# SYNTHETICAL ENLARGEMENT OF MFCC BASED TRAINING SETS FOR EMOTION RECOGNITION

Inma Mohino-Herranz[1], Roberto Gil-Pita[1], Sagrario Alonso-Diaz[2] and Manuel Rosa-Zurera[1]

[1]Department of Signal Theory and Communications, University of Alcala, Spain
inmaculada.mohino@edu.uah.es, roberto.gil@uah.es, manuel.rosa@uah.es
[2]Human Factors Unit, Technological Institute "La Marañosa" –MoD, Madrid (Spain)
salodia@et.mde.es

***ABSTRACT***

*Emotional state recognition through speech is being a very interesting research topic nowadays. Using subliminal information of speech, it is possible to recognize the emotional state of the person. One of the main problems in the design of automatic emotion recognition systems is the small number of available patterns. This fact makes the learning process more difficult, due to the generalization problems that arise under these conditions.*

*In this work we propose a solution to this problem consisting in enlarging the training set through the creation the new virtual patterns. In the case of emotional speech, most of the emotional information is included in speed and pitch variations. So, a change in the average pitch that does not modify neither the speed nor the pitch variations does not affect the expressed emotion. Thus, we use this prior information in order to create new patterns applying a pitch shift modification in the feature extraction process of the classification system. For this purpose, we propose a frequency scaling modification of the Mel Frequency Cepstral Coefficients, used to classify the emotion. This proposed process allows us to synthetically increase the number of available patterns in thetraining set, thus increasing the generalization capability of the system and reducing the test error.*

***KEYWORDS***

*Enlarged training set, MFCC, emotion recognition, pitch analysis*

## 1. INTRODUCTION

Emotional state recognition (ESR) through speech is being a very interesting research topic nowadays. Using subliminal information of speech, it is possible to recognize the emotional state of the person. This information, denominated "prosody", reflects some features of the speaker and adds information to the communication [1], [2].

David C. Wyld et al. (Eds) : CST, ITCS, JSE, SIP, ARIA, DMS - 2014
pp. 249–259, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4123

The standard scheme of an ESR system consists of a feature extraction stage followed by a classification stage. Some of the most useful features used in speech-based ESR systems are the Mel-Frequency Cepstral Coefficients (MFCCs), which are one of the most powerful features used in speech information retrieval [3]. The classification stage uses artificial intelligence techniques to learn from data in order to determine the classification rule. It is important to highlight that in order to avoid loss of generalization of the results, it is also necessary to split the available data in two sets, one for training the system and other for testing it, since the data must be different in order to avoid loss of generalization of the results.

One of the main problems in the design of automatic ESR systems is the small number of available patterns. This fact makes the learning process more difficult, due to the generalization problems that arise under these conditions [4], [5].

A possible solution to this problem consists in enlarging the training set through the creation the new virtual patterns. This idea, originally proposed in [6], consists in the use of auxiliary information, denominated hints, about the target function to guide the learning process. The use of hints have been proposed several times in several applications, like, for instance, automatic target recognition [7], or face recognition [8].

In the case of emotional speech, it is important to highlight that most of the information is included in speed and pitch variations [9]. So, a change in the average pitch value that does not modify neither the speed nor the pitch variations does not affect the expressed emotion.

In this work we propose the creation of new patterns by applying a pitch shift modification in the feature extraction process of an ESR system. For this purpose, we propose a frequency scaling modification of the MFCCs. This proposed process allows us to synthetically increase the number of available patterns in the training set, thus increasing the generalization capability of the system and reducing the test error.

## 2. MATERIALS AND METHODS

This section explains the two main stages of an ESR system: the feature extraction stage and the classification stage, describing the configuration of the ESR system used in the experiments.

