Comparison between Temporal and Time–Scale Information Measures applied to Speech Recognition*

MARÍA E. TORRES, HUGO L. RUFINER, DIEGO H. MILONE, ANALÍA S. CHERNIZ
Applied Research Group on Signal Processing and Pattern Recognition
Universidad Nacional de Entre Ríos, Facultad de Ingeniería
Universidad Nacional del Litoral, Facultad de Ingeniería y Ciencias Hídricas
CC 47 Suc 3 (3100) Paraná
ARGENTINA
metorres@ceride.gov.ar,{lrufiner, acherniz}@bioingenieria.edu.ar,dmilone@fich.unl.edu.ar

Abstract: While tested with noisy signals, it has been observed deterioration in the performance of automatic speech recognition systems trained with clean signals. In this paper we propose to introduce new parameters to a classical MFCC parametrization to overcome this situation. Continuous multiresolution entropy have shown to be robust to additive noise in applications to different physiological signals. In previous works temporal Shannon and Tsallis entropies, and their corresponding divergences, have been included in different speech related applications. Here we extend the continuous multiresolution entropy notion to different divergences. These parameters are introduced as new dimensions at the pre–processing stage of a speech recognizer and we compare the results obtained with the temporal measures. These new parametrizations are tested with speech signals corrupted with babble and white noise. Their performance are compared with the classical mel cepstra parametrization. Results suggest that information measures, specially those related to multiresolution divergences, provide valuable information that could be considered as an extra component in a pre–processing stage.

Key–Words: Entropy, Divergence, Automatic speech recognition, Continuous multiresolution entropy

1 Introduction

Considerable efforts has been made concerning Automatic Speech Recognition (ASR) over the past two decades. However, important performance deteriorations are observed when ASR systems, trained with clean speech signals, recorded with high quality audio systems, are tested with signals registered with simple home microphones or noisy signals. This is the scope of "robust" speech recognition, which aim is to obtain ASR systems that can be used in real environments, with noise, reverberation, home quality audio systems, etc [1].

There are several pre-processing methods to improve the ASR system's performance, in which it is often supposed that both, signal and noise, are generated by linear systems and that noise can be easily modeled. In practice none of them is a real assumption and the robustness problem of ASR systems is still "open", specially for low signal-to-noise ratio (SNR).

Entropy notions have been used to characterize the complexity degree of different physiological signals. The application of these quantitative measures provides information about the underlying dynamics of non linear systems and helps to gain a better understanding of them.

Shannon and Tsallis entropies and their corresponding divergences have been included in the pre–processing stage of an ASR system [2]. This dimension is computed by means of the evaluation of a temporal complexity measure of speech signals. The information of the temporal evolution of the complexity degree is concatenated to a classical mel frequency cepstral coefficients (MFCC) parameterization. For both, babble and white noise, improvements in the ASR system performance have been obtained. These preliminary results suggest that complexity measures could provide valuable information derived from vocal tract dynamical changes in speech recognition [2].

The multiresolution entropy, proposed by Torres *et al.* in [3], is a tool based on the wavelet transform which gives account of the temporal evolution of the wavelets coefficients' Shannon entropy. Combined with the continuous wavelet transform (CWT) [4], the tool known as continuous multiresolution entropy (CME) has shown to be robust to additive white

[1] This work is supported by A.N.P.C.yT., under Projects PICT #11-12700 and PICT #11-25984. CAID 12–72. CONICET.

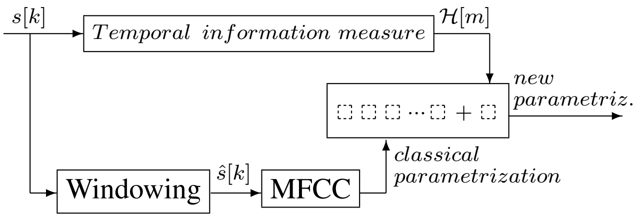

Figure 1: Scheme of a pre–processing stage where temporal information measure have been concatenated to standard MFCC speech processing.

noise in the detection of slight changes in the underlying nonlinear dynamics of physiological signals [5]. In applications to speech signals corrupted with additive noise, good results have been obtained using self-organizing map clustering [6]. These previous results motivate us to explore this tool over the parametrization of an ASR system in order to improve its robustness.

In this paper we compare temporal information measures with continuous multiresolution entropy and its extension to different divergences. These new parameters, taking into account information about the changes in the dynamics of speech signal in time domain and for different scales, will be concatenated to a classical MFCC parametrization. We contrast the performance of a classical ASR system front–end with the ones obtained including the information measures computed in time and time–scale plane, in the presence of noise.

2 Materials and Methods

A modification to the classical speech signal pre-processing stage will be here outlined. First, the temporal computation of entropies and divergences will be presented. Then, the algorithms based on the multiresolution information measures will be explained. At Fig.2 the corresponding block diagram is depicted.

