Analysis of Emotions in Vowels: a Recurrence Approach

Angela Lombardi, Pietro Guccione

Dipartimento di Ingegneria Elettrica e dell'Informazione
Politecnico di Bari
Email: angela.lombardi@poliba.it, pietro.guccione@poliba.it

***Abstract*—Emotional content in speech has been so far characterized by features based on linear source-filter models. However, the presence of nonlinear and chaotic phenomena in speech generation have been widely proven in literature. In this work, a novel framework has been developed to explore recurrence properties of vowels and describe nonlinear dynamics of speech. Experiments using a database of short spoken sentences emitted in the six basic emotions (anger, boredom, fear, happiness, neutral, sadness) show preliminary results of the approach.**

Keywords–Speech Emotion Recognition; Recurrence Plot; Recurrence Quantitative Analysis.

I. INTRODUCTION

Speech Emotion Recognition (SER) is a recent field of research that aims at identifying the emotional state of a speaker through a collection of machine learning and pattern recognition techniques [1].

As a classification problem, a SER system needs a set of features able to optimally reflect the emotional content in speech. According to the existing literature, it is possible to distinguish three main categories of features: prosodic, spectral, and quality-based [2]. Prosodic features such as the fundamental frequency (pitch), the energy of the signal and the rhythm/articulation rate, have been combined with spectral measures (Mel Frequency Cepstral Coefficients (MFCC), Linear Predictor Cepstral Coefficients (LPCC) and formants) in different ways to improve the performances of the classifier [3]. The third category includes acoustic cues related to the shape of glottal pulse signal, its amplitude variation (shimmer) and frequency variation (jitter) [4].

Although such features have been extensively used for the development of SER systems, they are based on a source-filter model [5], which represents a simplification of the process of voice production that ignores more complex physiological mechanisms. In fact, in the last two decades, nonlinear tools for speech signal processing have spread out after new findings concerning the occurrence of nonlinear phenomena during voice production [6]. In particular, the evidence of the chaotic behavior of certain processes involved in the speech generation (e.g., turbulent airflow) [7], made the *Chaos Theory* a favored approach for the study of nonlinear dynamics in the system voice.

To describe these dynamics it is necessary to find the set of the possible states that the system can take (to reconstruct the phase space). This approach assumes that the speech signal represents a projection of a higher-dimensional nonlinear dynamical system evolving in time, with unknown characteristics. Embedding techniques can be employed to reconstruct the attractor of the system in the phase space and provide a representation of its trajectories. Afterward, it is possible to describe the dynamic behavior of the system by studying the properties of the embedded attractor: chaotic measures such as Lyapunov exponents, correlation dimension and entropy, have been successfully applied to the analysis of vocal pathologies and speech nonlinearities [8] [9].

The behavior of the trajectories of a system in the phase space can also be modeled through the recurrence, a property that quantifies the tendency of a system to return to a state close to the initial one. A Recurrence Plot (RP) is a graphic tool that shows the recurrent behavior of the trajectories of a system even with high-dimensional phase space [10]. Recurrence Quantitative Analysis (RQA) has been introduced later to objectively evaluate the structures contained in a RP through nonlinear measurements. RQA has found extensive applications in many scientific fields, thanks to its effectiveness in the presence of short and non-stationary data [11] [12].

In this work, we have developed a framework to explore the recurrence properties of vowel segments taken from a set of spoken sentences of a publicly available database, for six categories of basic emotions (anger, boredom, fear, happiness, neutral, sadness). An automatic vowel extraction module has been built up to extract vowel segments from each sentence; then, their time evolutions have been analyzed by means of the RQA measures. To test the ability of these measures to characterize the different emotional contents, they have been grouped according to the emotion which they belong to and statistical tests have been performed to compare the six groups.

The rest of the paper is divided in four sections: the theoretical background is provided in Section II, the general framework of the approach is explained in Section III, results and conclusions are reported in Section IV and Section V, respectively.

II. THEORETICAL BACKGROUND

This section provides a general overview of the basic concepts related to the state space reconstruction of a dynamical system and of the main tools used for the analysis of its recurrence properties.

A. The Embedding Theorem

The state of a dynamical system is determined by the values of the variables that describe it at a given time.

However, in a real scenario, not all the variables of the system can be inferred and often only a time series $\{u_i\}_{i=1}^{N}$ is available as an output of the system.

