Problems with Nonparametric Entropy Estimation of Voice Signals

Paulo Rogério Scalassara, prscala@sel.eesc.usp.br
Carlos Dias Maciel, maciel@sel.eesc.usp.br
José Carlos Pereira, pereira@sel.eesc.usp.br
University of São Paulo, São Carlos, Brazil

Suely Oliveira, oliveira@cs.uiowa.edu
David Stewart, dstewart@math.uiowa.edu
The University of Iowa, Iowa City, USA

Abstract. *The use of information theoretic measurements has been applied to many different science fields nowadays. Among the tools used to evaluate the amount of information of a system, the most frequent is the entropy and its derivations such as the relative entropy, also known as the Kullback Leibler divergence, and the Renyi entropy family. The entropy measurements are related to the system's level of complexity, or stated differently, the intricacy of its generated signals. This characteristic may be very important when trying to obtain information from the system, mainly because it is related to the system's behavior. As an example, the condition of human systems, such as the cardiovascular, can be assessed by use of some entropy measures because the change of the electrocardiogram's complexity is related to disturbances, usually pathologies. Despite this great value for system analysis, the entropy estimation is not a simple task. This is because it relies on the probability density function (pdf) of the signal under study. There are basically three approaches for estimating the pdf of a signal: parametric, semiparametric and nonparametric. The parametric technique depends on an assumption about the distribution of the samples and then the parameters of this distribution is estimated using some method. It works really well when the data actually fits the distribution, otherwise the results are poor. The nonparametric technique is not based on any assumptions about the distribution, and the commonest use is simply estimating the pdf through the histogram of the samples. If there is not enough samples the estimate is also poor, then another method based on kernels could be used. In this paper we present an analysis about the estimation of the pdf based on the histogram of the data samples, we make no considerations about the distribution of the samples. Some tests are performed on data from known distributions and also on real voice signals. These signals are recorded samples of sustained vowels /a/ obtained from people with healthy throats. The study of the vocal system is relevant using information theoretic measurements due to its complexity that is related to the generation of sound by the interaction of several other systems. Besides that, many researches have shown that the voice production is also a nonlinear system with chaotic components. The analysis of this system using entropy measures is very promising but it must rely on good estimates of the signal's pdf. One other kind of measurement, called predictive power that has been presented on a previous paper, is based on the KL divergence between the signal and its error of prediction given by some model, like autoregressive or wavelet decomposition. However the KL entropy is more sensitive to pdf estimation errors, then in order to improve the method, we used some techniques to try to compensate this sensibility, especially around less probable samples, obtaining satisfactory results. In brief, this paper aims to show some problems with the nonparametric entropy estimation using the empirical histogram, also some methods to try to reduce the unwanted effects of low probable elements in the estimation of the KL divergence employed on the predictive power calculation of voice signals.*

Keywords: *Entropy, nonparametric estimation, complexity, voice.*

1. INTRODUCTION

Since the introduction of the entropy in telecommunications in 1948 (Shannon, 1948), as a measure of information, it has become one of the most important tools of information theory. However, entropy has a limited applicability in practical signal processing cases due to difficulties of estimation from real data (Bercher and Vignat, 2000). An interesting paper that shows the use of entropy measures is Pincus (1991), as a modification of the classical entropy form called approximate entropy. It shows that the entropy of a signal may be related to the complexity of the generating system. Pincus and Goldberger (1994) also present the relationship of the entropy and the signals' regularity with an application to physiological time-series analysis, especially heart rate data.

Another example, related to the voice production system, is presented in Moore et al. (2006), this paper shows the approximate entropy applied to the analyses of spectral patterns of voice signals. This entropy is considered as a reliable measure of voicing quality in vocal fold cancer patients. In (Richman and Moorman, 2000), another kind of entropy measure is presented, the sample entropy, that is aimed for short and noisy data sets, such as for the cardiovascular system.

Therefore, as presented, a measure of system's complexity is really interesting to several signal processing applications. Bercher and Vignat (2000) mention several kinds of applications such as, source separation, blind deconvolution, source coding, image alignment, and detection of abrupt changes .

An important contribution to the complexity studies came from Kolmogorov (Kolmogorov, 1965), that defined the algorithmic complexity of a given object, such as a sequence of numbers, as the length of the shortest computer program that describes the object (Cover and Thomas, 1991). Actually, these ideas were presented independently and almost simultaneously by Solomonoff and Chaitin, and the latter also presented some papers showing mathematical proofs of the Kolmogorov complexity (Cover and Thomas, 1991). The problem is that it is not possible to compute this measure directly, only obtain a boundary of its value (Wyner and Foster, 2003).

