FUZZY LOGIC BASED VOICE PROCESING

Wladimir Rodríguez
Departamento de Computación
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

ABSTRACT

A new approach for analyzing the similarity of dynamical systems is presented with applications to the analysis of voice. This approach is based on a fuzzy information space representation of the trajectories of the voice signal. The similarity between the segments of voice signal is determined based on similarity measures of the corresponding fuzzy information space representation. We present an application of the method to vowels recognition in the samples (amplitude-time) space.

Keywords: speech analysis, non-spectral analysis, fuzzy information space, similarity measure, fuzzy classification.

1. Introduction

Traditional approaches to system analysis - e.g. trying to find a mathematical model that describes the output as a function of the state variables and the input – perform poorly when dealing with complex systems. This may be due to their nonlinear, time-varying nature or to uncertainty in the available measurements. One approach to solving this problem is to use pattern matching techniques to generate a model of the dynamics of the system, based on a representation of the output trajectories using temporal fuzzy sets [1, 6, 7]. The technique proposed in this paper is based on a representation of system trajectories in a Fuzzy Information Space [4, 5]. A Fuzzy Information Space is characterized by a collection of temporal fuzzy sets [4], which decompose the observed trajectories into simple modes of dynamic activities. Each collection of temporal fuzzy sets is called the dynamic profile of a physical system, and can be used to construct a dynamic fuzzy set, which generates a system trajectory. This dynamic profile of a physical system represents a decomposition of the dynamic trajectory in feature space. A temporal fuzzy set belonging to the dynamic profile characterizes a region of attraction in feature space, and quantifies the extent to which the observed dynamic system is governed by that region at any time. Then, by measuring the similarity of these temporal fuzzy sets, we may infer some characteristics of the underlying dynamical system. This can be considered a dynamic approach, because it takes into account the temporal elements of the system in question.

2. Dynamics of the Speech Signal

The understanding of how humans produce, hear and recognize speech signals is important in computer science, communication and medicine. But several of the mechanisms in speech production, perception and understanding are still insufficiently understood.

The desire to exploit nonlinear techniques in analyzing intricate aspects of speech production has become significant. Also, there are reported researches on nonlinear processes in sound production by several musical instruments, mainly those with reeds. These processes are known to undergo dynamics similar to speech generation at the level of vocal cords, as well as at the level of the velum.

In this paper, we discuss how we can perceive the dynamics of the speech signal. By representing the signal in phase space we are able to graphically represent the attractors of the corresponding signals.

Drawing attractors of the acquired speech signal gives an intuitive picture of its dynamics. The phoneme attractors were constructed with the help of phase space reconstruction of the time series representing the speech signal. Commonly used maps for graphic representations in the phase space are:

$$x = x[k]; \qquad y = f(x[k-1]) \quad (2D)$$

Those maps work well in the case of systems of difference equations of relatively low order (lower than 3). Given the fact that the speech signal is the projection of a dynamic process with several degrees of freedom, we prefer to define the phase space in the following manner:

$$x = x[k]; \qquad y = x[k+1] - x[k-1] \quad (2D)$$

where x[k], k = 1, , N, denotes the samples of the vocal signal and k is the time.

The speech signals used here are 2-dimensional. Figure 1 represents the /e/ vowel attractor, built up using 8000 samples of the speech signal of a person saying *e* (360 msec. of speech signal, @22Ksamp/sec), one can see that the attractor's tube has a larger section, filling the phase space and tending to cover the small windings.

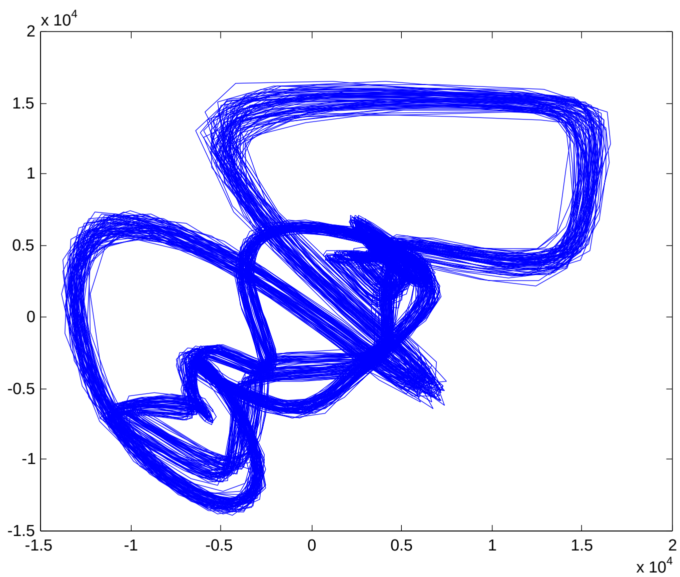

Figure 1. The /e/ vowel attractor, 8000 samples of speech signal (360 msec. of speech signal)

In the next section we will proposed a new technique to voice signal recognition based on the fuzzy information space representation of the voice signal.