### 2.1. Feature extraction: Mel-Frequency Cepstral Coefficients (MFCCs)

Obtaining MFCC coefficients [10] has been regarded as one of the techniques of parameterization most important used in speech processing. They provide a compact representation of the spectral envelope, so that most of the energy is concentrated in the first coefficients. Perceptual analysis emulates human ear non-linear frequency response by creating a set of filters on non-linearly spaced frequency bands. Mel cepstral analysis uses the Mel scale and a cepstral smoothing in order to get the final smoothed spectrum. Figure 1 shows the scheme for the MFCC evaluation.
The main stages of MFCC analysis are:

- *Windowed*: In order to overcome the non-stationary of speech, it is necessary to analyze the signal in short time periods, in which it can be considered almost stationary. So, time frames or segments are obtained dividing the signal. This process is called windowed. In order to maintain continuity of information signal, it is common to perform the windowed sample with frame blocks overlap one another, so that the information is not lost in the transition between windows.
- *DFT*: Following the windowed, DFT is calculated to $x_t[n]$, the result of windowing the $t$-th time frame with a window of length $N$.

$$X_t[k] = \sum_{n=0}^{N-1} x_t[n] \cdot e^{-j2\pi nk}, 0 \leq k \leq N-1 \quad (1)$$

From this moment, phase is discarded and we work with the energy of speech signal, $|X_t[k]|^2$.

- *Filter bank*: The signal $|X_t[k]|^2$ is then multiplied by a triangular filter bank, using Equation (2).

$$E_{nt} = \sum_{k=0}^{N/2} |X_t[k]|^2 H_m[k], \quad 1 \leq m \leq F \quad (2)$$

where $H_m[k]$ are the triangular filter responses, whose area is unity. These triangles are spaced according to the MEL frequency scale. The bandwidth of the triangular filters is determined by the distribution of the central frequency $f[m]$, which is function of the sampling frequency and the number of filters. If the number of filters is increased, the bandwidth is reduced.

So, in order to determine the central frequencies of the filters $f[m]$, the behaviour of the human psychoacoustic system is approximated through $B(f)$, the frequency in MEL scale, in Equation (3).

$$B(f) = 2595 \cdot log(1 + f/700) \quad (3)$$

where $f$ corresponds with the frequency represented on a linear scale axis.

Therefore, the triangular filters can be expressed using Equation (4).

$$H_m[k] = \begin{cases} 0, & k < f[m-1] \\ \frac{2(k-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])}, & f[m-1] \leq k < f[m] \\ \frac{2(f[m+1]-k)}{(f[m+1]-f[m-1])(f[m]-f[m-1])}, & f[m] \leq k < f[m+1] \\ f[m+1] - f[m] & k \geq f[m+1] \end{cases} \quad (4)$$

where $1 \leq m \leq$ F , being $F$ the number of filter, and furthermore we have the central frequency $f[m]$ of the $m$-th frequency band :

$$f[m] = \frac{N}{F_s} B^{-1}\left(m \frac{B(F_s/2)}{M+1}\right) \quad (5)$$

where $B^{-1}(b) = 700(e^{b/2595} - 1)$, and $F_s$ is the frequency sampling.

- *DCT (Discrete Cosine Transform)*: Through the DCT, expressed in Equation (6), the spectral coefficients are trans- formed to the frequency domain, so the spectral coefficients are converted to cepstral coefficients.

$$MFCC_{nt} = \sum_{k=1}^{F} log(E_{nk}) cos(n(k-1/2)\pi/N), \quad n = 1, \cdots, F \quad (6)$$

One MFCCs are evaluated, features are determined from statistics of each MFCC. Some of the most common used statistics are the mean and the standard deviation. In is also habitual to use statistics from differential values of the MFCCs, denominated, delta MFCC, or ΔMFCCs. These ΔMFCCs are determined using Equation (7),

$$\Delta MFCC_{nt} = MFCC_{nt} - MFCC_{n(t-d)} \tag{7}$$

where $d$ determines the differentiation shift. In this paper we use as features the mean and standard deviation of the MFCCs, and the standard deviation of ΔMFCCs with $d = 2$, since we have found that these values obtain very good results with a considerably low number of features.