Entropy can be defined for continuous or discrete variables, but for signal processing, we are more interested in the discrete case. Considering an alphabet events as $\mathcal{X}$, the entropy, as defined by Shannon, of a random variable X with $p(x)$ as probability density function (pdf) is given by Equation (1). The logarithm is to the base 2, therefore the entropy is expressed in bits. Based on analysis of the continuity of the equation, $0\,log_2\,0 = 0$, then the entropy is zero when $p(X) = 0$ (Cover and Thomas, 1991).

$$H(X) = -\sum_{X\in\mathcal{X}} p(X)\,log_2\,p(X) \tag{1}$$

A generalization of the Shannon entropy is the Rényi entropy, also called the α-entropy (Rényi, 1960), it is defined by Equation 2. When $\alpha \to 1$, the Rényi entropy tends to the Shannon entropy, that is why it is a generalization of the latter. More information about this kind of entropy can be found in Zyczkowski (2003).

$$H_\alpha(X) = -\frac{1}{1-\alpha}\sum_{X\in\mathcal{X}} p(X)^\alpha \tag{2}$$

Entropy measures are intimately related to the predictability of signals, they can be used to evaluate forecast skill of a system (Scalassara et al., 2009a). These concepts are well formalized in Schneider and Griffies (1999), by a parameter called predictive power which is based on the difference between the Shannon entropy of the state of a system and of its error of predictability. In DelSole (2004), there is a study about the use of different kinds of predictability measures, including the relative entropy and the mutual information. Some qualities of the use of the relative entropy is presented in Kleeman (2002), showing the use of this predictability measure in climate analysis.

The relative entropy, also known as the Kullback-Leibler (KL) entropy, may be interpreted as a measure of the difficulty to discriminate two distributions (DelSole, 2004). The KL entropy of two distributions, $p(X)$ and $q(X)$, is presented in Equation (3).

$$D(p\|q) = \sum_{X\in\mathcal{X}} p(X)\log_2\frac{p(X)}{q(X)} \tag{3}$$

The relative entropy presents some characteristics of a distance metric between distributions, such as being always positive or equal to zero $p(X) = q(X)$. Despite this fact, it is not a real distance because $D(p\|q)$ is different from $D(q\|p)$, it does not satisfy the triangle inequality (Cover and Thomas, 1991).

The aim of the paper is to present a discussion about some nonparametric entropy estimation methods, mainly focused on the histogram estimation. Three methods are approached: first, a method based on the discretization of the continuous pdf; second, one based on a fixed quantization step of the histogram; and, third, a method based on the equalization of the histogram in order to correct unwanted effects of low probable elements in the estimation of the KL divergence.

The paper is divided as follow: in the next section, Entropy Estimation, we present some theory about the several methods of estimation; in Histogram Methods, we discuss more about the methods used in the paper. In Voice Signal Analysis, we briefly talk about the voice signals used in the tests and their analysis. Finally, in Results, we present tests with noise signals, simulated and real voice signals; followed by the Conclusions.

2. ENTROPY ESTIMATION

As stated in (Bercher and Vignat, 2000), a signal usually has continuous pdf because of the noise inherent in the systems, at least in common signal processing applications. The problem is that this pdf is hardly ever known, then it must be estimated from the data available. This is especially problematic for the block entropy estimation, that is the entropy of consecutive blocks, as shown in Schürmann and Grassberger (1996). The block entropy is mainly obtained by means of the standard likelihood estimate, a estimation of all word probabilities up to some fixed length. The block entropy is used to obtain the entropy rate and its characteristics (Scalassara et al., 2009c, 2008b; Crutchfield and Feldman, 2003).

The entropy estimation is normally performed using one of two methods: parametric and nonparametric, although some may be characterized as semiparametric (Erdogmus and Principe, 2006). The parametric approach considers an assumption about the distribution of the data and then makes an estimation of its parameters, however the assumption is not very flexible (Viola et al., 1996). If the estimated pdf matches the data, good results are obtained, otherwise the method performs poorly. The assumption largely used is that of a Gaussian pdf. As explained in Viola et al. (1996), this is so because of three factors: first, finding the best Gaussian for the data is simple; second, the entropy of the Gaussian is straightforward; third, an affine transformation does not change the distribution. The parameters are usually estimated using the maximum likelihood method.

Another approach is the nonparametric estimation, Bercher and Vignat (2000) give some examples such as based on histograms (Moddemeijer, 1989), order statistics (Wachowiak et al., 2005), and kernel methods (Erdogmus et al., 2004). The methods based on histograms can be of fixed-bin histograms; variable bin size, which is in accordance with local data distribution; or rectangular sliding window. If instead of the rectangular window, there is another smoother function, the method is called kernel density estimation (Erdogmus and Principe, 2006). A similar method is the Parzen windowing entropy estimation (Duin, 1976; Erdogmus et al., 2004).

