Recognition of Emotions in Speech

Enrique M. Albornoz, María B. Crolla and Diego H. Milone

Grupo de investigación en señales e inteligencia computacional
Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral
Consejo Nacional de Investigaciones Científicas y Técnicas

Abstract. The recognition of the emotional state of the speaker is a research area that has received great interest in the last years. The main goal is to improve voiced-based human-machine interactions. Most of the recent research on this domain has focused the studies in the prosodic features and the speech signal spectrum characteristics. However, there are many other characteristics and techniques which have not been explored in emotion recognition systems. In this work, a study of the performance of Gaussian mixtures models and hidden Markov models is presented. For the hidden Markov models, several configurations have been used, including an analysis of the optimal number of states. Results show the influence of number of Gaussian components and states. The performance of the classifiers has been evaluated with 3 to 7 emotions in spontaneous emotional speech and with speaker independence. In the analysis of three emotions: neutral, sadness and anger, the recognition rate by the Gaussian mixture classifiers was 93% and with hidden Markov models it was 97%. In the recognition of seven emotions, the accuracy was 67% with the Gaussian mixtures models and 76% in the evaluation of hidden Markov models.

1 Introduction

Today, with the constant development of new information technologies, it becomes more necessary to improve the human-machine interaction. if machines would be able to automatically recognize the speaker emotional state through the speech, their interaction would be improved. Speech-based human-machine interaction systems can recognize "what was said" and "who said this" using speech recognition and speaker identification techniques. If an emotion recognition system would be added, it could know "what was the emotional state when she/he said" in order to act consequently, offering a more natural interaction for the human like result [1].

Most of the previous works in emotion recognition have been based in the analysis of speech prosodic features and spectrum [2, 1, 3]. Two of the most used methods in emotion recognition are the support vector machines (SVM) [4] and the Gaussian mixture models (GMM) [5]. Also the hidden Markov models (HMM) have been explored for this task, although to a lesser extent [1, 6].

The use of 39 candidate features, from which were selected the 5 more representatives to model the speech signal, was proposed [1]. The candidate features

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

were the energy, the fundamental frequency, the formant frequencies, mel frequency cepstrum coefficients and mel frequency sub-band energies, and their first and second derivatives. The support vector machines and hidden Markov models, to classify five emotional states were used. The second one achieves a 99.5% in five emotions, with a Danish database recorded by 2 female and 2 male actors, but it was never mentioned the number of utterances used. Also, authors conclude that the mel frequency cepstrum coefficients are not suitable for emotion recognition in speech, because they achieved less than 60% recognition rate in theirs experiments.

Nogueiras et al. [7] based their study in two prosodic features: the fundamental frequency and the energy, using hidden semi-continuous Markov models. The accuracy in recognising seven different emotions is 80%, using the best combination of low level features and HMM structure, in a speaker dependent domain. The corpus used was recorded by two professional people, an actress and an actor.

In other work [3], it was tested 3 different recognizers, changing the modeled features and the classification methods. In a first model, only six prosodic features and a GMM is used to obtain an accuracy of 86.71%. In the second model, they take six prosodic features and a SVM based classifier, obtaining a classification rate of 92.32%. In the last proposed, mel frequency cepstrum coefficients and their first and second derivatives were selected and used in a GMM of 512 components, this achieve a 98.4% recognition rate. An important drawback there, is that the used corpus was recorded by only one actress.

In this work, the relevance of implicit information in the speech signal and the design of an automatic emotion recognition system with this information are investigated. Two classification methods are proposed, one of them using Gaussian mixtures and the other using hidden Markov models. In the classification stage a 10 speakers corpus and up to 7 emotional states have been explored: happiness, anger, fear, boredom, sadness, disgust and neutral.

In the next section, the speech signal analysis and the classification methods are briefly introduced. Section 3 describes the emotional speech data base and the experiments. Section 4 deals with the classification performance and discussion. Finally, conclusions and future works are presented.