Figure 1: Scheme to MFCC calculate

## 2.2. Classification stage: Least Square Diagonal Quadratic Classifier

The Least Square Diagonal Quadratic Classifier is a classifier that renders very good results with a very fast learning process [11] and therefor it has been selected for the experiments carried out in this paper. Let us considerer a set of training patterns $x = [x_1, x_2, \ldots, x_L]^T$, where each of these patterns is assigned to one of the possible classes denoted as $C_i$, $i = 1, \ldots, k$. In a quadratic classifier, the decision rule can be obtaining using a set of $k$ combinations, as shows Equation (8)

$$y_k = w_{k0} + \sum_{n=1}^{L} w_{kn} x_n + \sum_{n=1}^{L} \sum_{m=1}^{n} x_m x_n v_{mnk} \tag{8}$$

where $w_{kn}$ and $v_{mnk}$ are the linear and quadratic values weighting respectively. Furthermore, Equation (8) can be expressed in matrix notation as shown in Equation (9).

$$\mathbf{y} = \mathbf{w}_0 + \mathbf{W}^T \mathbf{x} + \mathbf{V}^T \mathbf{x} \tag{9}$$

The particular case of a diagonal quadratic classifier is referred to the use of only the diagonal coefficients of $\boldsymbol{V}$. This leads to a simplification of Equation (8), giving Equation (10).

$$v_{mnk} = 0, \quad \forall m \neq n \tag{10}$$

With this last equation, the decision rule is obtained as shows Equation (11).

$$y_k = w_{k0} + \sum_{n=1}^{L} w_{kn} x_n + \sum_{n=1}^{L} x_n^2 v_{nnk} \tag{11}$$

The pattern matrix **Q,** which contains the input features for classification and his quadratic value, is expressed in Equation (12).

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L1} & x_{L2} & x_{L3} & \dots & x_{LN} \\ x_{11}^2 & x_{12}^2 & x_{13}^2 & \dots & x_{1N}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L1}^2 & x_{L2}^2 & x_{L3}^2 & \dots & x_{LN}^2 \end{bmatrix} \tag{12}$$

Being, **V** as is expressed the Equation (13)

$$\mathbf{V} = \begin{bmatrix} w_{10} & w_{11} & \cdots & w_{1L} & v_{111} & \cdots & v_{1LL} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{k0} & w_{k1} & \cdots & w_{kL} & v_{k11} & \cdots & v_{kLL} \end{bmatrix} \tag{13}$$

So, the output of the quadratic classifier is obtained according to Expression (14).

$$\mathbf{Y} = \mathbf{V} \cdot \mathbf{Q} \tag{14}$$

Let us now define the target matrix containing the labels of each pattern as:

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{K1} & t_{K2} & t_{K3} & \dots & t_{KN} \end{bmatrix} \tag{15}$$

where $N$ is the number of data samples, and $t_{kn}$=1 if the $n$-th pattern belongs to class $C_K$, and 0 in other case. Then, the error is the difference between the outputs of the classifier and the correct values, which are contained in the target vector:

$$\mathbf{E} = \mathbf{Y} - \mathbf{T} = \mathbf{V} \cdot \mathbf{Q} - \mathbf{T} \tag{16}$$

Consequently, the mean square error (*MSE*) is computed according to Equation (17).

$$MSE = \frac{1}{N} \|\mathbf{Y} - \mathbf{T})\|^2 = \frac{1}{N} \|\mathbf{V} \cdot \mathbf{Q} - \mathbf{T}\|^2 \tag{17}$$

In the least squares approach, the weights are adjusted in order to minimize the mean square error. The minimization of the *MSE* is obtained deriving expression (17) with respect **V** and, using the equations of *Wiener-Hopf*, the next expression for the weight values is obtained:

$$\mathbf{V} = \mathbf{T} \cdot \mathbf{Q}^T \cdot \left(\mathbf{Q} \cdot \mathbf{Q}^T\right)^{-1} \tag{18}$$

This expression allows to determine the values of the coefficients that minimize the mean square error for a given set of features.

## 3. PROPOSED MFCC-BASED ENLARGEMENT OF THE TRAINING SET

As we stated in the introduction, it is important to highlight that most of the information of emotional speech is included in speed and the pitch variations [9]. So, an average change in the pitch value that does not modify neither the speed nor the pitch variations does not affect the expressed emotion.

In this paper we propose to modify the MFCC extraction in order to implement frequency scaling, allowing to create new patterns for the training set. So, the MFCCs can be easily pitch-shifted through a scale factor applied in frequency domain. This modification is applied to each pattern in the database, allowing to enlarge the training set.