Takens demonstrated that it is possible to use time delayed versions of the signal at the output of the system to reconstruct

 ISBN: 978-1-61208-487-9

a phase space topologically equivalent to the original one. According to Takens' embedding theorem [13], a state in the reconstructed phase space is given by a m-dimensional time delay embedded vector:

$$\vec{x}_i = \left(u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}\right) \tag{1}$$

where m is the embedding dimension and τ is the time delay. For the embedded parameters estimation, several techniques have been proposed. As an example, the First Local Minimum of Average Mutual Information algorithm [14] can be used to determine the time delay, while the False Nearest-Neighbors algorithm [15] is usually employed to estimate the minimum embedding dimension.

B. Recurrence Plots

A Recurrence Plot is a graphical tool that provides a representation of recurrent states of a dynamical system through a square matrix:

$$R_{i,j}(\epsilon) = \Theta\left(\epsilon - ||\overrightarrow{x_i} - \overrightarrow{x_j}||\right), \qquad i,j = 1,...,N \tag{2}$$

with $\overrightarrow{x_i}, \overrightarrow{x_j}$ the system state at times i, j, Θ the Heaviside function, ϵ a threshold for closeness, N the number of considered states and $||\cdot||$ a norm function.

An entry of the matrix is set equal to one if the distance between the corresponding pair of neighboring states is below the threshold ϵ and zero elsewhere.

The resulting plot is symmetric and always exhibits the main diagonal, called line of identity (LOI). Apart for the general RP structure, it is often possible to distinguish small scale structures, which show local (temporal) relationships of the segments of the system trajectory (for a visual reference, see Figure 3). In details:

- single isolated points are related to rare states;
- diagonal lines parallel to the LOI indicate that the evolution of states is similar at different times;
- vertical lines mark time intervals in which states do not change.

C. Recurrence Quantitative Analysis

Several measures of complexity (RQA) have been proposed to obtain an objective quantification of the patterns in a Recurrence Plot [11] [12].

RQA can be divided into three major classes:

1) Measures based on recurrence density. Among these, the simplest measure is the *recurrence rate* (RR) defined as:

$$RR(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}(\epsilon) \tag{3}$$

It is a measure of the density of the recurrence points in the RP.

2) Measures based on the distribution $P(l)$ of lengths l of the diagonal lines. Among these, the *determinism* (DET) is the ratio of the recurrence points that form diagonal structures to all recurrence points and it is an index of the predictability of a system:

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{4}$$

The $RATIO$, defined as the ratio between DET and RR, combines the advantages of the these two categories of measures: it has been proven that it is able to detect some types of transitions in particular dynamics.

3) Measures based on the distribution $P(v)$ of vertical line lengths v. This distribution is used to quantify laminar phases during which the states of a system change very slowly or do not change at all. The ratio of recurrence points forming vertical structures to all recurrence points of the RP is called *laminarity* (LAM):

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{l=1}^{N} vP(v)} \tag{5}$$

From a Recurrence Plot, it is possible to extrapolate the *recurrence times*. The *recurrence times of second type* are:

$$\left\{T_k^{(2)} = j'_{k+1} - j'_k\right\}_{k\in\mathbb{N}} \tag{6}$$

The set of $T^{(2)}$ measure the time distance between the beginning of subsequent recurrence structures in the RP along the vertical direction and they can be considered as an estimate of the average of the lengths of white vertical lines in a column of the plot [12].

A great advantage offered by this analysis is that the calculation of the RQA measures for moving windows along the RP allows to identify the transitions of dynamical systems.

III. General Framework

The algorithm block scheme is represented in Figure 1. Since the voice has a non-stationary nature, we perform a short term analysis with a frame size of 40 ms and an overlap of 50%. Given an input track, an automatic vowel extraction module is used to detect and retain only the vowel frames and for each of them the optimal parameters (m and τ) for state space reconstruction are found. Then, RPs are generated using the time delay method, and some RQA measures extracted to describe RPs quantitatively. Since a set of RQA measures can be extracted, in principle, for each frame, statistics on these measures may be collected to give a general description of the emotional content of the input sentence.

Each step of the adopted framework is detailed in the following sections.