2.1 Information Measures computed in time

In previous works Shannon and Tsallis entropies, Kullback-Leiber distance and q–divergence have been used to characterize speech signals in time domain [2]). The parameters obtained were concatenated to a classical parametrization for its use in an ASR system. This process, depicted in the block diagram of Fig. 1, is briefly explained in what follows. For major details concerning this implementation, we remit the reader to [2].

Let $s[k]$ the discrete evolution of the speech signal at the time $k = 1, \ldots, K$. In order to calculate the different information measures, the probability distributions of $s[k]$ is evaluated. A set of sliding window $W(m, L, \Delta) = \{s[k], k = l + m\Delta, l = 1, ..., L\}$, with $m = 0, 1, 2, ..., M$, is defined. It depends on two parameters: width $L \in \mathbb{N}$ and shift $\Delta \in \mathbb{N}$, that are chosen such that $L \leq K$ (the signal length) and $(K - L)/\Delta = M \in \mathbb{Z}$. This is accomplished in agreement with the windowing performed to obtain the MFCC parametrization of speech signal (see Fig. 1). In this case, the windows length is directly related with maximum speed of significant vocal tract morphology modification [7].

Over each window $W(m, L, \Delta)$, a subset of N disjoint subintervals I_n is considered. With $p_m(I_n)$ is denoted the probability that a given $s[k] \in W(m, L, \Delta)$ belongs to one of such subintervals. Thus, for each window, a set $P[m]$ of N probabilities $p_m(I_n)$ is obtained:

$$P[m] = \{p_m(I_n),\ n = 1, \ldots, N\}, \tag{1}$$

The information measures over each window $W(m, L, \Delta)$ can be computed following the seminal ideas of multiresolution entropies in [3, 4]. Shannon entropy [8] can be writen as:

$$\mathcal{H}_s[m] = -\sum_{n=1}^{N} p_m(I_n) \ln \left(p_m(I_n)\right),$$

Tsallis entropy or q–entropy [9], with $q \neq 1$, is computed as:

$$\mathcal{H}^q{}_s[m] = (q-1)^{-1} \sum_{n=1}^{N} (p_m(I_n) - (p_m(I_n))^q).$$

Having in mind the probability set $P[m]$ mentioned above (1), corresponding to one window $W(m, L, \Delta)$, a second set $R[m] = \{r_m(I_n), n = 1, \ldots, N\}$, where $r_m(I_n)$ corresponds to the probability at the consecutive window $W(m+1, L, \Delta)$, is considered. Thus, Kullback-Leiber distance [10], the relative entropy associated with Shannon entropy, corresponding to two consecutive windows can be computed as:

$$D_s[m] = \sum_{n=1}^{N} p_m(I_n) \ln \left(\frac{p_m(I_n)}{r_m(I_n)}\right).$$

In a similar way the q–divergence [2, 11], related with q–entropy, is:

$$D^q{}_s[m] = \frac{1}{1-q} \sum_{n=1}^{N} p_m(I_n) \left[1 - \left(\frac{p_m(I_n)}{r_m(I_n)}\right)^{q-1}\right].$$

These information measures, $\mathcal{H}_s$, $\mathcal{H}^q{}_s$, D_s and $D^q{}_s$, are concatenated to the MFCC parametrization.

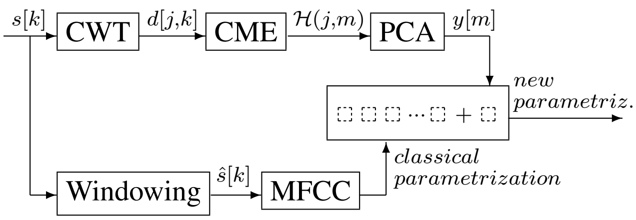

Figure 2: Scheme of the stages of multiresolution information measures computation concatenated to MFCC parametrization explained in the text.

2.2 Continuous Multiresolution Entropy

The complexity measures of $s[k]$ in the time–scale plane are obtained by performing first its "quasi-continuous" wavelet transform, $\Psi_s(a,b)$. This leads to a discretized distribution $\{d[j,k]\} \in \mathbb{R}^{J\times K}$, where $\{d[j,k]\} = \Psi_s(a = j\delta, b)$, $j = 1,\dots,J \in \mathbb{Z}$, $\delta \in \mathbb{R}^+$ and $b = k$ is the time–control. For each fixed j the CWT coefficient's temporal evolution will be denoted as $d_j[k]$ in what follows.

We consider again a set of rectangular sliding windows, but now $W_j(m,L,\Delta) = \{d_j[k], k = l + m\Delta, l = 1,...,L\}$, with $m = 0,1,2,...,M$, also depending of L and Δ chosen as above. We denote with $p_{j,m}(I_n)$ the probability that a given $d_j[k] \in W_j(m,L,\Delta)$ belongs to one of the N disjoint subintervals I_n considered over each window $W_j(m,L,\Delta)$. A set $P[j,m]$ of N probabilities $p_{j,m}(I_n)$ is obtained for each window:

$$P[j,m] = \{p_{j,m}(I_n),\ n = 1,\dots,N\}, \tag{2}$$

where m represents the time–evolution at the considered scale j.