2 Features Extraction and Classification Methods

2.1 Speech Signal Analysis

The usual hypothesis in speech analysis is that this signal remains stationary by frames. The maximum speed of morphological variation in the vocal tract explains its validity, therefore it is possible to consider that speech signal remains stationary in periods of approximately 20 ms [8].

There are several windows for the short term analysis: the rectangular window, the Hamming window and the Blackman window to name only a few [9]. The selection of window type depends on the application and it is based on a

tradeoff between Gibbs phenomena reduction and frequency resolution. Hamming window is the most used window in speech application [8].

Keeping in mind that transforms are applied on the windowed signal, frame by frame, a briefly review of the standard cepstral transforms is presented in the next paragraphs.

Cepstral Coefficients (CC): the cepstral analysis is a special case of the homomorphic processing methods [10]. It is applied in speech signal analysis to extract the vocal tract information. Based on the Fourier Transform (FT) [11], the cepstrum is defined as $cc(n) = FT^{-1}\{\mathsf{log}|FT\{x(n)\}|\}$.

Mel Frequency Cepstral Coefficients (MFCC): a *mel* is a unit of measure for perceived pitch or frequency of a tone. An analysis that combines the cepstrum properties and experimental results about the human perception of pure tones brings out the MFCC representation. The mel scale was determined as a mapping between real frequency scale (Hz) and the perceived frequency scale (mel) $F_{mel} = 1000\,\mathsf{log}_2\left[1 + \frac{F_{Hz}}{1000}\right]$.

To obtain the MFCC coefficients, the FT is calculated and this spectrum is filtered by a filter bank in the mel domain [8]. Then, the logs of the powers at each of the mel frequencies are taken. Finally, the FT^{-1} is replaced by the Cosine Transform (CT) in order to simplify the computation and is used to obtain the MFCC of the list of mel log powers.

2.2 Gaussian Mixtures

Even though Gaussian distributions have important analytical properties, they have limitations to model real data. Suppose that real data are concentrated in two well-separated groups, so a simple Gaussian distribution will not catch their structure properly, whereas a superposition of two distributions would fit better the real data distribution. Such superpositions, formed as a finite linear combination of more simple distributions can be called *mixture distribution model* or *mixture model*, and they are widely used in statistical pattern recognition. If each distribution is a simple Gaussian density, the model is called *mixture of Gaussians* [5] and read as

$$p(\mathbf{x}) = \sum_{k=1}^{K} \omega_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k), \tag{1}$$

where the mixing coefficients verify $\sum_k \omega_k = 1$ and $0 \leq \omega_k \leq 1$ for all k.

If the model is parametrically defined as $\lambda = \{\mu_k, \Sigma_k, \omega_k\}$ with $k = 1, 2, \ldots, K$, then it will be determined by the means vector μ, the covariance matrix $\mathbf{\Sigma}$ and the mixing coefficients vector ω. With the *expectation maximization* (EM) framework these parameters can be estimated [5]. The method starts with an initial estimation of the parameters $\lambda(0)$ from whose the new parameters $\lambda(1)$ of the model are estimated, and this is repeated iteratively up to the achievement of some convergence criterion.

Given an observation $\mathbf{x}_n$, the equation 1 could be expressed as

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

$$p(\mathbf{x}_n) = \sum_{k=1}^{K} p(k)p(\mathbf{x}_n|k), \tag{2}$$

where $p(k) = \omega_k$ and $p(\mathbf{x}_n|k)$ is the k-th normal distribution. Then, by Bayes' theorem, the posterior probability can be written as

$$\gamma_{nk} \equiv p(k|\mathbf{x}_n) = \frac{p(k)p(\mathbf{x}_n|k)}{\sum_l p(l)p(\mathbf{x}_n|l)} = \frac{\omega_k \mathcal{N}(\mathbf{x}_n|\mu_k, \Sigma_k)}{\sum_l \omega_l \mathcal{N}(\mathbf{x}_n|\mu_l, \Sigma_l)}. \tag{3}$$

To model the distribution of an observation set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ by means of GMM, the function $-\mathsf{log}\, p(\mathbf{X}|k)$ is maximized. The derivatives with respect to μ_k, Σ_k and ω_k are equaled to zero in order to obtain the re-estimation formulas

$$\tilde{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} \mathbf{x}_n \tag{4}$$

$$\tilde{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} (\mathbf{x}_n - \mu_k^{\mathsf{new}})(\mathbf{x}_n - \mu_k^{\mathsf{new}})^T \tag{5}$$

$$\tilde{\omega}_k = \frac{N_k}{N} \tag{6}$$

with $N_k = \sum_{n=1}^{N} \gamma_{nk}$.

