation algorithm.

Its concept comes from the commonly adopted model of reverberant room impulse response (RIR), which is composed by three parts: the direct path signal; the early reflections, which presents a non-flat frequency response that distorts the speech spectrum; and finally the late reverberation, which causes smearing of the speech spectrum, reducing the intelligibility and quality of the signal [3].

The distortions caused by the early reflections are commonly reduced by inverse filtering the reverberant speech, generating an estimate of the original speech signal. The inverse filter optimizes a higher order statistics of the linear prediction (LP) residue of the inverse filtered speech, which may be implemented as a block Least Mean Squares (LMS)-like adaptive algorithm [3], using the higher order statistics as adaptive cost function.

3. HIGHER-ORDER STATISTICS

According to [2, 8] higher order statistics as skewness, kurtosis and entropy can be used to estimate the reverberation amount in a speech signal. Intuitively, the more accurate is this estimation, the more efficient is expected to be the inverse filtering as indicated by [2], which shows that the kur-

tosis can express the reverberation, thus being applied to the inverse filtering.

According to [8], especially in short frames, the samples of the LP residual signal are less correlated than the samples of the speech signal. For this particular reason, the following statistics metrics are applied directly to the LP residue of $y(n)$, denoted by $y_r(n)$. The reverberant signal $z(n)$ is divided into M frames of N samples with $O\%$ overlap, where $N = N_{ms} \times \frac{F_s}{1000}$, N_{ms} is the frame size in milliseconds and F_s is the sampling frequency. The LP residues are calculated for each frame using an LP filter with L coefficients, generating the frame residue $y_r(n;m)$, where n is the sample index within the m^{th} frame.

3.1. Skewness

The skewness is a measure of asymmetry of the probability distribution around the sample mean. In this work it is defined as [4, 5, 6]

$$\mathcal{S}(m) = \frac{\mathbb{E}[{y_r}^3(n;m)]}{\mathbb{E}^{3/2}[{y_r}^2(n;m)]}, \tag{1}$$

where $\mathbb{E}[\cdot]$ denotes the statistical mean operator over the samples n.

3.2. Kurtosis

The kurtosis is a measure of the peakiness of a probability distribution of a real-valued random variable. There are several definitions found in the literature [2, 9, 10, 7, 11, 12, 13], from which two were chosen to be studied: The first is mathematically denoted by [2]

$$\mathcal{K}_1(m) = \frac{\mathbb{E}[{y_r}^4(n;m)]}{\mathbb{E}^2[{y_r}^2(n;m)]} - 3, \tag{2}$$

and the second by [7]

$$\mathcal{K}_2(m) = \mathbb{E}[{y_r}^4(n;m)] - 3\mathbb{E}^2[{y_r}^2(n;m)]. \tag{3}$$

3.3. Entropy

The entropy is a measure of a random variable uncertainty and it is usually referred to Shannon's entropy. According to [8], the entropy $\mathcal{H}(m)$ of $y_r(n;m)$ can be estimated from a B-bin ($B = 7$) histogram of its samples in each frame, generating the estimated probability $p(i;m)$ of the i^{th} bin for m^{th} frame and

$$\mathcal{H}(m) = -\sum_{i=1}^{B} p(i;m)\log(p(i;m)). \tag{4}$$

3.4. Quantifying reverberation

The reverberation amount in a speech signal can be roughly estimated by averaging the higher order statistics measures over the frames [2], providing a single measure for a given speech signal.

4. REVERBERANT SPEECH QUALITY ASSESSMENT

In order to study the capacity to relate to perceptual reverberation effects of a signal feature, a speech database composed of three distinct reverberation approaches, a subjective and an objective scores for reverberant speech quality were chosen and are described next.

4.1. Speech database

The experimental part of this work is entirely based on the so-called New Brazilian-Portuguese (NBP) database [14], which is comprised of 204 speech signals of $F_s = 48$ kHz sampling frequency and different reverberation types and intensity levels. The complete database was generated from (and includes) 4 anechoic speech signals (2 from a male speaker and 2 from a female speaker) contaminated with three distinct reverberation approaches: artificial, natural and real.

The artificial and natural approaches convolves 6 artificially generated and 17 directly recorded RIRs, respectively, with the 4 anechoic speech signals. In the real approach, the reverberant speech signals were obtained from direct recording the 4 anechoic signals played in 7 distinct rooms with a total of 27 different configurations. The whole database is composed of 51 different scenarios with reverberation times in the range $T_{60} = [120, 920]$ ms.

4.2. Quality assessment scores

The 204 speech signals perceptual quality were assessed by an absolute category rate mean opinion score (MOS) test with 30 listeners. Due to the cost of subjective tests, it is common to use an objective score highly correlated to the MOS score to asses the quality of a speech signal.