A. Database

The German Berlin Emotional Speech Database (EmoDB) [16] has been employed for all the experiments carried out in this work. The database contains ten sentences pronounced by ten actors (five males and five females) in 7 different emotional states: neutral, anger, fear, happiness, sadness, disgust and boredom. The audio tracks were sampled as mono signals at 16 KHz, with 8 bit/sample. Most of the sentences were recorded several times in different versions and the resulting corpus was subjected to a perception test where the degree of recognition of emotions and their naturalness were evaluated by a group of listeners. Utterances with an emotion recognition rate better than 80% and a naturalness score greater than 60% were included in the final database. As shown in Table I, among the 535 available sentences, some emotions prevail over

 ISBN: 978-1-61208-487-9

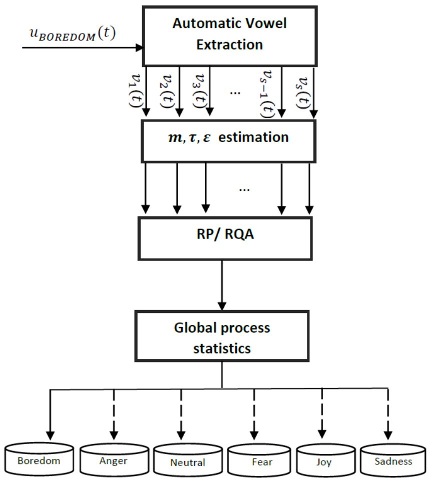


Figure 1. The algorithm block scheme for an example input sentence.

the others. The emotion disgust has been excluded from our analysis because of the too low number of tracks belonging to this group.

TABLE I. NUMBER OF UTTERANCES IN EMODB

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

B. Automatic Vowel Extraction

The vocal tract acts as a resonator filter which has its own resonance frequencies known as formants: by varying the shape of the vocal tract to produce different sounds, the formant frequencies of the filter change too [5]. Therefore, lots of characteristics of speech sounds can be detected by analyzing the spectral content of their waveforms. In detail, vowels, unlike consonants, show quasi-periodic waveforms and this can be proved by differences in the first three formant frequencies [5].

For these reasons, the estimation of the vowel segments has been carried on by extracting spectral features from the formant frequencies estimated from the power spectral density of the audio track. Then, the features have been used to train a classifier that automatically detects vowel segments in the signal.

Supposing each frame the output of a stationary process, an autoregressive model (AR) has been used to estimate the power spectral density. First, the order of the model has been identified with the Akaike's Information Criterion (AIC) [17] to avoid splitting line and spurious peaks in the final spectrum. Subsequently, the Burg's method [18] has been employed to find the parameters of the AR model. This technique has been preferred over the simple linear prediction analysis as the former identifies the optimal set of parameters by minimizing the sums of squares of the forward and backward prediction errors while the latter uses only the backward errors. Furthermore, as compared with other parametric methods, the Burg's algorithm ensures more stable models and a higher frequency resolution [19].

The peaks of the power spectral density are in correspondence of the formants position. The first three peaks have been identified in the estimated spectrum and for each of them the following characteristics have been collected:

- the frequency at which they occur;
- the amplitude of the peak;
- the area under the spectral envelope within the -3dB bandwidth.

To distinguish the vowel sounds from all other types of phonemes (including silence intervals) a one-class classification approach has been adopted. This method was introduced by Schölkopf [20] as a variant of the two-class SVM to identify a set of outliers amongst examples of the single class under consideration. Thus, according to this approach, the outlier data are examples of the negative class (in this case, the not-vowels frames). A kernel function is used to map the data into a feature space F in which the origin is the representative point of the negative class. So, the SVM returns a function f that assigns the value $+1$ in a subspace in which the most of the data points are located and the opposite value -1 elsewhere, in order to separate the examples of the class of interest from the origin of the feature space with the maximum margin.

Formally, let us consider $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l$, l training vectors of the one class X, where X is a compact subset of $\mathbb{R}^N$. Let $\Phi : X \rightarrow F$ be a a kernel function that map the training vectors into another space F. Separating the data set from the origin is equivalent to solving the following quadratic problem:

$$\min_{w \in F, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \frac{1}{2}||w||^2 \frac{1}{\nu l} \sum_{i=1}^{l} \xi_i - \rho \tag{7}$$

subject to

$$(w \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \qquad \xi_i \geq 0 \tag{8}$$

where $\nu \in (0; 1]$ is a parameter that controls the decision boundary of the classification problem, ξ_i are the nonzero slack variables, w a weight vector and ρ an offset that parametrizes a hyperplane in the feature space associated with the kernel. If w and ρ solve for this problem, then the decision function:

$$f(\mathbf{x}) = \mathrm{sign}(w \cdot \Phi(\mathbf{x}) - \rho) \tag{9}$$

will be positive for the most of the examples $\mathbf{x}_i$ contained in the training set.