3. Similarity Analysis of the Speech Signal

We will analyze speech signals by measuring the similarity of their Fuzzy Information Space representation. We will use a similarity measure, and the Gustafson-Kessel version of the fuzzy c-means algorithm, in order to verify the proposed methodology.

We will use a total of six sets of speech data, uttered by two speakers, each pronouncing the vowels /a/, /e/, and /i/. The data was taken from two female Romanian speakers, one having a low pitch and the other one having a medium pitch. We will use the graphic representation presented in the above section to represent the vowel attractors and the similarity analysis will be performed using that representation.

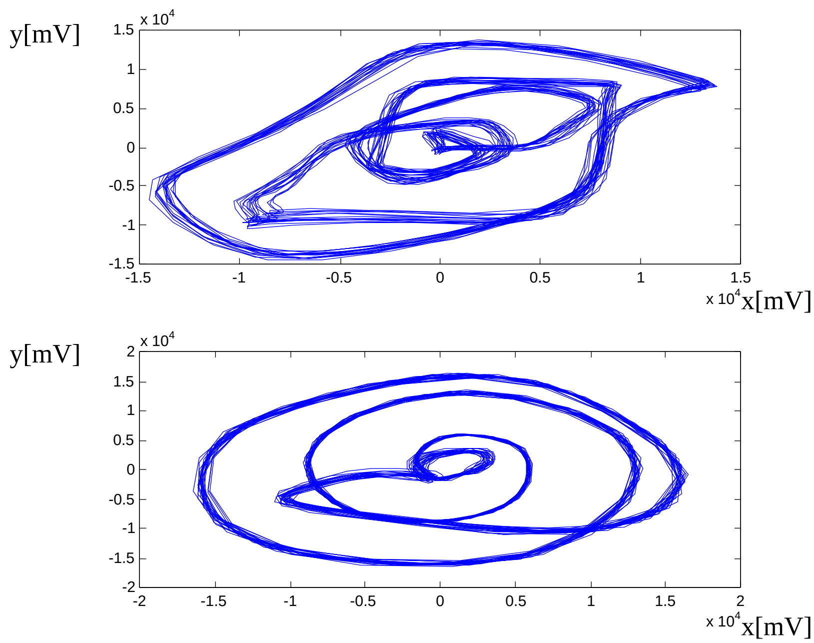

Figure 2. A comparison between two speakers uttering the /a/ vowel (2000 samples, 90 msec. of speech signal)

Figure 2 shows the /a/ vowel uttered by the two speakers, the Figure 3 shows the vowels /a/ and /e/ uttered by the same speaker and the Figure 4 shows two phonemes (/a/ and /e/) uttered by the different speakers.