At each fixed scale j and for each fixed m, we can compute the information measures over each window $W_j(m,L,\Delta)$ of wavelet coefficients, in a similar way as the previous section. In this way,

$$\mathcal{H}_d[j,m] = -\sum_{n=1}^{N} p_{j,m}(I_n)\ln\left(p_{j,m}(I_n)\right),$$

provides the entropy value corresponding to the wavelet coefficients on the window $W_j(m,L,\Delta)$. The Shannon entropy evolution at the time–control m is a matrix of elements $CME(a = j\delta, m) = \mathcal{H}_d[j,m]$, wich is named as the continuous multiresolution entropy.

The evolution of Tsallis entropy or q–entropy, $q \neq 1$, computed over each window of $d_j[k]$ is:

$$\mathcal{H}^q{}_d[j,m] = (q-1)^{-1}\sum_{n=1}^{N}\Big(p_{j,m}(I_n) - (p_{j,m}(I_n))^q\Big).$$

$CME_q = \Big(\mathcal{H}^q{}_d[j,m]\Big)$ is the corresponding continuous multiresolution q-entropy matrix.

2.3 Continuous Multiresolution Divergence

In this section, we extend the ideas of multiresolution entropy to the relative information measures. We use the Kullback-Leiber distance [10] and the q–divergence [2, 11].

Considering the probability sets $P[j,m]$ and $R[j,m]$, corresponding to two consecutive window, the Kullback–Leiber divergence can be computed as:

$$D_d[j,m] = \sum_{n=1}^{N} p_{j,m}(I_n)\ln\left(\frac{p_{j,m}(I_n)}{r_{j,m}(I_n)}\right).$$

This procedure accomplished for all scales gives the corresponding continuous multiresolution divergence matrix $CMD = \Big(D_d[j,m]\Big)$.

The relative q–entropy is:

$$D^q{}_d[j,m] = \frac{1}{1-q}\sum_{n=1}^{N} p_{j,m}(I_n)\left[1 - \left(\frac{p_{j,m}(I_n)}{r_{j,m}(I_n)}\right)^{q-1}\right]$$

and the Continuous Multiresolution q–Divergence is $CMD_q = \Big(D^q{}_d[j,m]\Big)$.

2.4 Different CME and CMD–based parametrization approaches

Once the multiresolution information measures are obtained, principal component analysis (PCA) is performed in order to keep a relative low dimension for the final coefficients vector [12]. It is used here in three different ways, described in what follows for the $\mathbf{CME}$. This can be easily extended to the other multiresolution information measures.

Method 1. First PC ($\mathbf{PC_1}$): Given $\mathbf{CME}$, we obtain the matrix of principal component as:

$$\mathbf{Y} = \mathbf{Q}^T\left(\mathbf{CME}\right)^*, \tag{3}$$

where $(\mathbf{CME})^*$ is the statistical normalized matrix and $\mathbf{Q}$ contain the eigenvector of its correlation matrix.

The row of $\mathbf{Y}$ corresponding to the maximum eigenvalue of $\sigma_{\mathbf{CME}}$ is its principal component and we denote it as $\mathbf{y_1}$. This vector evolves with the time–control m and it will be concatenated to the classical MFCC to obtain our new parametrization.

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
M. E. Torres, H. L. Rufiner, D. H. Milone & A. S. Cherniz; "Comparison between temporal and time-scale information measures applied to speech recognition"
WSEAS Transactions on signal Processing, Vol. 2, pp. 1153--1159, Sep, 2006.