Of course, the type of kernel function, the operating parameters of the kernel and the correct value of ν must be estimated to build the one-class SVM classifier. As suggested by the author, we have chosen a Gaussian kernel with Sequential Minimal Optimization (SMO) algorithm to train the classifier, since the data are always separable from the origin in the feature space. For generic patterns $\mathbf{x}$ and $\mathbf{y}$, a Gaussian kernel is expressed as:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{||\mathbf{x} - \mathbf{y}||^2}{c}\right) \tag{10}$$

where the parameter c is the kernel scale that controls the tradeoff between the over-fitting and under-fitting loss in the feature space F.

Regarding the choice of the value ν, it should be taken into account that it represents an upper bound on the fraction of outliers and, at the same time, a lower bound on the fraction of support vectors. It is then necessary to find a value that on the one hand is able to to describe the whole dataset for training and on the other hand avoids the over-training of such data. Results on the tuning of the parameters on real data and classification performances are in Section IV-A.

C. RP/RQA

In this work, the dynamics of each vowel were treated as local descriptions of the overall process of expression of a particular emotion. Therefore, after extraction of vowel segments from a sentence, a frame-level analysis is applied to monitor such dynamics. First, time delays and embedding dimensions are estimated to allow a correct reconstruction of the dynamics in the phase space. Hence the Recurrence Plots are obtained and the Recurrence Quantitative Analysis is performed on RPs. In order to explore the time dependent behavior of the recurrence measures, the computation is performed using sliding windows of length W (less than the duration of a frame) with an offset of W_s samples along the main diagonal of the RP of each vowel frame. The values of these two parameters are calculated accounting for the scale of the dynamics to be investigated (local/global) and for the temporal resolution to be achieved. The overall trend of each RQA measure is finally reconstructed considering the various vowel segments neatly placed in the sentence. For an experimental dataset of sentences, the trends of each RQA measures are grouped by emotion and some statistics are computed to explore the general characteristics of the emotions expressed in the sentences.

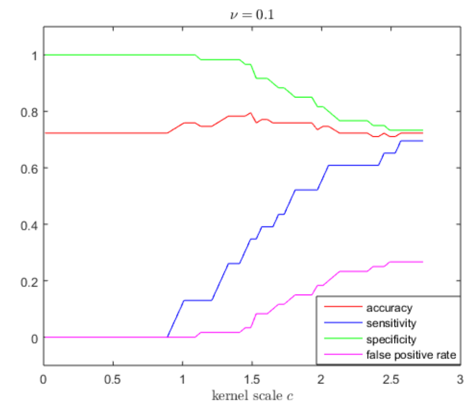
IV. RESULTS

The following sections report the performances achieved by the one-class SVM classifier and both qualitative and quantitative results of the recurrence analysis.

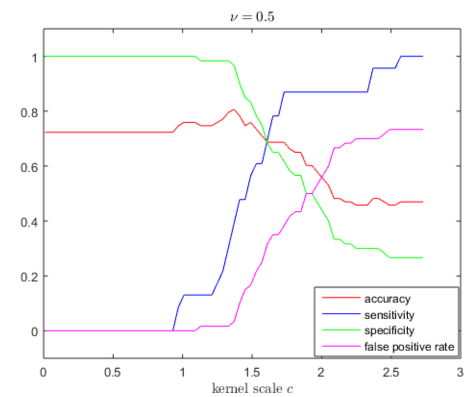
A. Automatic Vowel Extraction

To train the one class SVM classifier, a dataset was used of 128 segments of German vowels of duration equal to 40 ms, extracted from several sentences spoken by four people (two men and two women) for the six emotions. In order to identify the optimal values for the parameters c and ν, the classifier was trained and validated several times. In particular, due to the nature of the classification problem, an holdout validation scheme has been adopted. So, another set of 83 speech segments including vowels, consonants and pauses, has been used to tune the parameters and identify the most effective model. Keeping fixed the value of ν, the classifier was retrained by varying the value of the kernel scale in a predetermined range. For each model obtained, the performances on the validation set were evaluated in terms of accuracy, sensitivity (or true positive rate), specificity (or true negative rate) and false positive rate. The curves that illustrate the behavior of such measures for three values of ν and by varying the kernel scale from 0 to 2.7 are shown in Figure 2.