Let us define the Pitch Shift Factor ($P_{SF}$) as a global change of the pitch, measured in semitones. Then, this shift in the pitch is equivalent to scaling the frequency with a Frequency Scale Factor ($F_{SF}$). So, the relationship between $P_{SF}$ and $F_{SF}$ can be expressed using Equation (19).

$$F_{SF} = 2^{\frac{P_{SF}}{12}} \tag{19}$$

In order to apply this frequency scaling in the MFCC process, the central frequencies $f[m]$ of the triangular filters are modified, taking into account the scaled frequency factor. So, in Equation (20) we can observe the relationship between the original and synthetic frequency.

$$f'[m] = F_{SF} \cdot f[m] \tag{20}$$

Being the new frequency scale, as shows in Equation (21)

$$f'[m] = F_{SF} \cdot \frac{N}{F_s} B^{-1}\left(m \frac{B(F_s/2)}{M+1}\right) \tag{21}$$

Figure 2 shows in linear axis, that is, the frequency is scaled the central frequency for each MFCC. In this Figure, we can observe the difference between the Normal relationships between center frequency for each coefficient, and the difference when the frequency has been scaled, that is, the center frequency is reduced when increase the number of cepstral coefficients.

As an example, the difference between the MFCCs calculate with $P_{SF}=0$ and $P_{SF}=1$ are shown in Figure 3. So, we can observe that the filter responses in logarithmic scale without frequency shift of MFCCs with a shifting in frequency of one semitone.

In order to implement the enlargement of the database using pitch shifting, two factors must be taken into account: the range of the pitch shifting ($R$) and the step of the pitch shifting ($S$).

- *Range (R):* The range defines the maximum absolute variation in the pitch modification process in semitones. With this parameter it is possible to change the upper and lower limits of the shift variations.

- *Step (S):* The step defines the smallest change in the pitch that is produced in the pitch shifting process in semitones.

Taking into account these two factors, it is possible to determine the enlargement factor ($EF$), that is, the number of times that the size of the training set is increased.

$$EF = \frac{2R}{S} + 1 \tag{22}$$

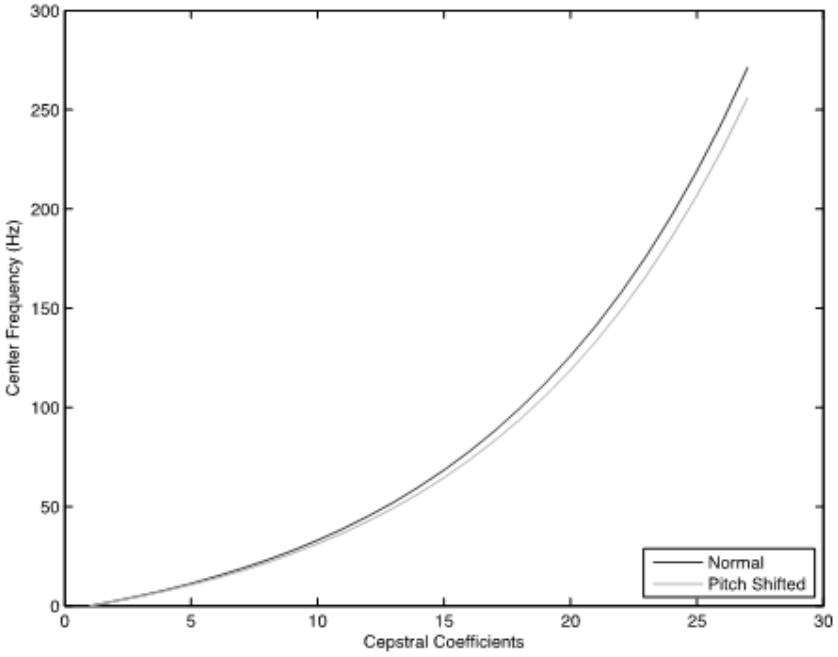


Figure 2: MFCCs for different factors. Lineal Scale.