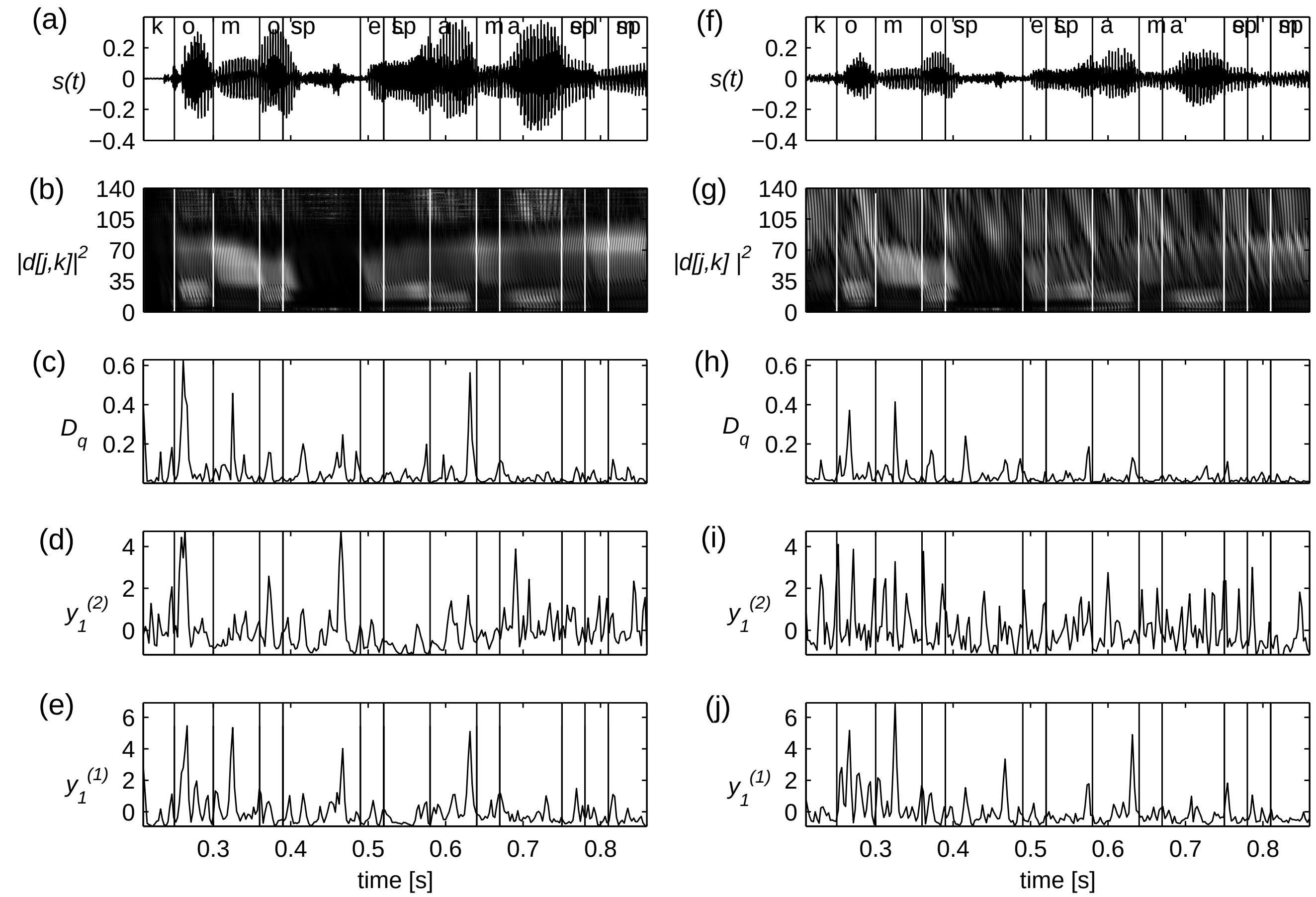

Figure 3: Fist column shows processing corresponding to clean speech signal and second column processing corresponding to noisy signal. (a) Labeled speech signal. (b) Scalogram corresponding to the signal displayed in (a). (c) q–divergence ($q = 0.5$) of temporal signal showed in (a). (d) Component $\mathbf{y}_1^{(1)}$ of method PC$_{SD}$ using CMD_q ($q = 0.2$) in scalogram (b). (e) Component $\mathbf{y}_1^{(2)}$ of previous method. (f) The same signal shown in (a) with additive background conversation noise (10 dB SNR). (g) Scalogram corresponding to the signal displayed in (f). (h) q–divergence ($q = 0.5$) of temporal signal showed in (f). (i) Component $\mathbf{y}_1^{(1)}$ of method PC$_{SD}$ applied over scalogram (g) using CMD_q ($q = 0.2$). (j) Component $\mathbf{y}_1^{(2)}$ corresponding to the previous PC$_{SD}$ method.

Method 2. First and second PC (PC$_{12}$): Now we consider the two first components $\mathbf{y}_1$ and $\mathbf{y}_2$ of $\mathbf{Y}$. Both vectors are concatenated to the MFCC to generate the new parametrization vector.

Method 3. Scale dependent PC (PC$_{SD}$): Here we consider two submatrices from $\mathbf{CME}$: $\mathbf{U}^{(1)} = \{\mathbf{CME}^*[j\delta, m], = 1, ..., J/2\}$ and $\mathbf{U}^{(2)} = \{\mathbf{CME}^*[j\delta, m],\ j = 1 + J/2, ..., J\}$. Applying (3) to each of them, with the corresponding matrix $\mathbf{Q}^{(i)}$ for $i = 1, 2$, we obtain the principal component matrices $\mathbf{Y}^{(i)}$. From them we obtain $\mathbf{y}_1^{(1)}$ and $\mathbf{y}_1^{(2)}$, which are concatenated to the classic MFCC parametrization.