The Parzen windowing is a method of approximating the pdf of a signal. According to Erdogmus and Principe (2000), the pdf is estimated by a sum of even and symmetric kernels that are placed over each sample of the signal. Again, the Gaussian distribution is commonly used, basically due to its characteristic of continuous differentiation. With the estimated pdf, the entropy may be obtained by the direct plug-in of the pdf, as presented in (Erdogmus et al., 2004; Beirlant et al., 1997).

Viola et al. (1996) presents three advantages of using the Parzen estimate instead of the parametric approach. First, the former can model any distribution as long as its has a smooth pdf; second, there is no need to search for parameters, because it used the signal directly; third, it is easy to calculate the derivative of the entropy.

In practical terms, the pdf of a system is a finite function that has length N, Equation (1) can be rewritten as Equation (4), given that $p(x_i)$, or simply p_i, is the probability that the random variable X takes values x_i.

$$H(X) = -\sum_{i=1}^{N} p_i \, log_2 \, p_i \tag{4}$$

In the next section, we give some more details about the histogram technique of entropy estimation, that is a nonparametric method, with both fixed and variable bin size.

3. HISTOGRAM METHODS

The histogram based entropy methods can be basically of two kinds depending on the type of bins: fixed or variable bin length. A simple method is presented in Moddemeijer (1989), which shows a discretization of the entropy estimation of continuous distributions with bias correction. The procedure is performed by dividing the pdf function in a rectangular grid with I equally spaced Δx-sized cells, then the samples' occurrences in each cell, k_i, are summed. The probability of sample x_i is $p(x_i)$, but the approximation is $p_i = p(x_i)\Delta x$. Then, with independent samples, the probability p_i may be replaced by the estimative k_i/N where N is the total number of samples. Substituting this approximation to Equation (4), the result is shown in Equation (5) with the correction $log_2 \Delta x$ due to finite Δx.

$$\hat{H}(X) = -\sum_{i=1}^{I} \frac{k_i}{N} \, log_2 \, \frac{k_i}{N} + log_2 \Delta x \tag{5}$$

The bias correction presented by Moddemeijer (1989) is shown in Equation (6).

$$E\{\hat{H}(X)\} = H(X) - \frac{I-1}{2N} \tag{6}$$

A similar derivation for the relative entropy is presented in Scalassara et al. (2009a) where it is applied to voice pathology discrimination. A related paper that uses the Shannon entropy formulation of Equation (6) for the same purpose is Scalassara et al. (2008a). In Paninski (2003), the discretization and bias correction are also shown, where it is called the maximum likelihood (or plug-in) estimator. However, depending on the kind of signal under study, this might not be the best approach, a better one would be one that did not rely on discretization, but considered directly the discrete case.

Another histogram algorithm is based on a variable number of bins depending on the chosen quantization step as shown in Scalassara et al. (2008b). The digital signals have an amplitude quantization that are usually 16 bits and normalized amplitude that range from -1 to 1, what means that the ideal quantization step (amplitude range divided by the number of combinations) is $q = 2/2^{16} = 3.052 \times 10^{-5}$ (Scalassara et al., 2008b). But with such quantization step is that the histogram would be almost empty if the signal has small number of samples. Then, in order to obtain a good precision, the smallest possible quantization step should be used, there is a trade off between the histogram precision and the available number of samples. This method is especially interesting when there are several signal with different amplitude ranges to analyze, because the number of bins are different for each one but the quantization step is the same. Other techniques for selecting the appropriate bin size are presented in Shimazaki and Shinomoto (2007).