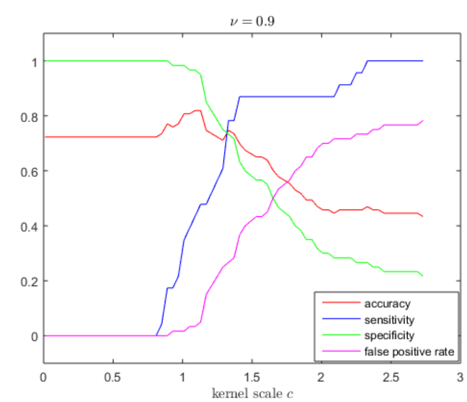
In Figure 2b and 2c only one point can be identified to guarantee high performances of the classifier, since the values of accuracy, sensitivity and specificity are high (around 0.7), while the false positive rate remains low. For kernel scale



(a)



(b)



(c)

Figure 2. Accuracy, sensitivity, specificity and false positive rate of the one class SVM classifier in function of the kernel scale c for the fixed parameter (a) $\nu = 0.1$ (b) $\nu = 0.5$ (c) $\nu = 0.9$.

values greater than this optimum, specificity and accuracy decrease rapidly, while sensitivity and false positive rate increase. These results suggest that there is a rapid growth of the number of false positives, i.e., the percentage of the not-vowels frames incorrectly predicted as vowels by the classifier increases.

For our purposes, the system critically depends on the percentage of false positives, since the classifier acts properly if it is capable of rejecting the greatest amount of not-vowel frames. Therefore, even at the expense of a lower number of true positives and higher percentage of false negatives