As an example, we show in Fig. 3 the behavior of q–divergence, computed in time domain and in time–scale plane, while applied to a speech signal with and without noise. In (a) a part of the labeled speech signal of sentence: "¿Cómo se llama el mar que baña Valencia?" (*What is the name of the sea that border Valencia?*) is shown. Fig. 3(b) shows the scalogram ($|d[j, k]|^2$) corresponding to the signal showed in (a), obtained with the Daubechies wavelet of order 16. In Fig. 3(c) the temporal q–divergence, with $q = 0.5$, of clean speech signal is shown. Fig. 3(d) and 3(e) depict the components $\mathbf{y}_1^{(1)}$ and $\mathbf{y}_1^{(2)}$, corresponding to method PC$_{SD}$ used with $\mathbf{CMD}_q$, for $q = 0.2$. Figs. 3(f–j) show results obtained for the same signal but corrupted with additive background conversation noise at 10dB SNR. It can be observed from Figs. 3(e) and 3(j) that higher values in component $\mathbf{y}_1^{(1)}$ are coincident for clean and noisy signal, in contrast with the

sinc(*i*) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
M. E. Torres, H. L. Rufiner, D. H. Milone & A. S. Cherniz; "Comparison between temporal and time-scale information measures applied to speech recognition"
WSEAS Transactions on signal Processing, Vol. 2, pp. 1153--1159, Sep, 2006.

temporal measure. In component $\mathbf{y}_1^{(2)}$ of Figs. 3(i) appears more peaks than in Figs. 3(d), which are related with noise. This is associated with the fact that this measure was computed from the upper half of scalogram of Fig. 3(g), where, comparing with Fig. 3(b), it can be observed the structures related with noise in this part. This suggests that an appropriate inclusion of this tool to represent dynamical changes of vocal tract could improve the ASR system performance in the presence of noise, making it more robust.

2.5 Automatic speech recognition system:

In order to compare the classical parametrization with the alternatives presented here, we build a state of the art ASR system [13] for Spanish speech corpus.

A 3 state semi-continuous HMMs (SCHMMs) have been used for context–independent phonemes and silences. Observations probability density functions have been modeled with Gaussian mixtures. A complete model was built for all the phrases and four reestimations have been accomplished using the Baum-Welch algorithm. Parameters tying was accomplished using a pool of 200 Gaussians for each model state [14]. Finally the remaining reestimations have been computed in order to complete the total of sixteen. For language modeling, backing-off smoothed bigrammars have been estimated with transcriptions of the training database [15].

For the reference system, each phrase has been normalized in mean, pre-emphasized and Hamming windowed in segments of 25 ms length, shifted 10 ms. Each segment have been parameterized with 28 coefficients: 13 MFCC, 1 energy coefficient (E) and their temporal derivatives (Δ MFCC+ΔE) [7].

For the parametrization accomplished with temporal entropies and divergences we have:

- TE method: MFCC | E | $\mathcal{H}$ | ΔMFCC | ΔE | $\Delta\mathcal{H}$,
- TD method: MFCC | E | D | ΔMFCC | ΔE | ΔD,

where $\mathcal{H}$ can be either, Shannon entropy ($\mathcal{H}_s$) or q–entropy (${\mathcal{H}^q}_s$), and D the Kullback–Leiber distance (D_s) or the q–divergence (${D^q}_s$).

For each of the multiresolution methods the following parametrizations were considered:

- PC$_1$ method: MFCC | E | $\mathbf{y}_1$ | ΔMFCC | ΔE | $\Delta\mathbf{y}_1$,
- PC$_{12}$ method:MFCC | E | $\mathbf{y}_1$ | $\mathbf{y}_2$ | ΔMFCC | ΔE | | $\Delta\mathbf{y}_1$ | $\Delta\mathbf{y}_2$,
- PC$_{SD}$ method: MFCC| E | $\mathbf{y}_1^{(1)}$ | $\mathbf{y}_1^{(2)}$ | ΔMFCC| | ΔE | $\Delta\mathbf{y}_1^{(1)}$ | $\Delta\mathbf{y}_1^{(2)}$.

2.6 Database and cross validation tests

A subset of the Albayzin speech corpus [16], consisting of 600 sentences, 200 words vocabulary, related to Spanish geography, was used. Speech utterances, registered in a recording study, had 3.55 secs. phrase duration average, and they were spoken by 6 males and 6 females from the central area of Spain (average age 31.8 years). In order to test robustness of the ASR system, speech signals corrupted with background conversation and white noise from the NOISEX-92 database were used [17]. Both noises have been mixed additively with data at different SNR levels. Data was re–sampled at 8 kHz and 16 bits of resolution.

Tests were accomplished using the *leave-k-out* cross validation method. Ten models were built and trained with different partitions on the same subset of data. For each partition, 80% of sentences have been randomly selected for system training and the remaining 20% has been used for testing. Recognition has been evaluated computing the word error rate (WER), considering as errors the word deletion and substitutions [13]. The percentage of relative error improvement has been computed as:

$$\Delta\varepsilon_{\%} = \left(\varepsilon_{ref} - \varepsilon\right)/\varepsilon_{ref},$$

where ε is the WER value of the different methods and ε_{ref} is the reference WER of baseline front–end.