The objective score used in this work is estimated by the measure Q_{MOS} for quality assessment of reverberation proposed in [14]. The measure Q_{MOS} is derived from the measure Q, which results from the combination of three features estimated from the room impulse response (RIR). These features are the reverberation time (T_{60}), the room spectral variance (σ_I^2) and the direct-to-reverberant energy ratio (R). The measure Q is expressed as

$$Q = -\frac{T_{60}\,\sigma_I^2}{R^{\gamma}}, \tag{5}$$

where the exponent $\gamma = 0$ corresponds to Allen's original measure and the best system performance was empirically obtained using $\gamma = 0.3$.

The Q score is mapped to MOS by a third order mapping function

$$Q_{\mathrm{MOS}} = \alpha(x_1 Q^3 + x_2 Q^2 + x_3 Q + x_4) + \beta, \tag{6}$$

where the coefficients $x_1, x_2, x_3, x_4, \alpha$ and β are empirically determined during the system training. In practice, the coefficients obtained for the NBP database were $x_1 = 0.0017$, $x_2 = 0.0598$, $x_3 = 0.7014$, $x_4 = 4.5387$, $\alpha = 1.0000$ and $\beta = 1.85 \times 10^{-10}$.

5. EXPERIMENTAL RESULTS

5.1. Comparative Analysis

The performance of the higher order statistics measures were calculated by the Pearson's correlation (ρ) between the measures and MOS scores. The perceptual quality of the reverberant speech signals are represented by the subjective and objective MOS scores, and the correlation analyzes the dependency between these perceptual measures and the higher order statistics measures used in this work.

Table 1 presents the correlations between the subjective MOS and the higher order statistical measures $\mathcal{S}$, $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{H}$, represented by $\rho_{\mathcal{S}}$, $\rho_{\mathcal{K}_1}$, $\rho_{\mathcal{K}_2}$ and $\rho_{\mathcal{H}}$, respectively. These correlations were calculated to several different values of LP filter order, frame size and overlap percentage, where the ranges were $L = [10, 30, 50]$, $N_{ms} = [10, 32]$ and $O = [0\%, 50\%]$, thus a setup can be described by $(L, N_{ms}, O\%)$. The best correlation performance was 40% for $\mathcal{S}$ with the setups $(50, 10, 0\%)$ and $(50, 10, 50\%)$; -28% for $\mathcal{K}_1$ with $(10, 10, 0\%)$ and $(30, 10, 0\%)$; 50% for $\mathcal{K}_2$ with $(30, 10, 0\%)$; and -72% for $\mathcal{H}$ with $(30, 32, 0\%)$ and $(50, 32, 0\%)$. The performance of $\mathcal{H}$ is clearly superior than the other higher order statistical measures, indicating that the entropy can be more appropriated than the other measures as the cost function of the adaptive inverse filtering process.

Table 1. *Correlation coefficients $\rho_{\mathcal{S}}$, $\rho_{\mathcal{K}_1}$, $\rho_{\mathcal{K}_2}$ and $\rho_{\mathcal{H}}$ between MOS and $\mathcal{S}$, $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{H}$, respectively.*

Setup			Correlation coefficients (%)			
L	N_{ms}	$O\%$	$\rho_{\mathcal{S}}$	$\rho_{\mathcal{K}_1}$	$\rho_{\mathcal{K}_2}$	$\rho_{\mathcal{H}}$
10	10	0	37	**-28**	48	-63
10	10	50	37	-27	49	-62
10	32	0	21	-19	32	-69
10	32	50	22	-18	31	-67
30	10	0	35	**-28**	**50**	-67
30	10	50	36	-27	49	-66
30	32	0	20	-17	35	**-72**
30	32	50	21	-16	35	-71
50	10	0	**40**	-26	49	-68
50	10	50	**40**	-26	49	-68
50	32	0	25	-15	35	**-72**
50	32	50	27	-14	35	-71

Table 2 presents the same layout and characteristics of Table 1, except for use of objective MOS (Q_{MOS}), instead of subjective MOS as perceptual quality measure to calculate the correlation performance of the higher order statistical measures $\mathcal{S}, \mathcal{K}_1, \mathcal{K}_2$ and $\mathcal{H}$. The best correlation performance was 45% for $\mathcal{S}$ with the setup $(50, 10, 0\%)$; -32% for $\mathcal{K}_1$ with $(10, 10, 0\%)$ and $(30, 10, 0\%)$; 51% for $\mathcal{K}_2$ with $(30, 10, 0\%)$ and $(50, 10, 0\%)$; and -75% for $\mathcal{H}$ with $(30, 32, 0\%)$ and $(50, 32, 0\%)$. The results of this table confirm the performances found in Table 1.