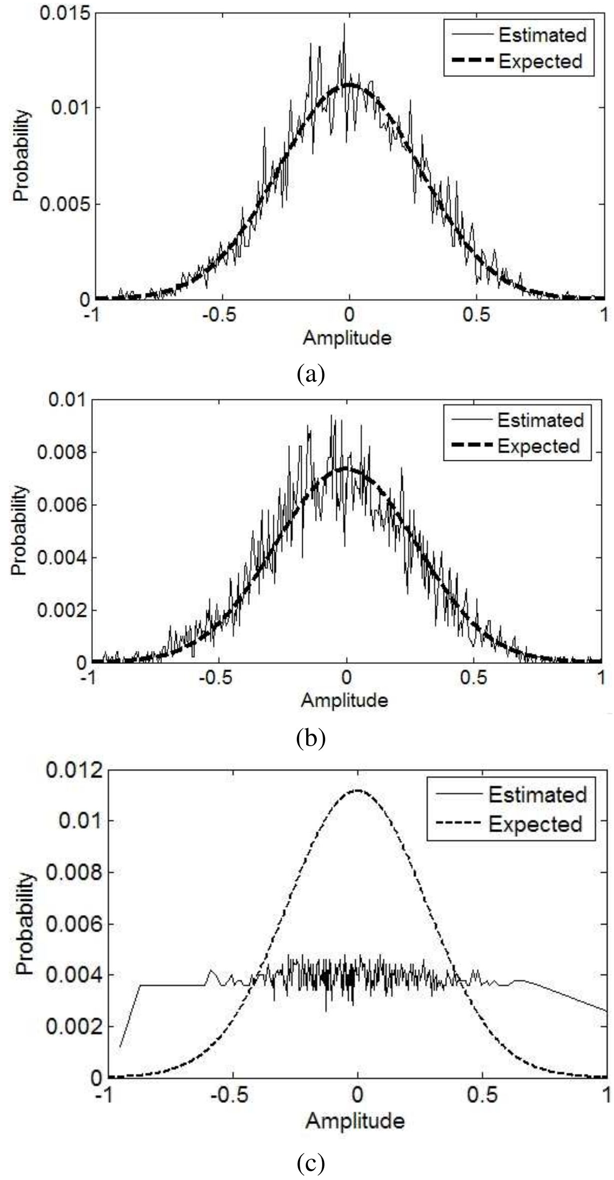

Figure 1. Estimated probability density function of Gaussian noise. (a) Using the method of discretized continuous distribution histogram according to Moddemeijer (1989). (b) Using the method of fixed quantization step histogram. (c) Using the method of equalized histogram.

Although this method has a variable bin size, it is constant for each signal under study. One other method has variations of the bin size according to the distribution of the signal. This technique is based on the equalizations of the histograms in image processing for contrast and color correction of digital pictures (Lim, 1990). The basis is that the algorithm tries to obtain a flat histogram by increasing the length of the bins as the number of samples in each starts to decrease. Some simple codes were created to perform this procedure, but other more sophisticated techniques can be employed such as the one shown in Engel (1997), where the wavelet transform is used to create an adaptive histogram that has variable bin size.

An example is now presented to illustrate the difference of these algorithms. The pdf of an white noise signal is estimated using the three methods presented: histogram based on the discretization of a continuous distribution, based on a fixed quantization step, and with equalized distribution (flat histogram). Figures 1 (a), (b), and (c) present the pdfs of these three methods respectively. In all the figures, the dashed line represents the expected pdf of a Gaussian random variable with the same mean and variation of the noise signal.

In this example, the first two methods had similar results, the real difference appears when several signals with different amplitude ranges have to be analyzed. The last pdf is complete different from the expected one, but its applicability depends on the intended use of the entropy, especially for the estimation of KL entropy.

4. VOICE SIGNAL ANALYSIS

Since the 1960s, several papers were published focused on the acoustical analysis of digital voice signals, like Atal and Schroeder (1967), Markel (1972), and Lieberman (1963) for instance. Then, several studies showed the relationship between the acoustical properties of voice signals and the characteristics of larynx pathologies (Davis, 1979). Also, nonlinear techniques are being employed to study voice signals, one example is Zhang et al. (2005).

Another approach is the use of Information Theory measurements to voice signals. An example is Moore et al. (2006), in which voice electro-glottogram signals from healthy and larynx cancer patients are analyzed with the approximate entropy in order to obtain a complexity measure. In Scalassara et al. (2009a), the relative entropy between healthy and pathological voice signals are used to study their properties, the pathological signals presented the highest entropy values. In Scalassara et al. (2008a) a similar work is presented using the Shannon entropy of healthy and nodule affected signals.

This study uses some healthy voice samples to illustrate the presented methods of entropy estimation. They are composed of sustained vowel /a/ with approximately half a seconds of duration. They were collected along the past ten years and have been used in several studies such as Rosa et al. (2000); Scalassara et al. (2007, 2008b). The signals are taken from the voice database (not public) of the Signal Processing Laboratory of the School of Engineering of São Carlos at the University of São Paulo, Brazil. All signals have normalized amplitude and recorded in mono-channel WAVE format using amplitude quantization of 16 bits and sampling frequency of 22,050 Hz. Figure 2 presents a 50 ms example of a normalized healthy voice signal for illustration.

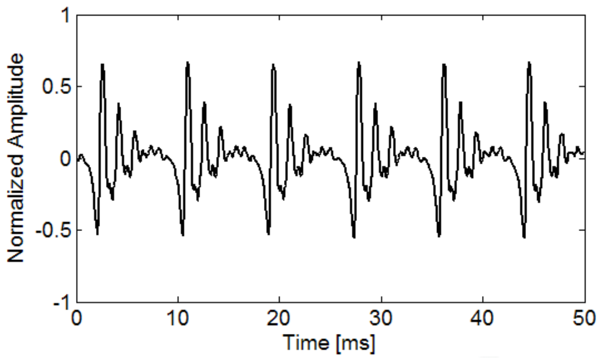

Figure 2. Example of a healthy voice signal.