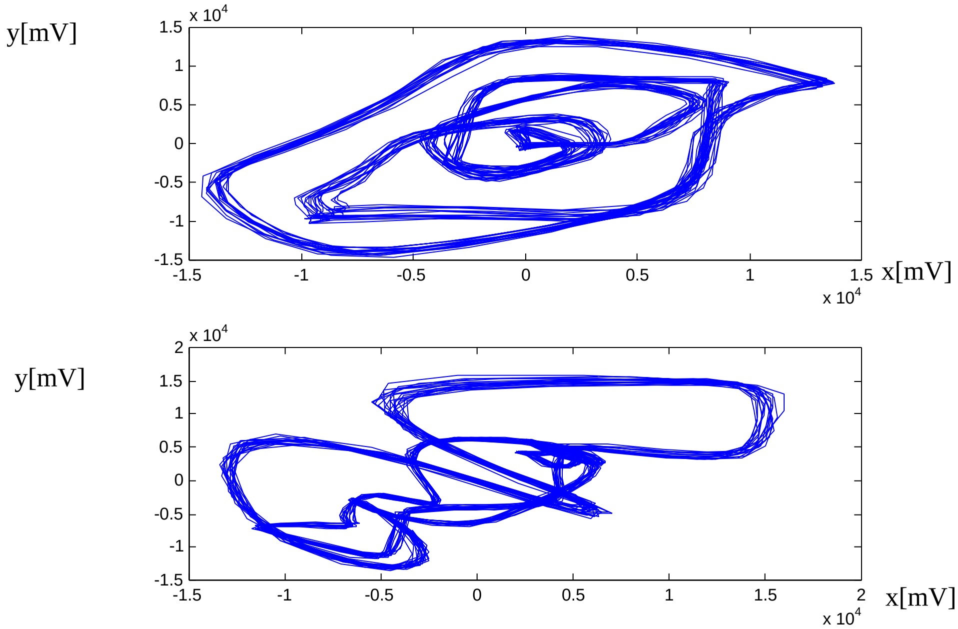

Figure 3. Two phonemes (/a/ and /e/) uttered by the same speaker.

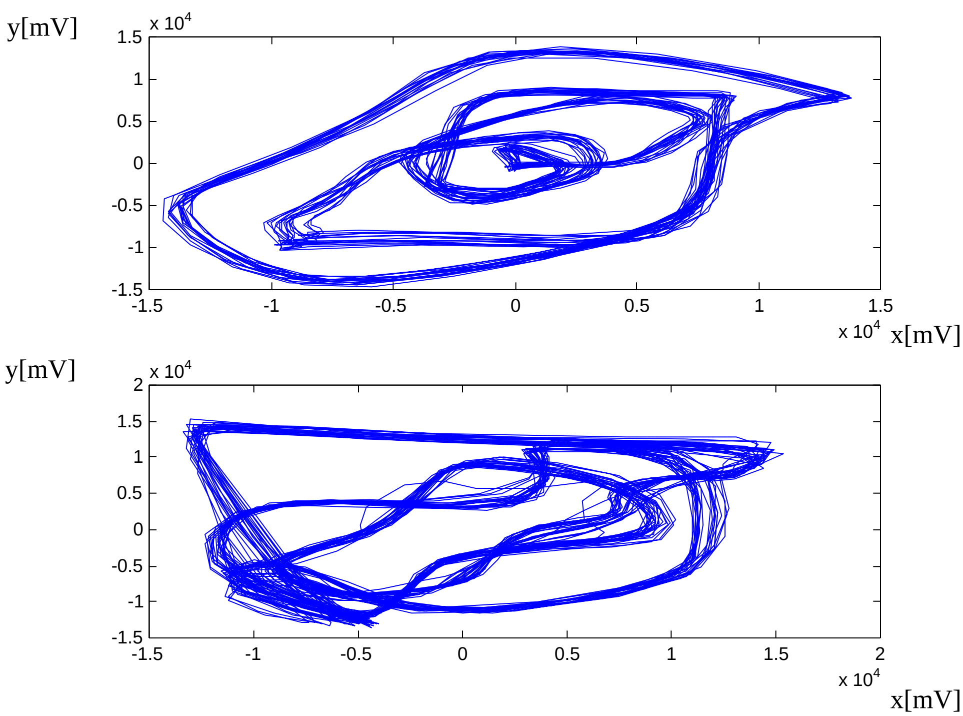

Figure 4. The two phonemes (/a/ and /e/) uttered by the different speakers.

In order to represent the vowel attractors in the Fuzzy Information Space we must perform a clustering of the data. As stated before, we are using fuzzy clustering as the clustering method and in particular we are using Gustafson-Kessel fuzzy c-means [3]. One goal of the fuzzy c-means algorithm is to find the “optimal” partitioning of the feature space. For the six data sets the “optimal” number of partitions is three, according to use of the partition and correlation index S [8].