By using a sufficient number of Gaussians, and by adjusting their means and covariances as well as the coefficients in the linear combination, almost any continuous density can be approximated to arbitrary accuracy [5].

2.3 Hidden Markov Models

Hidden Markov models are basically statistical models that describe sequences of events. For classification tasks, a model is estimated for every signal type. Thus, it would be take into account as many models as signal types to recognize. During classification, a signal is taken and the probability for each signal given the model is calculated. The classifier output is based on the maximum probability that the model has generated this signal. A hidden Markov model has two basic elements: a Markov process and a set of output probability distributions [10].

An HMM is defined by an algebraic structure $\Theta = \langle \mathcal{Q}, \mathbb{O}, \mathbf{A}, \mathcal{B} \rangle$, where $\mathcal{Q}$ is the set of possible states, $\mathbb{O}$ is the observable space, $\mathbf{A}$ is the matrix of transition probabilities and $\mathcal{B}$ is the set of observation (or emission) probability distributions [10]. It will never be able to determine the present state, looking only the output, because every state can emit one of the symbols Therefore, the internal behavior of the model remains "hidden" and this is the motivation of the name for the model.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

In the continuous HMM (CHMM), instead of discrete probability distribution $b_j(i)$, for each symbol i, a probability distributions expressed by a mixture is modeled as

$$b_j(\mathbf{x}) = \sum_{k=1}^{K} c_{jk} b_{jk}(\mathbf{x}) \tag{7}$$

where K is the number of mixture components and b_{jk} is the probability density given by the k component of the mixture (generally a normal distribution).

Given a sequence of acoustic evidences $\mathbf{X}^T$, the training is a maximization of the probability density function

$$\Pr\left(\mathbf{X}^T \left|\Theta\right.\right) = \sum_{\forall \mathbf{q}^T} \left\{ \Pr\left(\mathbf{X}^T \left|\mathbf{q}^T, \Theta\right.\right) \Pr\left(\mathbf{q}^T \left|\Theta\right.\right) \right\} \tag{8}$$

In a CHMM, the parameters are efficiently estimated with the forward-backward algorithm [8]. In models where all the distributions are Gaussians and using an auxiliar function, the re-estimation formulas of state transitions $\tilde{a}_{ij}$, the distribution weights $\tilde{c}_{jk}$, the mean vectors $\tilde{\mu}_{jk}$ and the covariance matrices $\tilde{U}_{jk}$ are [10]

$$\tilde{a}_{ij} = \frac{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_{t-1} = i, q_t = j \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}}{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_{t-1} = i, \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}} \tag{9}$$

$$\tilde{c}_{jk} = \frac{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j, k_t = k \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}}{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}} \tag{10}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j, k_t = k \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)} \mathbf{x}_t}{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j, k_t = k \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}} \tag{11}$$

$$\tilde{\mathbf{U}}_{jk}^{-1} = \frac{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j, k_t = k \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)} (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})^{\mathrm{T}}}{\sum_{t=1}^{T} \frac{p\left(\mathbf{X}^T, q_t = j, k_t = k \left|\Theta\right.\right)}{p\left(\mathbf{X}^T \left|\Theta\right.\right)}} \tag{12}$$

Table 1. Distribution of emotional corpus.

Emotion	Anger	Boredom	Disgust	Fear	Joy	Sadness	Neutral
Number of utterances	127	81	46	69	71	62	79

3 Emotional Speech Corpus and Experiments

The emotional speech signals used were taken from an emotional speech data base [12], developed by the Communication Science Institute of Berlin Technical University. This corpus had been used in numerous studies [13, 14] and allows an analysis with speaker independence and is freely available on Internet[1].