5. RESULTS

In this section, we present some tests performed with the histogram methods shown before in order to obtain the Shannon and the KL entropy. The KL entropy is used as part of these tests because of its applications in several analysis as shown previously. This is especially true for voice signal analysis as presented in Scalassara et al. (2009a) and Scalassara et al. (2009b). These tests are done with three kinds of signals: Gaussian noise, simulated voice signals, and healthy voice signals. All the values shown in the next figures are based on a mean over 10 different trials with new random noise.

The simulated signal is created using a simple voice simulator that uses a single period of a female healthy voice signal to create a whole signal. Also, it adds Gaussian noise, jitter, and shimmer according to given parameters. An example of a simulated signal is presented in Figure 3.

Firstly, we show a simple test with Gaussian noise that has five different amplitude levels, therefore five different variance levels. The entropy of a Gaussian noise is directly proportional to its variance. Figure 4 shows the normalized entropies of the noise signals for the three histogram methods.

When the discretization method, the one based on Moddemeijer (1989), is used to estimate the signals' histograms, the entropy behaves as expected, it increases with the variance. However, the variance sensitivity is too high, what does not happen with the quantization step method. The histogram equalization method is interesting only for comparisons of histograms, such as the KL entropy, therefore, for a direct measurement such as the Shannon entropy, its results are poor. The entropy with this method shows no sensitivity to the variance. This test shows that the two histogram methods, discretization and quantization, are corresponding to the expected.

As a more interesting test, a simulated voice signal is used instead of the noise signal. This signal has the advantage of being simple and also modeling some characteristics of voice signals. One of these characteristics is the presence of noise,

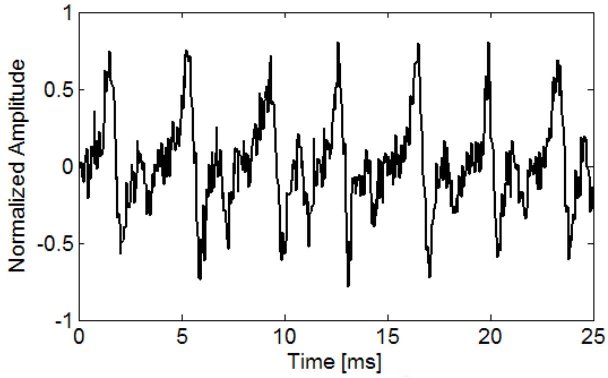

Figure 3. Example of a simulated voice signal.

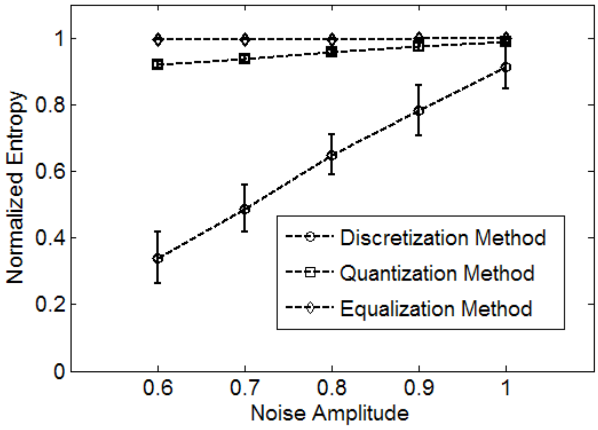

Figure 4. Normalized entropy values of the Gaussian noise signals estimated by the three histogram methods presented.

if it has a high amplitude, it is commonly related to a pathological condition Davis (1979). Therefore, several values of noise levels are tested, as shown in Figure 5, ranging from 6 to 20 dB. As higher noise decreases the regularity of the signal, the entropy increases, as shown for lower signal-to-noise ratios. Again, the same two methods correspond to the expected behavior and the discretization method shows high sensibility to the noise level variation.

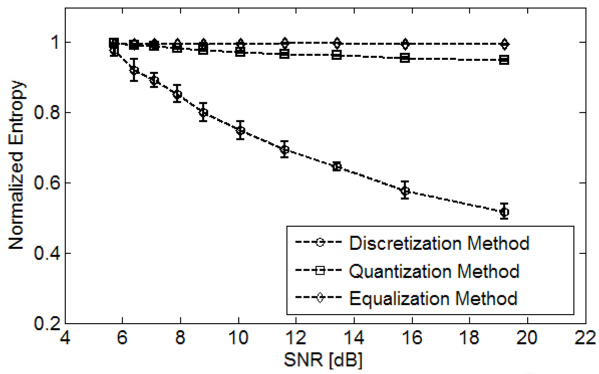

Figure 5. Normalized entropy values of the simulated voice signals estimated by the three histogram methods presented.