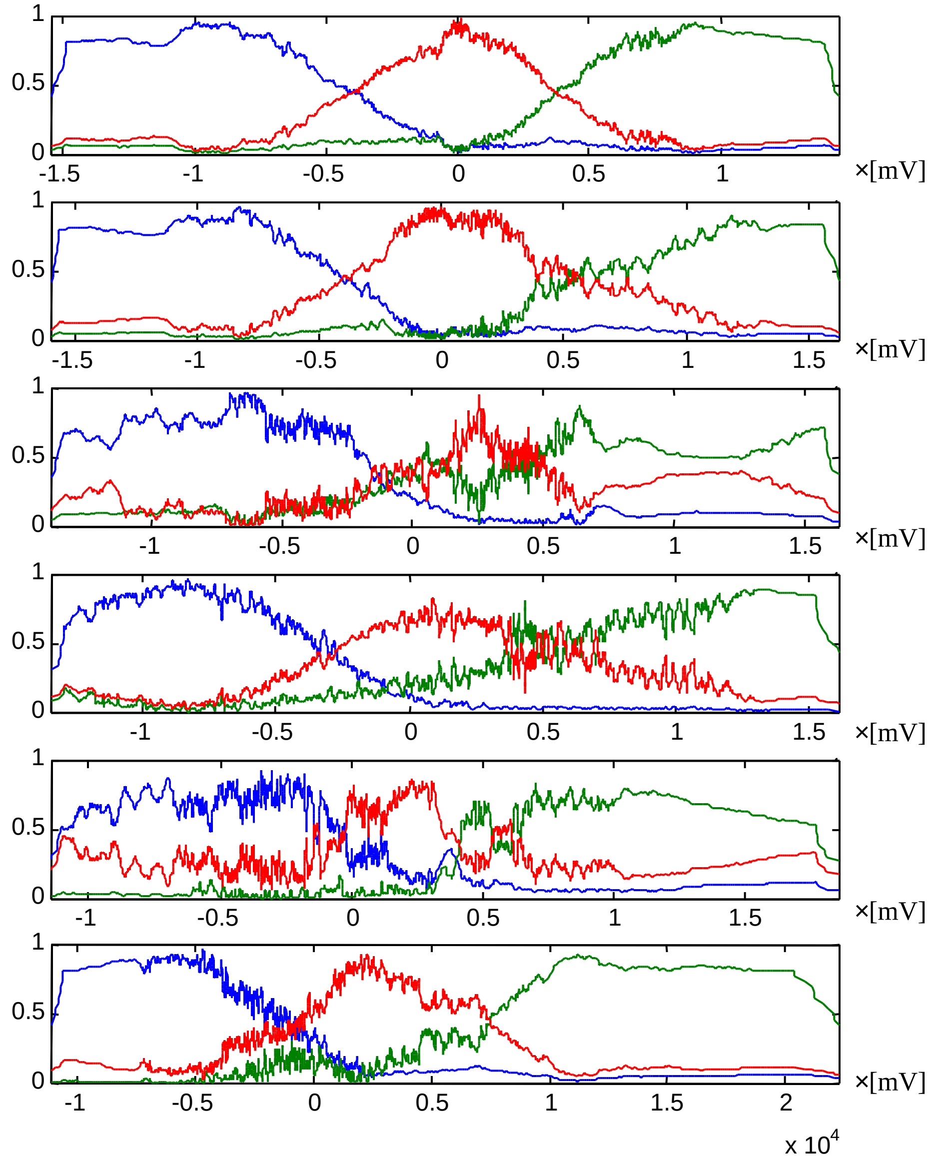

Figure 5. Representación de todos los atractores de las vocales utilizando GK-fuzzy c-means. a) *a-low* b) *a-med* c) *e-low* d) *e-med* e) *i-low* f) *i-med*

In Figure 5 we show the smoothed version of the temporal fuzzy sets of all six data sets using the Gustafson-Kessel's version of fuzzy c-means.

The last phase of the speech signal analysis is to measure the similarity between the different vowels, and attempt to recognize or classify them. In this paper we will use the similarity measure proposed by Dubois and Prade in [2], which is based on the set theoretic operations of intersection and union:

$$S(A,B) = \frac{\sum_{j=1}^{m}(\mu_A(x_j) \quad \mu_B(x_j))}{\sum_{j=1}^{m}(\mu_A(x_j) \quad \mu_B(x_j))} \qquad (3.1)$$

which it is one of the most widely used similarity measures.

In order to measure the similarity between the fuzzy information space representation of two dynamical systems, we will measure the similarity between each of the temporal fuzzy sets that represent the grade of membership of the signal samples of each cluster. Then we will take the mean value as the similarity value. We use the mean value instead of the maximum or the minimum. In the case of the maximum value the similarity will be only based on the most similar clusters. In the case of the minimum value the similarity will be based only on the most dissimilar clusters.