The corpus, formed by 535 utterances, includes sentences performed in 6 ordinary emotions and sentences in neutral emotional state. These emotions are the most frequently used in this domain and allow comparisons with others works. These are labeled as: happiness (joy), anger, fear, boredom, sadness, disgust and neutral (Table 1). The same texts were recorded in german by ten actors, 5 female and 5 male, which allows conclusions over the whole group, comparisons between emotions and comparisons between speakers. The corpus consist of 10 utterances for each emotion type, 5 short and 5 longer sentences, from 1 to 7 seconds. To achieve a high audio quality, these sentences were recorded in an anechoic chamber with 48 kHz sample frequency (later downsampled to 16 kHz) and were quantized with 16 bits per sample. A perception test with 20 peoples was carried out to ensure the emotional quality and naturalness of the utterances.

Speech signals parametrization using the first 12 MFCC was conducted at first by using Hamming windows of 25 ms with a 10 ms frame shift. Then, the first 12 MFCC including the first and second derivatives were taken [15]. In order to implement GMM and HMM, the *Hidden Markov Toolkit* (HTK) [15] was used.

The transcriptions of the utterances are not considered and each utterance has one label deal with the emotion expressed. Then, each utterance is a train or a test pattern according to the case.

The estimation of recognition can be biased if only one train and test partition is used. To avoid these estimation biases, a cross-validation with the leave-k-out method was performed [16]. Ten data partitions were generated, for every one a 80% of data was randomly selected to the training and the remainder 20% was left for test.

The tests of GMM altering the Gaussian components number in the mixture, increasing in two every time, were performed. To evaluate the hidden Markov models, a two states model was defined and it was undergoing to similar previously mentioned tests, increasing the Gaussian components too. After that, one state was added to the model, the system tests were repeated, ando so on until arriving at a seven states model.

[1] The information is accessible from http://pascal.kgw.tu-berlin.de/emodb/.

Table 2. Confusion matrix for 3 emotions and GMM with 22 Gaussians.

Emotion	Joy	Anger	Neutral
Joy	**99**	38	3
Anger	50	**192**	8
Neutral	11	4	**135**

Table 3. Confusion matrix for 7 emotions and a GMM with 32 Gaussians.

Emotion	Joy	Fear	Disgust	Sadness	Anger	Boredom	Neutral
Joy	**101**	9	2	0	25	0	0
Fear	21	**62**	2	16	17	4	4
Disgust	4	12	**67**	8	3	3	0
Sadness	0	0	1	**100**	0	14	5
Anger	23	6	1	0	**220**	0	0
Boredom	0	6	15	26	0	**63**	50
Neutral	2	4	6	21	3	30	**84**

The evaluation started with three emotions (neutral, joy and anger) and the emotions were added one-by-one up to reach the seven emotions.

4 Results and Discussions

Although experiments were carried out for all the combinations of states and Gaussian components of the states in the HMM analysis, for brevity only the best results are shown here.

The confusion matrix is a good representation to analyse the performance and to find the main classification errors. In the confusion matrices shown in Tables 2, 3, 4 and 5, the columns have the emotions uttered by the speakers, and the rows are the outputs of the recognizer. The main diagonal shows the right recognized emotions and the other values are the sustitution errors between emotions.

Table 2 shows the confusion matrix for the recognition of the three most standard emotions with GMM of 22 components. An accuracy of 79% was achieved for this test, with the most important confusion between *Joy* and *Anger* emotions. In Table 3, a confusion matrix for the GMM with 32 Gaussian components and seven emotions is shown. There, the percentage of correctly identified emotions was 67%.

Figure 1 shows an analysis of the number of states in HMM, taking the average recognition rates for $K \in [1, 2, 4, \ldots, 32]$. It is possible to observe that the is no need to increase the number of states beyond two states.