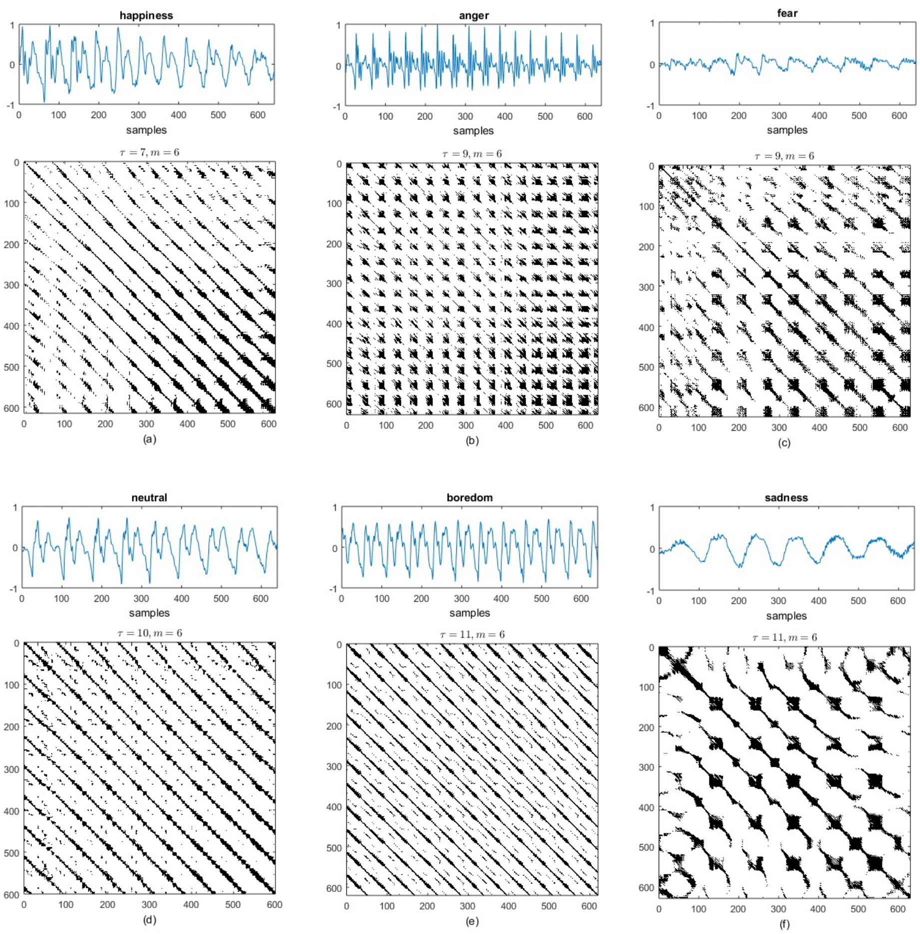


Figure 3. RPs of vowel /a/ in the track $08a02$ for emotions: (a) happiness, (b) anger, (c) fear, (d) neutral, (e) boredom, (f) sadness; ϵ is setting to 10% of maximum space diameter with a maximum norm.

(vowel frames incorrectly rejected), we have preferred to set $\nu = 0.1$ and consequently chosen the value of c at which the classifier returns high values of accuracy and specificity, while maintaining a false positive rate less than 15% (see Figure 2a).

To assess the performances of the one class SVM with the chosen parameter settings ($\nu = 0.1$ and $c = 1.75$), we performed a final test on a set of 40 speech segments independent of both the training and the validation sets. The confusion matrix is shown in Table II. As it can be seen, the low rate of false positives (not-vowels incorrectly predicted as vowel frames) confirms the validity of the model for the selected parameters (represented in Figure 2a for $\nu = 0.1$ and $c = 1.75$).

TABLE II. CONFUSION MATRIX OF THE ONE CLASS SVM ON THE TEST SET COMPOSED OF 20 VOWEL AND 20 NOT-VOWEL FRAMES.

		Predicted condition	
		Vowels	Not-vowels
True conditions	Vowels	9	11
	Not-vowels	4	16

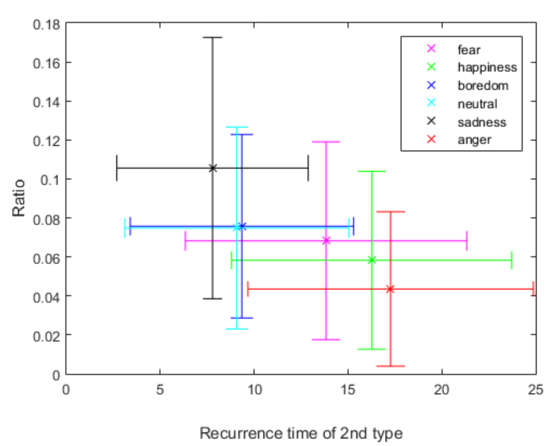


Figure 4. Median and iqr values of $RATIO$ and $T^{(2)}$ measures for the six groups of emotions.

B. Qualitative and quantitative results: RP-RQA

The patterns in RPs can reveal typical behaviors of the system and so they can be used to provide a general description of the time evolution of the dynamic trajectories. Figure 3 shows the RPs of the vowel /a/ extracted in the same sentence and approximately in the same position, pronounced by a female subject for different emotions. As it can be seen, all RPs have a topology with periodic patterns that are regularly repeated, with the exception of the emotion fear in which there are discontinuities and white bands that indicate the presence of abrupt changes in the dynamics of the system. Another distinctive feature is the length of the diagonal lines: the RPs of boredom and neutral, besides being very similar each other, have the longest diagonal lines; on the other hand, anger and fear show very short diagonal lines. Moreover, a drift can be noted in the emotion sadness: the RP fades away from LOI indicating that the system varies very slowly. The examples show that certain measures are most distinctive for some emotions and that certainly the density of points in the RPs, the length of the lines present in them and measures that are able to differentiate the different kinds of time periodicity (such as $T^{(2)}$), can effectively distinguish among different emotional levels.

On the basis of such considerations we performed the analysis described in Section III-C on a set of tracks in EmoDB (obtained by excluding from the entire set of tracks in the database, those used to train the one class SVM), to extract the collection of $RATIO$ and $T^{(2)}$ values along time. The measures were then grouped by emotions for the same RQA, obtaining 2 sets of measures (each set consists of 6 groups of data, one for each emotion). The non-parametric Kruskal-Wallis test was employed for testing whether the 6 different data groups of each RQA measure originate from the same distribution (null hypothesis), at a significance level $\alpha = 0.05$. Both tests returned a p-value < 0.0001, so the null hypothesis was rejected.