In order to test the technique we performed an experiment, using a total of six sets of speech data uttered by two speakers, with each speaker pronouncing the vowels /a/, /e/, and /i/. The six sets of speech data will be denoted as follows: *a-low*, *a-med*, *e-low*, *e-med*, *i-low* and *i-med*. The purpose of the experiments is to determine if the method allows us try to recognize a vowel despite the fact that they were pronounced by different speakers. Notice that the method is based on the amplitude-time representation of the speech signal. Classic methods used to perform vowel recognition in the amplitude-time space are widely used to be unreliable.

The experiment was performed using Gustafson-Kessel fuzzy c-means as the clustering algorithm. Table 1 shows the results of the experiment and they also show that the use of the Gustafson-Kessel fuzzy c-means as the clustering algorithm gives the best results with the selected similarity measure.

Table 1. Results using Gustafson-Kessel fuzzy c-means and the similarity measure based on the intersection and union.

	a-low	a-med	e-low	e-med	i-low	i-med
a-low		0.7529	0.5568	0.6340	0.5459	0.5322
a-med	0.7529		0.5143	0.5608	0.5491	0.5983
e-low	0.5568	0.5143		0.7466	0.6381	0.6251
e-med	0.6340	0.5608	0.7466		0.6049	0.6371
i-low	0.5459	0.5491	0.6381	0.6049		0.6450
i-med	0.5322	0.5983	0.6251	0.6371	0.6450	

Note in Table 1 that all vowels are recognized correctly, indicating that the Gustafson-Kessel fuzzy c-means is an appropriate clustering algorithm for this data.

From the experiment we can conclude that the fuzzy information space representation is a valid approach to study nonlinear signals such as the speech signals.

4. Conclusions

We found that it is reasonable to study the behavior of dynamical systems using fuzzy information space paradigm. Under certain circumstances, like dealing with complex systems (nonlinear, time-varying nature and uncertainty in the available measurement data) this could be the only available path to follow.

The fuzzy information space based technique simplifies the study of complex systems by focusing on the most relevant dynamic changes occurring in the observed behavior. This technique provides qualitative information about the system motion in the feature space. Another advantage is that this technique is based on the physical properties of the regions of attraction, which are characterized by temporal fuzzy sets. This technique will perform well, even on single observations of a system's behavior. As compared to other methods (e.g. NN recognition), it requires no training and only minor preprocessing of the speech signal (i.e. finding the number of clusters). Signal unification (normalizing) is automatically obtained by considering the fuzzy approach.

Further research is necessary to see if the results can be scaled to a more complex set of data, more vowels, more speakers, other languages, other clustering algorithms, other similarity measures etc. Nevertheless, the results are encouraging, and give a new direction for the study of nonlinear dynamical systems.

References

[1] Y. C. Chen and T. C. Tsao, "A Description of the Dynamical Behavior of Fuzzy Systems". *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, No. 4, July/August 1989.

[2] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

[3] E. E. Gustafson and W. C. Kessel, "Fuzzy Clustering with a Fuzzy Covariance Matrix", in Proceedings of the IEEE CDC, pp. 761-766, 1979.

[4] B. R. Kosanovic, L. F. Chaparro, M. Sun, and R. J. Sclabassi. "Physical System Modeling Using Temporal Fuzzy Sets". *Proceedings of the International Joint Conference of NAFIPS/IFIS/NASA*. San Antonio, TX, pp. 429-433, December 1994.

[5] B. R. Kosanovic, L. F. Chaparro, and R. J. Sclabassi. "Signal Analysis In Fuzzy Information Space", *Fuzzy Sets and Systems*, vol. 77, pp. 49-62, 1996.

[6] W. Rodríguez, H. N. Teodorescu, A. Kandel, H. Bunke, "A Fuzzy Information Space Approach to Movement Nonlinear Análisis", *International Journal of Chaos Theory and Applications*, vol.3, No. 1-2, pp. 29-37, 1998.

[7] W. Rodríguez, "Fuzzy Sets Based Similarity of Dynamical Systems", *Proceedings of the 5th International Conference on Information Systems Analysis and Synthesis ISAS´99*, Orlando, Florida, p.p. 297-302.

[8] X. L. Xie and G. Beni, "A Validity Measure for Fuzzy Clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 841-847, 1991.