Since the equalization method is not intended for this kind of analysis, a test with KL entropy is shown next. In order to do this, the technique employed in Scalassara et al. (2009b) is used, that shows a quantity based on the KL entropy between the signal and its error of prediction. This quantity called predictive power is valuable for the predictability analysis of signals as shown in Schneider and Griffies (1999).

A common predictor for the voice signal is the autoregressive model (AR) as shown in Scalassara et al. (2007). Then an AR of order 10 is used to give predictions of the simulated signals. As shown in Scalassara et al. (2009b), this KL entropy value is proportional to the predictability of the signals, therefore as the noise level increases, the KL entropy is supposed to decrease, as shown in Figure 6, for the three methods. The equalization method showed a smaller sensitivity to the variations, however for low noise levels, it also showed a great variability. The other methods presented basically identical results.

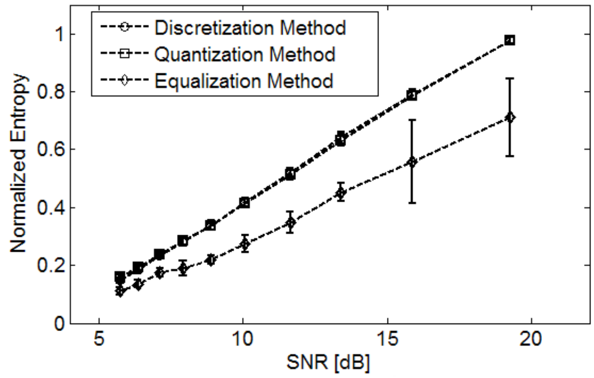

Figure 6. Normalized KL entropy values between the simulated voice signals and their AR(10) prediction error estimated by the three histogram methods presented.

After these tests with noise and simulated signals, we present some results for real voice signals. First, a single voice signal added with several levels of Gaussian noise is used with the three method of entropy estimation. The results are shown in Figure 7 and are similar to the ones obtained for simulated signals (Figure 5).

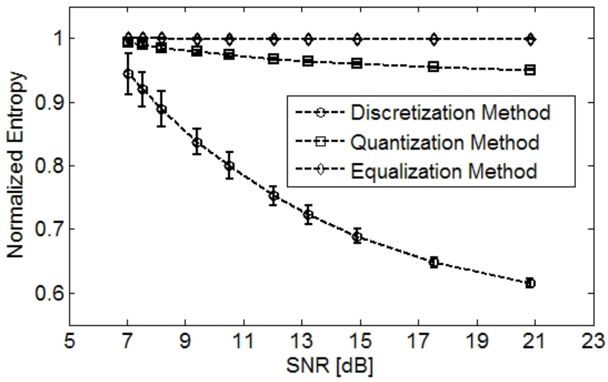

Figure 7. Normalized entropy values of a healthy voice signal with several levels of Gaussian noise. The three methods of histogram estimated are presented.

The values of the KL entropy between the voice signal with several noise levels and their AR(10) error of estimation are shown in Figure 8. The difference between this result and the one of the simulated signals (Figure 6) is that the methods presented higher variability and the equalization method (with less variability than before) still has a higher variability than the other methods.

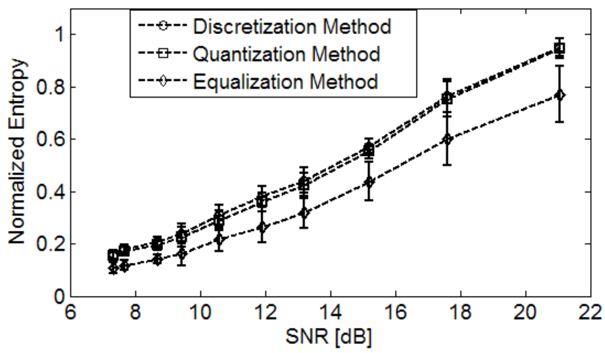

Figure 8. Normalized KL entropy values between a healthy voice signal with several levels of Gaussian noise and their AR(10) prediction error. The three methods of histogram estimated are presented.

Lastly, a test is performed with seven healthy voice signals added with Gaussian noise. Again, ten samples are created for each signal by adding new random noise. The mean, for each noise level, of the KL entropy value of the seven signals are presented in Figure 9. As can be seen, the three methods obtained similar results, showing high variability, probably due to the signals' variability.