3 Results and discussion

We present and discuss here the results obtained while contrasting the recognition obtained with methods presented in this work and classical front–end. We also compare the ASR system performance obtained when information measures, concatenated to MFCC parametrization, are computed in time and time–scale domain.

In Fig. 4 are compared the WER obtained with classical parametrization *vs.* parametrization with both, temporal and multiresolution information measures concatenation, using babble and white noise at different SNRs. We do not shows WER corresponding to method PC$_1$ because the results obtained were not satisfactory.

Fig. 4 (a) shows that entropy measures, using signals corrupted with babble noise, provides good results, specially for 10 and 15 dB SNRs, where divergences also present a lower WER. A similar performance was achieved when temporal information measures were used with white noise, as can be seen in Fig. 4 (d).

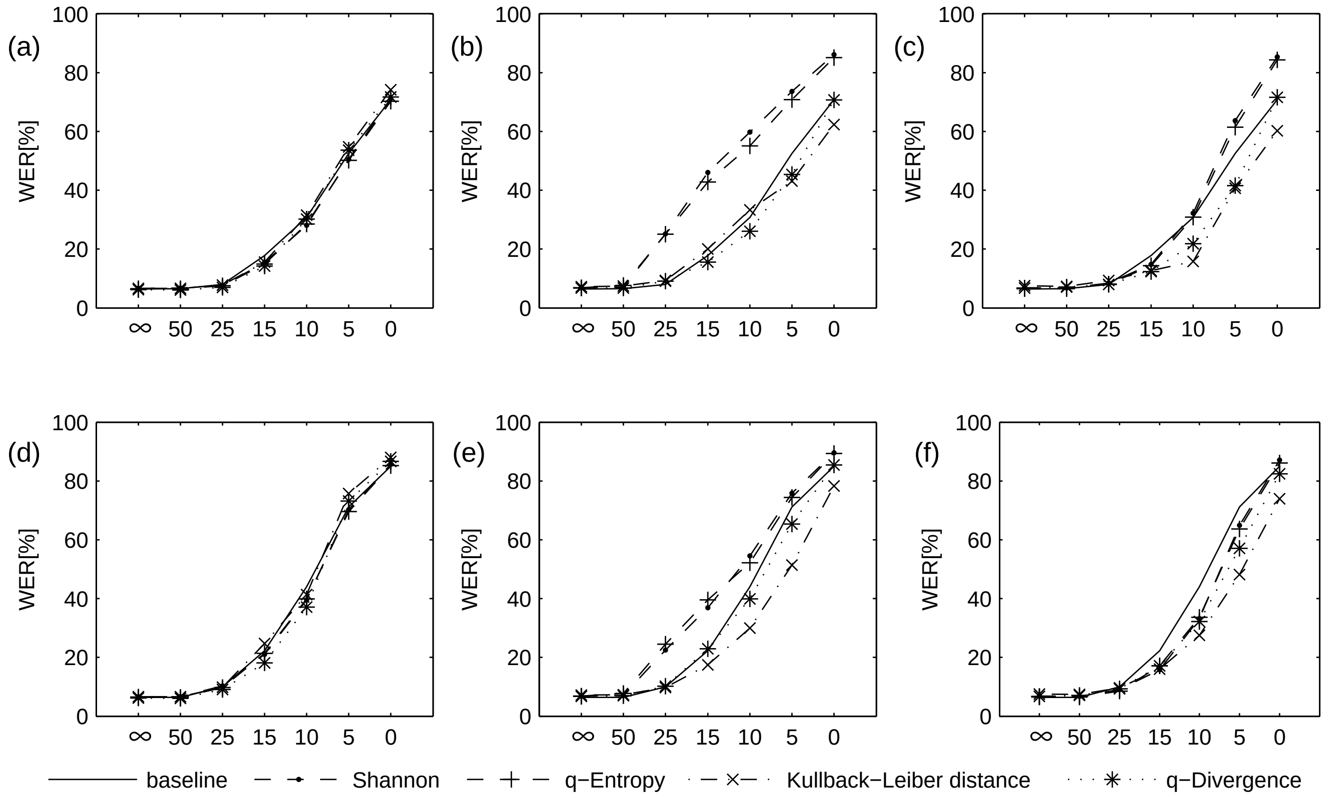

Figure 4: Word error rates of ASR system *vs.* SNR. Comparison between the classical pre–processing (solid line) and parametrizations with different information measures: Shannon entropy, q–entropy, Kullback-Leiber distance and q–divergence; computed in: (a) and (d) time domain, (b) and (e) time–scale plane with method PC_{12}, (c) and (f) time–scale plane with method PC_{SD}. Signals corrupted with babble additive noise were used in (a), (b) and (c) and signals corrupted with white additive noise in (d), (e) and (f).

When method PC_{12} is applied to signals corrupted with babble noise (Fig. 4 (b)) the concatenation of q–divergence measure provides a WER lower than baseline for 5, 10 and 15 dB SNRs. Kullback–Leiber distance gives good results for the SNRs 5 and 0 dB in this case. When white noise is used, 4 (e), this divergence presents the best result, especially for low SNRs.