For presentation purpose, MOS, Q_{MOS} and $\mathcal{H}$ scores average and standard deviation for the 4 instances (one for each anechoic signal) of each of the 51 scenarios were calculated. Figures 2 and 3 show the relation between MOS and $\mathcal{H}$ scores for the 51 reverberation scenarios of NBP using with standard deviation of $\mathcal{H}$ as error bar. Both figures clearly exemplifies the negative correlation, because as the subjective MOS values increase, the entropy values decrease, not affecting their high correlation.

Table 2. *Correlation coefficients $\rho_{\mathcal{S}}$, $\rho_{\mathcal{K}_1}$, $\rho_{\mathcal{K}_2}$ and $\rho_{\mathcal{H}}$ between Q_{MOS} and $\mathcal{S}$, $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{H}$, respectively.*

Setup			Correlation coefficients (%)			
L	N_{ms}	$O\%$	$\rho_{\mathcal{S}}$	$\rho_{\mathcal{K}_1}$	$\rho_{\mathcal{K}_2}$	$\rho_{\mathcal{H}}$
10	10	0	41	**-32**	49	-64
10	10	50	41	-31	49	-63
10	32	0	24	-22	29	-72
10	32	50	26	-21	30	-70
30	10	0	41	**-32**	**51**	-68
30	10	50	40	-31	50	-67
30	32	0	24	-20	34	**-75**
30	32	50	25	-20	34	-73
50	10	0	**45**	-30	**51**	-69
50	10	50	44	-29	50	-68
50	32	0	28	-18	34	**-75**
50	32	50	31	-18	35	-73

This reduced data approach was adopted to better show the entropy performance with error bar, once the 204-signals figure made the plot confusing and difficult to analyze.

The $\mathcal{H}$ score was the only higher order statistical measure used in this work that could associate the appropriated score (position in the plot) for the anechoic speech scenario, which is the point in the most right position, around 4.5 MOS in Figure 2. In the reduced data approach, the correlation coefficient between MOS and $\mathcal{H}$ and between Q_{MOS} and $\mathcal{H}$ are -86% and -89%, respectively.

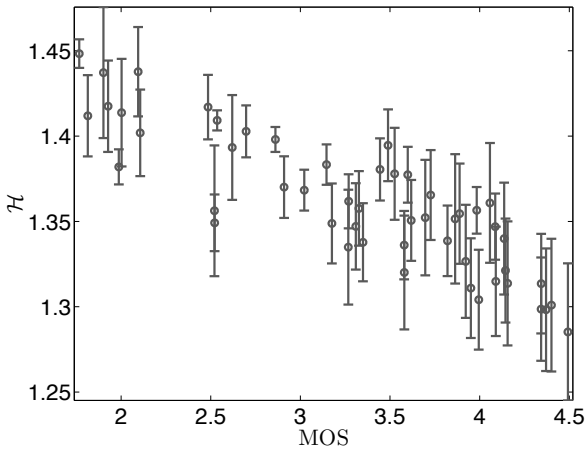

Fig. 2. Relation between the subjective MOS and $\mathcal{H}$ scores for the 51 reverberation scenarios of NBP with standard deviation as error bar.

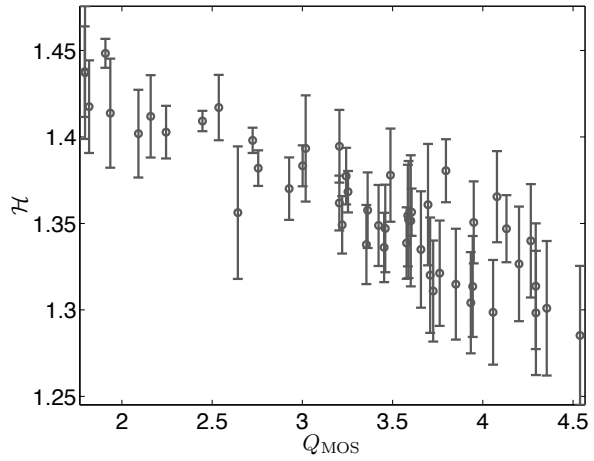

Fig. 3. Relation between the objective MOS (Q_{MOS}) and $\mathcal{H}$ scores for the 51 reverberation scenarios of NBP with standard deviation as error bar.

5.2. Discussion

Observing the correlation performances presented in Tables 1 and 2, it can be concluded that the overlap percentages $O\%$ make no difference in efficiency of any of the four higher statistical measures. The LP filter order L also seems not to affect significantly the measures performances, reaching a maximum of 6% difference in correlation for all the measures, when comparing setups with the same frame size N_{ms}. The only parameter that seems to be relevant for the efficiency of the measures in the limited experimental scenario adopted in this work is the frame size N_{ms}, due to the fact that different values of N_{ms} generate significantly different correlation performances for all measures. In addition, the best performances for $\mathcal{S}$, $\mathcal{K}_1$ and $\mathcal{K}_2$ were reached using $N_{ms} = 10$; and for $\mathcal{H}$ were reached using $N_{ms} = 32$.