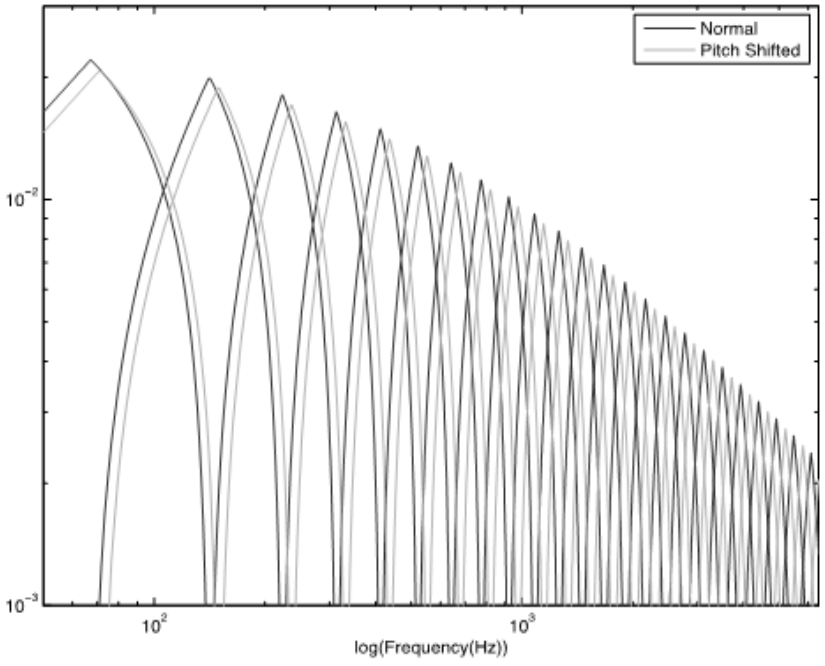


Figure 3: MFCCs for different factors. Logarithmic Scale.

# 4. RESULTS

## 4.1. Experimental setup

In this study, we have used the public database "The Berlin Database of Emotional Speech" [12]. This database consists of 800 files of 10 actors (5 males and 5 females), where each actor produces 10 German utterances (5 short and 5 longer sentences) simulating seven different emotions. These emotions are: Neutral, Anger, Fear, Happiness, Sadness, Disgust, and Boredom. The recordings were using a sampling frequency of 48 kHz and later downsampled to 16 kHz.
Although this database consists of 800 files, almost 300 were eliminated, since only those utterances with a recognition rate better than 80% and naturalness better than 60% were finally chosen. So, the database consists of 535 files.

In order to evaluate the results and to ensure that they are independent of the partition between training set and test set, we have used the validation method denominated Leave One Out [13] [14]. This is a model validation technique to evaluate how the results of a statistical analysis generalize to an independent data set. This method is used in environments where the main goal is the prediction and we want to estimate how accurate is a model that will be implemented in practice.

This technique basically consists in tree stages:

- First, the database is divided into complementary subsets called: training set and test set, where the test sets contains only one pattern in the database.
- Then, the parameters of the classification system are obtained using the training set.
- Finally, the performance of the classification system is obtained using the test set.

In order to increase the accuracy of the error estimation while maximizing the size of the training set, multiple iteration of this process are performed using a different partitions each time, and the test results are averaged over the different iterations.

In this paper we use an adaptation of this technique to the problem at hand, which we denominate *Leave One Couple Out*. So, we have worked with the database discussed above, which consists of 5 male and 5 female. In this case, we used 4 male and 4 female for each training set and 1 male and 1 female for each test set. This division guarantees complete independence between training and test data, keeping a balance in the gender. Therefore, our leave one couple out, is repeated 25 times, using each iteration different training and test sets. Concerning the features, a window size of N = 512 has been used, which implies time frames of 32ms. We have then selected mean and standard deviation of 25 MFCCs, and standard deviation of 2-ΔMFCCS, resulting in a total of 75 features, which has been used to design a quadratic classifier. In order to complete the comprehension of the results obtained, it is necessary to analyze the error probability for training set, the error probability for test set and the enlargement factor (EF). Table 1 shows the error probability for the training set. And in Table 3 it is possible to observe that the enlargement factor is 1 for the lowest error probability, which implies that to obtain the lowest error probability for the training set, the new patterns are not needed.