Fig. 4 (c) shows that method PC_{SD} with babble noise presents good results when divergences were concatenated to the classical parametrization. When white noise was used (Fig. 4 (f)) all the information measures have better performance than baseline for 5, 10 and 15 dB SNRs.

These results suggest that method PC_{SD} offers the best performance, specially in combination with relative information measures. This could be related to the characteristics of $\mathbf{y}_1^{(1)}$ and $\mathbf{y}_1^{(2)}$ already observed in Fig. 3, where the structures belonging to speech signal are mainly at the lowest scales ($\mathbf{y}_1^{(1)}$). In presence of noise the structures at the higher scales are highly modified, suggesting that babble noise information is more concentrated at these scales.

Finally, in Fig. 4also the WER curves obtained using the q–divergence computed in time, methods PC_{12} and PC_{SD}, and baseline are compared. We can appreciate that method PC_{SD} has the best performance of parametrization proposed for both, babble and white noise, specially in the range of 5 to 25 dB.

In order to evaluate the statistical significance of these results, we have estimated the probability that a given recognizer is better than the reference system ($\Pr(\epsilon < \epsilon_{ref})$). To perform this test we assumed the statistical independence of the recognition errors for each word and we approximated the binomial distribution of the errors by means of a Gaussian distribution. This is possible because we have a high number of words (11077 words, if we take into account all the test partitions).

Table 1 shows the relative errors $\Delta\varepsilon_{\%}$ obtained with each method using signal corrupted with babble noise at differents SNR. In Table 2 the results corresponding to additive white noise are presented. A positive value means that the results provide lower error than the baseline. We have marked in bold fonts those results with higher statistical significance than 99.99%. It can be observed that for babble noise, method PC$_{SD}$ with the Kullback and the parametric divergence provides $\Pr(\epsilon < \epsilon_{ref}) > 99.99\%$, for SNR between 5 and 15 dB. For white noise it is obtained for 0 to 15 dB, for the same methods.

4 Conclusions

In this work different information measures, computed in time and time–scale plane were included in the parametrization of an ASR system. Methods presented here were tested with speech signals corrupted with babble and white noise. Performance of these approaches were compared between them and with classical MFCC parametrization.

Method PC$_{SD}$ provided shows the most important improvements of the recognition rates over the baseline. This behavior was observed both in babble and white noise, specially for 15, 10 and 5 dB SNRs and for the relative informations.

These results suggest that the CME related measures gives valuable information to the ASR system in order to perform the recognition, even in the presence of additive noise. This information allows to the ASR system to track dynamical changes of the system. Because of that they would be considered to be included as an extra component in the pre–processing stage.

References:

[1] O. Viiki, editor, Noise Robust ASR, Speech communication, Special Issue, 34, 1–2, 2001.

[2] H. L. Rufiner, M. E. Torres, L. Gamero, and D. H. Milone, Physica A **332**, 496 (2004).

[3] M. Torres, L. Gamero, and E. D'Attellis, Latin American Applied Research **53**, 53 (1995).

[4] M. E. Torres, L. Gamero, P. Flandrin, and P. Abry, in *SPIE'97 Wavelet Applications in Signal and Image Processing V*, edited by A. F. L. Akram Aldroubi and M. Unser (SPIE Int. Soc. for Optical Engineering, Washington, 1997), Vol. 3169, pp. 400–407.

[5] M. M. Añino, M. E. Torres, and G. Schlotthauer, Physica A **324**, 645 (2003).

[6] H. M. Torres, J. A. Gurlekian, H. L. Rufiner, and M. E. Torres, Physica A: Statistical Mechanics and its Applications **361**, 337 (2006).

[7] J. Deller, J. Proakis, and J. Hansen, *Discrete Time Processing of Speech Signals* (Macmillan Publishing, New York, 1993).

[8] C. Shannon, The Bell System Technical Journal **27**, 379 (1948).

[9] C. Tsallis, Chaos, Solitons and Fractals **6**, 539 (1995).

[10] T. M. Cover and J. A. Thomas, *Information Theory* (John Wiley and Sons, NY, 1991).

[11] M. E. Torres, Ph.D. thesis, Universidad Nacional de Rosario - Argentina, 1999, (Math. D. Thesis).

[12] M. Akay, *Detection and Estimation Methods for Biomedical Signals* (Academic Press, San Diego, CA, 1996).

[13] S. Young *et al.*, *The HTK Book (for HTK Version 3.1)*, Cambridge University Engineering Department, `http://htk.eng.cam.ac.uk`, 2002.

[14] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, New Jersey, 1993).

[15] F. Jelinek, *Statistical Methods for Speech Recognition* (The MIT Press, Cambrige, Masachussets, 1999).