In order to better appreciate the possible differences among populations, median and interquartile range (iqr) values of the 2 RQA measures for all the groups of emotions were computed and are reported in Figure 4. It is noteworthy that boredom and neutral exhibit very similar values and that there is a relationship between the position of the emotion (based on the median values) on the 2D plot and their levels of activation (the so called arousal).

V. Conclusions

In this work, we have investigated the dynamic behavior of vowels taken from a set of spoken sentences of the EmoDB database, for the six emotions anger, boredom, fear, happiness, neutral and sadness.

To extract only the vowel frames, an automatic vowel extraction module was implemented. It consists essentially in a one class SVM classifier that processes the not-vowels frame as outliers. The tuning of the parameters of the classifier and an accurate validation step allowed us to identify a model able to achieve the 79% of accuracy.

Supposing that the expression of a particular emotional content in a spoken sentence is a gradual complex process, we exploited some properties of the local dynamics of the vowels in it to understand certain aspects of the overall process. The behavior of the trajectories of vowels dynamics was explored by means of RPs. Two kinds of RQA measures were extracted to describe RPs quantitatively. Statistical tests confirm that the considered RQA measures result statistically significant for discriminating the six groups of emotions.

In conclusion, it can be observed that certain RQA measures can better discriminate among the basic emotions examined. However, a further development could include a multivariate analysis to identify a specific subset of measures that perform a better and more complete characterization of the different emotional levels.

References

[1] C. N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," Artificial Intelligence Review, vol.43(2), 2015, pp. 155-177.

[2] M. E. Ayadi, M .S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol. 44(3), 2011, pp. 572-587.

[3] I. Luengo, E. Navas, and I. Hernáez, "Feature Analysis and Evaluation for Automatic Emotion Identification in Speech," IEEE Trans. Multimedia, vol. 12(6), 2010, pp. 490-501.

[4] M. Lugger and B. Yang, "The relevance of voice quality features in speaker independent emotion recognition," in Proc. Int. Conf. Acoustics, Speech and Signal Processing, Honolulu, HI, vol. 4, Apr. 2007, pp. 17-20.

[5] G. Fant, Acoustic Theory of Speech Production, Mouton & Co., the Hague, 1960.

[6] I. R. Titze, R. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," New Frontiers in Basic Science, I.R. Titze. Ed. Vocal Fold Physiology, San Diego, CA: Singular Publishing Group, 1993, pp. 143-188.

[7] H. Herzel, "Bifurcations and Chaos in Voice Signals," Applied Mechanics Reviews, vol. 46(7), 1993, pp. 399-413.

[8] P. Henriquez, et al., "Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17(6), 2009, pp. 1186-1195.

[9] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients," IEEE Transactions on Biomedical Engineering, vol. 58(2), 2011, pp. 370-379.

[10] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," Europhys. Lett. vol. 5, 1987, pp. 973-977.

[11] J. P. Zbilut and C. L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," Phys. Lett. A vol. 171(3-4), 1992, pp. 199-203.

[12] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence Plots for the Analysis of Complex Systems," Physics Reports vol. 438(5-6), 2007, pp. 237-329.

[13] F. Takens, "Detecting strange attractors in turbulence," Lectures Notes in Mathematics, Vol. 898, 1981, pp. 366-381, Springer.

[14] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A, vol. 33, 1986, pp. 1134-1140.

[15] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," Phys. Rev. A, vol. 45, 1992, pp. 3403-3411.

[16] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of german emotional speech," Proceedings on Interspeech, Lisbon, Portugal, 2005, pp. 1517-1520.

[17] H. Akaike, "A new look at the statistical model identification," IEEE Transaction on Automatic Control, vol. 19(6), 1974, pp. 716-723.

[18] J. P. Burg, "Maximum entropy spectral analysis," in Proc. 37th Meet. Soc. Explorational Geophys., Oklahoma City, 1967.

[19] B. I. Helme and Ch. L. Nikias, "Improved spectrum performance via a data-adaptive weighted Burg technique," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33(4), 1985, pp. 903-910.

[20] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," Technical report, Microsoft Research, MSR-TR-99-87, 1999.