However, in the Table 2, we observe the minimum error probability for the test set is 27.36% with $S$ = 1/8 and $R$ = 4. Comparing these results with the one associated the Table 3, the enlargement factor in this case is 65. This implies that an $EF$ = 65 is required to achieve the lowest test error rate. This error probability is much smaller than the one is obtain for Factor equal to 1.

Table 1: Error probability for the training set

| Error Probability | | Range | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 0.5 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 |
| Step | 1/16 | 1,74% | 2,67% | 4,11% | 7,21% | 9,62% | 13,23% | 14,80% | 16,50% | 18,61% | 20,34% | 22,12% |
| | 1/8 | 1,74% | 2,76% | 4,21% | 7,29% | 9,69% | 13,31% | 14,88% | 16,58% | 18,70% | 20,38% | 22,18% |
| | 1/4 | 1,74% | 2,93% | 4,37% | 7,45% | 9,81% | 13,39% | 14,86% | 16,57% | 18,69% | 20,33% | 22,17% |
| | 1/2 | 1,74% | 3,22% | 4,71% | 7,72% | 10,03% | 13,53% | 15,00% | 16,70% | 18,87% | 20,50% | 22,25% |
| | 1 | 1,74% | 1,74% | 5,07% | 8,14% | 10,34% | 13,77% | 15,13% | 16,73% | 18,59% | 20,32% | 22,07% |
| | 2 | 1,74% | 1,74% | 1,74% | 8,37% | 8,37% | 13,36% | 13,36% | 16,21% | 18,02% | 19,66% | 21,43% |

Table 2: Error probability for the test set

| Error Probability | | Range | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 0.5 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 |
| Step | 1/16 | 33,94% | 31,77% | 29,64% | 27,73% | 27,92% | 27,47% | 28,03% | 28,93% | 30,91% | 32,71% | 33,64% |
| | 1/8 | 33,94% | 31,77% | 29,45% | 27,77% | 27,88% | **27,36%** | 28,07% | 28,71% | 30,84% | 32,85% | 33,71% |
| | 1/4 | 33,94% | 31,73% | 28,97% | 27,73% | 27,88% | 27,62% | 28,18% | 29,12% | 31,40% | 32,97% | 33,71% |
| | 1/2 | 33,94% | 31,51% | 29,34% | 27,88% | 27,88% | 27,55% | 28,29% | 29,30% | 31,36% | 33,00% | 33,68% |
| | 1 | 33,94% | 33,94% | 29,08% | 27,92% | 27,66% | 28,18% | 28,82% | 29,60% | 31,25% | 33,08% | 33,79% |
| | 2 | 33,94% | 33,94% | 33,94% | 28,33% | 28,33% | 29,27% | 29,27% | 29,68% | 31,70% | 33,53% | 33,60% |

Table 3: Enlargement factor

| Enlargement factor | | Range | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 0.5 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 |
| Step | 1/16 | 1 | 17 | 33 | 65 | 97 | 129 | 161 | 193 | 257 | 321 | 385 |
| | 1/8 | 1 | 9 | 17 | 33 | 49 | **65** | 81 | 97 | 129 | 161 | 193 |
| | 1/4 | 1 | 5 | 9 | 17 | 25 | 33 | 41 | 49 | 65 | 81 | 97 |
| | 1/2 | 1 | 3 | 5 | 9 | 13 | 17 | 21 | 25 | 33 | 41 | 49 |
| | 1 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 17 | 21 | 25 |
| | 2 | 1 | 1 | 1 | 3 | 3 | 5 | 5 | 7 | 9 | 11 | 13 |

## 5. CONCLUSIONS

One of the main problems in the design of ESR systems is the small number of available patterns. This fact makes the learning process more difficult, due to the generalization problems in the learning stage. In this work we propose a solution to this problem consisting in enlarging the training set through the creation the new virtual patterns. In the case of emotional speech, most of the emotional information is included in speed and pitch variations. Thus, a change in the average pitch value that does not modify neither the speed nor the pitch variations does not affect the expressed emotion. So, we use this prior information in order to create new patterns applying a pitch shift modification in the feature extraction process of the classification system. For this purpose, we propose a frequency scaling modification of the Mel Frequency Cepstral Coefficients. This proposed process allows us to synthetically increase the number of available patterns in the training set, thus increasing the generalization capability of the system and reducing the test error.