The recognition performance of the HMM was also investigated in relation to the number of Gaussian components. Figure 2 shows how the number of Gaussians affects the performance in the 2 states HMM. The range from 14 to 22 Gaussian components provided the best performance.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

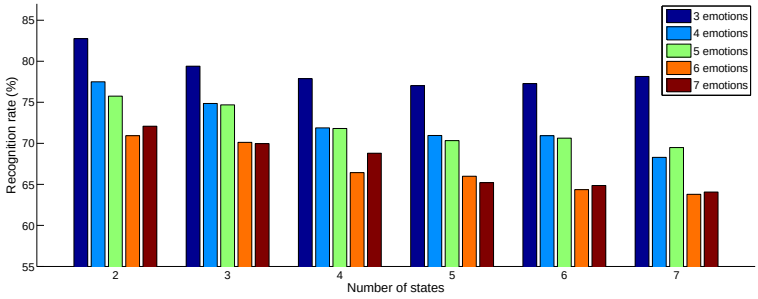

Fig. 1. Recognition as a function of number of states in HMM.

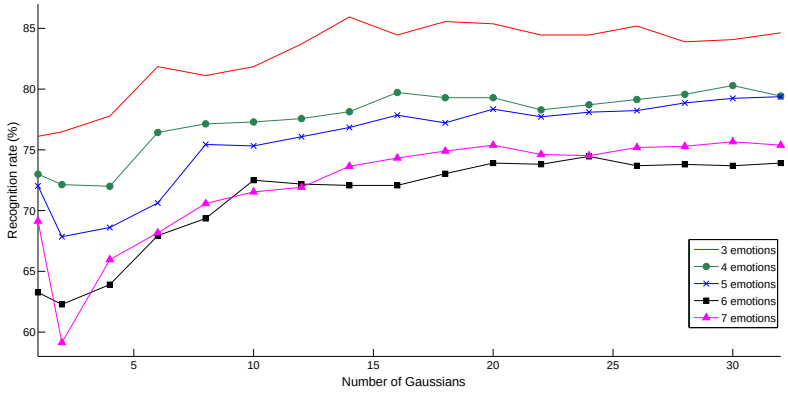

Fig. 2. Recognition as a function of number of Gaussians for a 2 state HMM.

Table 4. Confusion matrix for 3 emotions and a 2 states HMM (14 Gaussians).

Emotion	Joy	Anger	Neutral
Joy	**101**	39	0
Anger	32	**216**	2
Neutral	2	1	**147**

Tables 4 and 5 show the confusion matrices for three and seven emotions recognized with a two states HMM. The percentages of correctly identified emotions, for three and seven emotions tests, were 86% and 76% respectively. These results confirm both the usefulness of HMM and the convenience of MFCC parameterization.

The first remark about the results is the fact that in all the cases, a corpus uttered by 10 speakers was used. Given the multiple speakers in the speech corpus and the cross-validation used to evaluate the performance, it may be arguable that results can be generalised to other speakers. This is an important

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

Table 5. Confusion matrix for 7 emotions and a 2 states HMM (30 Gaussians).

Emotion	Joy	Fear	Disgust	Sadness	Anger	Boredom	Neutral
Joy	**93**	13	0	0	34	0	0
Fear	13	**93**	5	6	7	3	3
Disgust	4	7	**70**	0	4	3	2
Sadness	0	0	3	**93**	0	23	1
Anger	12	6	1	0	**231**	0	0
Boredom	0	3	7	14	0	**94**	42
Neutral	0	5	5	0	0	27	**113**

improvement in comparison with previous (c.f. [1, 3, 7]). For example, in [3] an accuracy of 95% is reported for a single professional speaker corpus. Then, if the here reported improvement from GMM to HMM is taken into account, a similar improvement over this single speaker corpus can be expected. However, a recognizer trained with only one speaker is not suitable for practical purposes.

Here, the presented results were selected in order to compare these with others works, despite an interesting relation between others evaluated emotions was found. Although *Anger*, *Happiness* and *Neutral* are generally used as extreme emotion patterns [17], it was observed that a similar test done with *Anger*, *Sadness* and *Neutral* accomplished an accuracy of 97% with HMM and 93% with GMM. The traditional definition of primary and secondary emotions is founded in phycological analysis and it is not related with the classification difficulties of themselves. Then, the automatic recognition results here presented could be important in order to re-defining the primary and secondary emotions for each language.