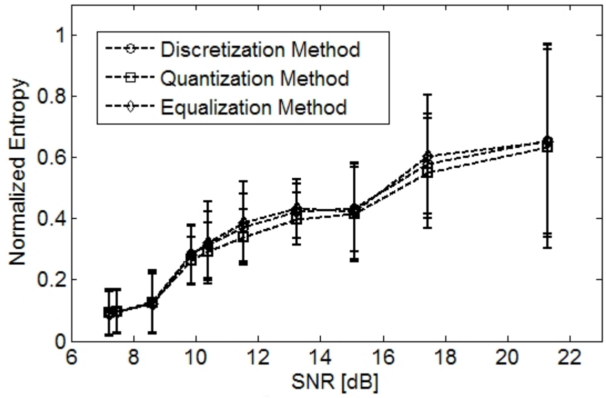

Figure 9. Normalized KL entropy values between seven healthy voice signals with several levels of Gaussian noise and their AR(10) prediction error. The three methods of histogram estimated are presented.

6. CONCLUSIONS

In this paper, we presented a discussion about three methods of nonparametric entropy estimation that are based on histogram approximations of the signals under study. Two of them, namely the discretization method and the quantization method, are true approximations of the histogram, the third, the equalization method, is a modification of the classical histogram to correct the effect of low probable elements in the estimation of the KL divergence.

We presented several tests with signals composed of varying levels of Gaussian noise, simulated voice signals, and healthy voice signals. The Shannon entropy estimations showed high variability of the discretization method and low variability of the quantization method what can be advantageous in some situations. The equalization method is not suitable for this kind of entropy estimation, due to its lack of sensibility to the parameters involved. Also, for real voice signals, the first method presented a great variability of the entropy values for low SNR values.

For the KL entropy of simulated signals, the equalization method presented a great variability of the values for high SNR, what also happens for the real voice signal added with noise. For the simulated signals the first two methods have a better response, but for the real voice signal the difference is smaller. But for the test with the seven voice signals, the results are similar.

The purpose of the equalization method is to reduce the effects of the low probable elements that contaminate the KL entropy estimation of voice signals analysis like the one in Scalassara et al. (2009b). The last test showed that, at a first look, the equalization method is similar to the other ones for real voice signals, but more tests must be performed, especially with pathological signals, that are the ones with more problems of entropy estimation.

Besides that, the AR model has some problems for the prediction of voice signals as shown in Scalassara et al. (2009b), a wavelet model is proposed in that paper, that could prove to be a better test. Also, tests with signals without added noise, in real situations, such as in Scalassara et al. (2009a), should be performed to evaluate the efficacy of the methods.

From all the methods, the discretization is the simplest of all, because the parameters are easy to adjust. The quantization is also simple, but requires more tests with the signals in order to find the better quantization step for analysis. But the equalization method shows some difficulties of parameter adjustment, then more tests could present better results.

The fact is that the KL entropy for predictive power estimation in voice analysis suffer with the effects of low probable elements and the histogram equalization method is a step towards a solution of this problem, or at least present better results.

7. ACKNOWLEDGMENTS

The authors acknowledge the Research Foundation of the State of São Paulo (FAPESP - *Fundação de Amparo à Pesquisa do Estado de São Paulo*) for the support and scholarship. We also thank The University of Iowa, especially the Department of Computer Science, for being so helpful to the first author by the occasion of his six months visit.

8. REFERENCES

Atal, B. S. and Schroeder, M. R. (1967). Predictive coding of speech signals. In *Proceedings of the International Conference on Speech Communications and Processing*, pages 360–361.

Beirlant, J., Dudewicz, E. J., Györfi, L., and Meulen, E. C. (1997). Nonparametric entropy estimation: An overview. *International Journal of the Mathematical Statistics Sciences*, 6:17–39.

Bercher, J. F. and Vignat, C. (2000). Estimating the entropy of a signal with applications. *IEEE Transactions on Signal Processing*, 48(6):1687–1694.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley and Sons, Inc., New York.

Crutchfield, J. P. and Feldman, D. P. (2003). Regularities unseen, randomness observed: levels of entropy convergence. *Chaos*, 13(1):25–54.

Davis, S. B. (1979). *Speech and language: advances in basic research and practice*, chapter Acoustic characteristics of normal and pathological voices, pages 271–314. Academic Publishers, New York.

DelSole, T. (2004). Predictability and information theory - Part I: Measures of predictability. *Journal of the Atmospheric Sciences*, 61(20):2425–2440.

Duin, R. (1976). On the choice of the smoothing parameters for parzen estimators of probability density functions. *IEEE Trans Comput*, 25(11):1175–1179.

Engel, J. (1997). The multiresolution histogram. *Metrika*, 46:41–57.

Erdogmus, D., Hild, K. E., Principe, J. C., Lazaro, M., and Santamaria, I. (2004). Adaptive blind deconvolution of linear channels using renyi's entropy with parzen window estimation. *IEEE Transactions on Signal Processing*, 52(6):1489–1498.