The use of $\mathcal{S}$ and $\mathcal{K}_1$ in the inverse filtering part of a two-stage dereverberation algorithm has shown to be adequate [2, 3, 6, 15]. However the use of other higher order statistical measures that better estimate the perceptual quality of a reverberant speech such as $\mathcal{K}_2$ and $\mathcal{H}$ indicated in this work, suggests that these measures are appropriated to be applied in the adaptive inverse filtering for speech dereverberation.

6. CONCLUSIONS

This work analyzes the capability to assess the perceptual quality of a reverberant speech signal of four higher order statistical measures: skewness ($\mathcal{S}$), two versions of kurtosis ($\mathcal{K}_1$ and $\mathcal{K}_2$), and entropy ($\mathcal{H}$). The analysis was performed by verifying the correlation coefficients between these four measures and two different perceptual reverberation quality scores, the subjective MOS and an objective MOS (Q_{MOS}), applied to a 204-signals reverberant speech database, called NBP database, which consists of 51 reverberation scenarios combined with 4 anechoic speech signals. The best performances were -72% and -75% reached by $\mathcal{H}$, whose performances were more than 20% higher than the performances of $\mathcal{K}_2$, which was the measure that reached the second best scores. Future research developments may include the use of entropy as cost function for improving the adaptive inverse filtering block for speech dereverberation algorithm.

7. REFERENCES

[1] R. Appel and J. Beerends, "On the Quality of Hearing One's Own Voice," *J. Audio Engineering Society*, vol. 50, no. 4, pp. 237–248, April 2002.

[2] B. W. Gillespie, H. S. Malvar and D. A. F. Florêncio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* Salt Lake, USA, May 2001.

[3] M. Wu and D. Wang, "A Two-Stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 14, May 2006.

[4] P. Pääjärvi and J. P. LeBlanc, "Skewness Maximization for Impulsive Sources inBlind Deconvolution," *Proc. of the 6th Nordic Signal Processing Symposium*, Espoo, Finland, June 2004.

[5] Q. Shi, R. Wu and S. Wang, "A Novel Approach to Blind Source Extraction Based on Skewness" *IEEE International Conference on Signal Processing*, Beijing, China, Nov. 2006.

[6] S. Mosayyebpour, A. Sayyadiyan, M. Zareian, and A. Shahbazi, "Single Channel Inverse Filtering of Room Impulse Response by Maximizing Skewness of LP Residual," *IEEE Int. Conf. on Signal Acquisition and Processing*, Bangalore, India, Feb. 2010.

[7] O. Tanrikulu and A.G. Constantinides, "The LMK Algorithm with Time-varying Forgetting Factor for Adaptive System Identification in Additive Output-noise," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Atlanta, USA, May 1996.

[8] B. Yegnanarayana and P. S. Murthy, "Enhancement of Reverberant Speech Using LP Residual Signal," *IEEE Trans. on Speech and Audio Proc.*, vol. 8, May 2000.

[9] J. P. LeBlanc and P. L. De Leòn, "Speech Separation by Kurtosis Maximization," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Seattle, USA, May 1998.

[10] Z. Xiong, Z. Dongfeng, J. Zhigang and W. Anhong, "A Modified Blind Equalization Algorithm Based on Kurtosis of Output Signal" *Proc. IEEE Asia-Pacific Radio Science Conference*, Qingdao, China, Aug. 2004.

[11] O. Tanrikulu and A.G. Constantinides, "Least-mean Kurtosis: A Novel Higher-order Statistics Based Adaptive Filtering Algorithm," *IET Electronic Letters*, vol. 30, Feb. 1994.

[12] D. I. Pazaitis and A. G. Constantinides, "A Novel Kurtosis Driven Variable Step-Size Adaptive Algorithm," *IEEE Trans. on Signal Processing*, vol. 47, Mar. 1999.

[13] B. Sällberg, N. Grbić, and I. Claesson, "Online Maximization of Subband Kurtosis for Blind Adaptive Beamforming in Realtime Speech Extraction," *Proc. IEEE Int. Conf. on Digital Signal Processing*, Cardiff, Wales, July 2007.

[14] A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker and M. Fozunbal, "On the quality-assessment of reverberated speech," *Speech Communication*, (Available online 20 October 2011, http://www.sciencedirect.com/science/article/pii/S0167639311001415), 2011.

[15] T. de M. Prego, A. A. de Lima, and S. L. Netto, "Perceptual improvement of a two-stage algorithm for speech dereverberation," *Proc. Interspeech*, Florence, Italy, pp. 209–212, Aug. 2011.