5 Conclusions and Future Works

In this paper two approaches to emotion recognition were studied, GMM and HMM, and up to 7 emotions styles were recognized. Tests with different number of components in GMM and with different number of states and Gaussian components in HMM were performed.

The high performance of MFCC has been showed and it could be considered a very useful key in emotion recognition. The vocal tract morphology changes because of different emotional states and these changes are properly captured in the cepstral features.

Results can be generalized to other speakers because the multi-speaker corpus and the cross-validation method used in the experiments. It is an important point for comparison with other previous works on speaker-dependent emotion recognition.

Gaussian mixtures had an acceptable performance, but its performance degrades as emotions are added. However, HMM obtain better performances because they allow more complex models. HMM provided better performance than GMM for all the tests.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.

The prosody and the spectral features play an important role in the emotion recognition task. Therefore, in future works will be important to study the combination of these features to increase the performance in emotion recognition systems.

Also, it is planned to carry out similar analyses on other languages. For the Spanish case we are working in the development of a corpus with speech signals from hispanic speech actors. Then, a subjective evaluation of the emotional speech corpus by humans could be carried out with this corpus. Thus, the ability of listeners to correctly classify the emotional utterances could be compared with the results of the automatic recognizer.

References

1. Lin, Y.L., Wei, G.: Speech emotion recognition based on HMM and SVM. Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on **8** (August 2005) 4898–4901
2. Dellaert, F., Polzin, T., Waibel, A.: Recognizing emotions in speech. In: Proc. ICSLP '96. Volume 3., Philadelphia, PA (1996) 1970–1973
3. Gil, L., et al.: Reconocimiento automático de emociones utilizando parámetros prosódicos. Procesamiento del lenguaje natural (35) (September 2005) 13–20
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2 sub edn. Wiley-Interscience (oct 2000)
5. Bishop, C.M.: Pattern Recognition and Machine Learning. 1 edn. Springer (2006)
6. Nwe, T.L., Foo, S.W., Silva, L.C.D.: Speech emotion recognition using hidden markov models. Speech Communication **41** (November 2003) 603–623
7. Noguerias, A., Moreno, A., Bonafonte, A., no, J.M.: Speech Emotion Recognition Using Hidden Markov Models. Eurospeech 2001 (2001) 2679–2682
8. Deller, J.R., Proakis, J.G., Hansen, J.H.: Discrete-Time Processing of Speech Signals. Macmillan Publishing, New York (1993)
9. Kuc, R.: Introduction to digital signal processing. McGraw-Hill Book Company (1988)
10. Rabiner, L.R., Juang, B.H.: Fundamentals of Speech Recognition. Prentice-Hall (1993)
11. Oppenheim, A.V., Wilsky, A.S.: Señales y Sistemas. Prentice Hall (1998)
12. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B.: A Database of German Emotional Speech. Proc. Interspeech 2005 (September 2005) 1517–1520
13. Paeschke, A.: Global Trend of Fundamental Frequency in Emotional Speech. In: ISCA - Speech Prosody, Nara, Japan (March 2004) 671–674
14. Burkhardt, F., Sendlmeier, W.F.: Verification of Acoustical Correlates of Emotional Speech Using Formant-Synthesis. In: ISCA - Speech Prosody, Newcastle, Northern Ireland, UK (September 2000) 151–156
15. Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Valtchev, V., Woodland, P.: The HTK Book (for HTK Version 3.1). Cambridge University Engineering Department., Cambridge, Inglaterra. (Dic. 2001)
16. Michie, D., Spiegelhalter, D., Taylor, C.: Machine Learning, Neural and Statistical Classification. Ellis Horwood, University College, London (1994)
17. Cowie, R., Cornelius, R.: Describing the emotional states that are expressed in speech. Speech Communication **40**(1) (2003) 5–32

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
E. M. Albornoz, M. B. Crolla & D. H. Milone; "Recognition of emotions in speech"
XXXIV Conferencia Latinoamericana de Informática, pp. 1120-1129, Sep, 2008.