[16] J. E. D. Verdejo *et al.*, in *Proceedings of the First International Conference on Language Resources and Evaluation* (European Language Resources Association, Granada, 1998), Vol. 1, pp. 497–502.

[17] A. Varga and H. Steeneken, Speech Communication **12**, 247 (1993).

sinc(i) Laboratory for Signals and Computational Intelligence (http://fich.unl.edu.ar/sinc)
M. E. Torres, H. L. Rufiner, D. H. Milone & A. S. Cherniz; "Comparison between temporal and time-scale information measures applied to speech recognition"
WSEAS Transactions on signal Processing, Vol. 2, pp. 1153--1159, Sep, 2006.

Table 1: Relative error improvement ($\Delta\epsilon_{\%}$) of the different measures compared with the classical front–end for speech signal corrupted with *babble* noise. Bold numbers indicate $\Pr(\epsilon < \epsilon_{ref}) > 99.99\%$.

	SNR_{dB}	∞	50	25	15	10	5	0
TE		0.89	1.48	6.46	17.98	**14.83**	1.52	-4.35
$\text{TE}_{q=0.5}$		2.55	3.79	6.84	20.75	13.46	2.18	-2.77
TD		0.96	1.66	4.31	17.19	4.13	-6.73	-8.56
$\text{TD}_{q=0.5}$		**8.04**	**10.71**	**14.11**	**24.31**	8.16	-4.53	-4.96
	PC_1	1.00	-2.31	-57.24	-59.67	-51.04	-30.82	-18.19
CME	PC_{12}	-7.14	-7.29	-68.55	-61.25	-48.52	-28.74	-17.74
	PC_{SD}	-6.28	-3.57	-5.81	**20.63**	-4.46	-17.60	-16.97
	PC_1	-0.62	-2.09	-59.42	-59.10	-48.23	-29.11	-16.52
$\textbf{CME}_{q=0.2}$	PC_{12}	-7.46	-13.59	-68.37	-58.34	-44.13	-25.87	-16.72
	PC_{SD}	-2.20	0.78	-2.28	**24.27**	-0.23	-14.65	-15.97
	PC_1	-5.71	-2.71	-6.01	-16.62	-24.08	-10.99	-29.16
CMD	PC_{12}	-10.77	-11,10	-15.07	-10.83	-7.50	**21.77**	**13,74**
	PC_{SD}	-14.67	-9.99	-14.78	**40.86**	**94.52**	**29.07**	**17.60**
	PC_1	-2.70	-1.87	-4.50	-13.53	-15.40	-5.85	-3.90
$\textbf{CMD}_{q=0.2}$	PC_{12}	-5.43	-2.58	-12.52	**14.63**	**18.01**	**15.76**	0.16
	PC_{SD}	-5.43	-7.76	-1.11	**44.75**	**41.39**	**26.18**	-1.02

Table 2: Relative error improvement ($\Delta\epsilon_{\%}$) of the different measures compared with the classical front–end for speech signal corrupted with *white* noise. Bold numbers indicate $\Pr(\epsilon < \epsilon_{ref}) > 99.99\%$.

	SNR_{dB}	∞	50	25	15	10	5	0
TE		0.89	-1.88	1.20	8.60	13.02	2.25	-1.26
$\text{TE}_{q=0.1}$		2.55	1.48	7.19	7.30	12.29	3.13	-2.00
TD		0.96	-2.01	5.90	-7.43	8.99	-5.45	-4.07
$\text{TD}_{q=0.5}$		**8.04**	**7.23**	**12.83**	**21.25**	**18.38**	-1.91	-2.44
	PC_1	1.00	-1.26	-47.16	-45.17	-30.91	-14.89	-7.44
CME	PC_{12}	-7.14	-9.78	-55.35	-39.67	-19.17	-6.18	-5.10
	PC_{SD}	-6.28	-5.15	10.61	**40.68**	**32.51**	**9.55**	-2.39
	PC_1	-0.62	-1.91	-52.48	-46.62	-27.69	-14.65	-6.91
$\textbf{CME}_{q=0.2}$	PC_{12}	-7.46	-18.01	-59.04	-43.70	-15.51	-4.31	-4.87
	PC_{SD}	-2.20	-1.68	**18.02**	**30.21**	**31.00**	**11.77**	-1.25
	PC_1	-5.71	-4.30	7.58	-4.23	**16.31**	**6.94**	**4.62**
CMD	PC_{12}	-10.77	-12.37	2.29	**27.57**	**47.07**	**38.50**	**8.58**
	PC_{SD}	-14.67	-13.65	3.92	**39.16**	**60.27**	**47.60**	**14.92**
	PC_1	-2.70	-1.85	-7.56	-9.95	-9.98	-4.87	-0.60
$\textbf{CMD}_{q=0.2}$	PC_{12}	-5.43	-8.66	-1.69	-2.68	**10.55**	**8.78**	-0.47
	PC_{SD}	-5.43	-10.65	7.22	**30.02**	**36.76**	**24.51**	**3.13**