Using MFCC-based enlargement of the training set, the system has a number of patterns appropriate, and it is possible train to the system correctly. In this case, it is possible reduce the error probability in emotion recognition near 7%, which is a considerable improvement in the performance. This percentage value is very important in emotion recognition.

## ACKNOWLEDGEMENTS

This work has been funded by the Spanish Ministry of Education and Science (TEC2012-38142-C04-02), by the Spanish Ministry of Defense (DN8644-ATREC) and by the University of Alcala under project UAH2011/EXP-028.

## REFERENCES

[1] Verderis, D. & Kotropoulos, C., (2006) "Emotional speech recognition: Resources, features, and method", Elsevier Speech communication, Vol. 48, No. 9, pp1162-1181.
[2] Schuller, B., Batliner, A., Steidl, S. & Seppi, D., (2011) "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge", Elsevier Speech Communication, Vol. 53, No. 9, pp1062-1087
[3] Mohino, I. & Goñi, M. & Alvarez, L. & Llerena, C. & Gil-Pita, R., (2013) "Detection of emotions and stress through speech analysis", International Association of Science and Technology for Development.
[4] Öztürk, N., (2003) "Use of genetic algorithm to design optimal neural network structure", MCB UP Ltd Engineering Computations, Vol. 20, No.8, pp979-997.
[5] Mori, R., Suzuki, S. & Takahara, H., (2007) "Optimization of Neural Network Modeling for Human Landing Control Analysis", AIAA Infotech@ Aerospace 2007 Conference and Exhibit, pp7-10.
[6] Abu-Mostafa, Yaser S, (1995) "Hints", MIT Press Neural Computation, Vol. 7, No. 4, pp639-671.
[7] Gil-Pita, R & Jarabo-Amores, P & Rosa-Zurera, M & Lopez-Ferreras, F, (2002) "Improving neural classifiers for ATR using a kernel method for generating synthetic training sets", IEEE Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on, pp425-434.
[8] Niyogi, Partha & Girosi, Federico & Poggio, Tomaso, (1998) "Incorporating prior information in machine learning by creating virtual examples", IEEE Proceedings of the IEEE, Vol. 86, No. 11, pp2196-2209.
[9] Vroomen, J., Collier, R. & Mozziconacci, Sylvie JL, (1993) "Duration and intonation in emotional speech", Eurospeech.
[10] Davis, S. & Mermelstein P., (1980) "Experiments in syllable-based recognition of continuous speech", IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 28, pp357-366.
[11] Gil-Pita, R. & Alvarez-Perez, L. & Mohino, Inma, (2012) "Evolutionary diagonal quadratic discriminant for speech separation in binaural hearing aids", Advances in Computer science, Vol. 20, No. 5, pp227-232.
[12] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F. & Weiss, B., (2005) "A database of German emotional speech", Interspeech, pp15717-1520.
[13] Chang, M.-W. and Lin, C.-J., (2005) "Leave-one-out bounds for support vector regression model selection", MIT Press Neural Computation, Vol. 17, No. 5, pp1188-1222.
[14] Cawley, G. C. & Talbot, N. L., (2004) "Fast exact leave-one-out cross-validation of sparse least-squares support vector machines", Vol. 16, No. 10, pp1467-1475

## Authors

**Inma Mohino-Herranz**

Year in which an academic degree was awarded: Telecommunication Engineer, Alcalá University, 2010. PhD student about Information and Communication Technologies. Area of research: Signal Processing.

**Roberto Gil-Pita**

Year in which an academic degree was awarded: Telecommnunication Engineer, Alcalá University, 2001. Possition: Associate Professor. Polytechnic School in the Department of Signal Theory and Communications. Some of his research interest include, audio, speech, image, biological signals.

**Sagrario Alonso-Díaz**

PhD Psicologist. Researcher in the Human Factors Unit. Technological Institue “La Marañosa” –MoD.

**Manuel Rosa-Zurera,**

Year in which an academic degree was awarded: Telecomunication Engineer, Polythecnic University of Madrid, 1995. Possition: Full professor and dean of Polytechnic School. University of Alcalá. His areas of interest are audio, radar, speech source separation.