Erdogmus, D. and Principe, J. C. (2000). Comparison of entropy and mean square error criteria in adaptive system training using higher order statistics. In *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 75–80.

Erdogmus, D. and Principe, J. C. (2006). From linear adaptive filtering to nonlinear information processing - the design and analysis of information processing systems. *IEEE Signal Processing Magazine*, 23(6):14–33.

Kleeman, R. (2002). Measuring dynamical prediction utility using relative entropy. *Journal of the Atmospheric Sciences*, 59(13):2057–2072.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:4–7.

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustic Society of America*, 35(3):344–353.

Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*. Prentice Hall, Inc., Upper Saddle River, NJ, USA.

Markel, J. D. (1972). Digital inverse filtering-a new tool for formant trajectory estimation. *IEEE Transactions on Audio and Electroacoustics*, 20(2):129–137.

Moddemeijer, R. (1989). On estimation of entropy and mutual information of continuous distributions. *Signal Processing*, 16(3):233–248.

Moore, C., Manickam, K., and Slevin, N. (2006). Collective spectral pattern complexity analysis of voicing in normal males and larynx cancer patients following radiotherapy. *Biomedical Signal Processing and Control*, 1:113–119.

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, 15:1191–1253.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88:2297–2301.

Pincus, S. M. and Goldberger, A. R. (1994). Physiological time-series analysis: what does regularity quantify? *Am J Physiol*, 266(4 Pt 2):H1643–H1656.

Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, 278(6):H2039–H2049.

Rényi, A. (1960). On measures of information and entropy. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561.

Rosa, M. O., Pereira, J. C., and Grellet, M. (2000). Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Transactions on Biomedical Engineering.*, 47(1):96–104.

Scalassara, P. R., Dajer, M. E., Maciel, C. D., Guido, R. C., and Pereira, J. C. (2009a). Relative entropy measures applied to healthy and pathological voice characterization. *Applied Mathematics and Computation*, 207(1):95–108.

Scalassara, P. R., Dajer, M. E., Maciel, C. D., and Pereira, J. C. (2008a). Voice signals characterization through entropy measures. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, BIOSIGNALS2008*, volume 2, pages 163–170, Funchal, Madeira, Portugal.

Scalassara, P. R., Dajer, M. E., Marrara, J. L., Maciel, C. D., and Pereira, J. C. (2008b). Analysis of voice pathology evolution using entropy rate. In *Proceedings of the Tenth IEEE International Symposium on Multimedia (ISM 2008)*, pages 580–585, Berkeley, CA, USA. IEEE Computer Society.

Scalassara, P. R., Maciel, C. D., Guido, R. C., Pereira, J. C., Fonseca, E. S., Montagnoli, A. N., Barbon, S., Vieira, L. S., and Sanchez, F. L. (2007). Autoregressive decomposition and pole tracking applied to vocal fold nodule signals. *Pattern Recognition Letters*, 28(11):1360–1367.

Scalassara, P. R., Maciel, C. D., and Pereira, J. C. (2009b). Predictability analysis of voice signals: Analyzing healthy and pathologic samples. *IEEE Engineering in Medicine and Biology Magazine*. Article in press.

Scalassara, P. R., Maciel, C. D., Pereira, J. C., Stewart, D., and Oliveira, S. (2009c). Practical problems with the entropy rate estimation. In *Proceedings of the 8th Brazilian Conference on Dynamics, Control and Applications*, page Accepted for publication.

Schneider, T. and Griffies, S. M. (1999). A conceptual framework for predictability studies. *Journal of Climate*, 12(10):3133–3155.

Schürmann, T. and Grassberger, P. (1996). Entropy estimation of symbol sequences. *Chaos*, 6(3):414–427.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Jounal*, 27:379–423, 623–656.

Shimazaki, H. and Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural Computation*, 19:1503–1527.

Viola, P., Schraudolph, N. N., and Sejnowski, T. J. (1996). Empirical entropy manipulation for real-world problems. In *Proceedings of the Neural Information Processing Systems 8 (NIPS'96)*, pages 851–857. MIT Press.

Wachowiak, M. P., Smolíková, R., Tourassi, G. D., and Elmaghraby, A. S. (2005). Estimating of generalized entropies with sample spacing. *Pattern Anal Applic*, 8:95–101.

Wyner, A. J. and Foster, D. (2003). On the lower limits of entropy estimation. *Tech. Rep. Dept. Statistics, Wharton School, Univ. Pennsylvania*.

Zhang, Y., Jiang, J. J., Biazzo, L., and Jorgensen, M. (2005). Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *Journal of Voice*, 19(4):519–528.

Zyczkowski, K. (2003). Rényi extrapolation of shannon entropy. *Open Sys Information Dyn*, 10:297–310.