Using Vector of Fractal Dimensions for Feature Reduction and Phoneme Recognition and Classification

S. Abolfazl Hosseini ,*student Member, IEEE,* Hassan Ghassemian, *Senior Member ,IEEE* , and Roya Alizadeh

***Abstract* — Difference between Hausdorff fractal dimensions of phonemes gave us a motivation to use this feature as input of a statistical Bayesian classification system and a nearest neighborhood (NN) classifier for speech waveform recognition. We divide phoneme waveforms to adjacent segments and calculate Hausdorff fractal dimension of each segment and using them as the input of a Bayesian/Nearest Neighborhood classifier. The power point of algorithm is in consideration of order of samples information in contrast of other non-supervised feature extraction algorithms.**

***Keywords* — Classification, Feature extraction, Fractal dimension, Phoneme, Speech recognition.**

I. Introduction

Classification of Audio signals has many applications in a variety of research fields such as broadcast browsing, audio content analysis, and information retrieval. Although speech processing has been developed for many years, it still suffers from some problems like human and environmental factors. In different areas like speaker identification, phoneme recognition, speech coding and many other domains various speech models are used to obtain the parameters required for audio signal analysis. The coefficients of the linear predictive coder (LPC), cepstral coefficients and formant energies are some widely used parameters [1]. Also a fractal theory based approach is used by some researchers [2]-[4]. Fractal geometry could be applied successfully to some problems involving complex natural shapes like speech waveforms [5]. Speech waveforms are two dimensional open curves which have sever fluctuations and seem very crumple. If amplitude fluctuations are viewed as irregularities, the fractal dimension of the signal is a measure of this irregularity [6].The main argument is that if fractals can effectively model chaotic nature of phenomena and if speech is a natural chaotic phenomenon, then fractals should be a promising model for speech [7]. Phonemes, the basic parts of human speech, show a high degree of repetition and from a fractal point of view, this high degree of repetition could be view as self-similarity and frequency content is the roughness, which is measured by the Fractal Dimension [8]. Boshoff and Grotepass determined dimension of fricative speech sounds using Hurst's rescaled range analysis and showed that voiced fricatives tend to have a lower dimension than their unvoiced counterparts and this fact was used as a criterion for making the voiced/unvoiced decision in recognition of fricatives, and generating balanced LPC code book samples to match all fricatives[5]. Langi presented consonant characterization using correlation fractal dimension [9],[10]. Fractal Dimension of the speech signals and eigenvalue of the co-variance matrix of the IFS parameters a,c,d, appears effective in phoneme recognition. Erik and Bohez presented a speech recognition method based on two-level Clustering using these features[8]. Bohez and Van Winden used fractal dimension of speech waveform and the IFS parameters that represent the waveform as the parameters for recognition [6]. Personal authentication for speaker recognition based on Mel-scale spectral dimension and Mel Frequency Cepstral Coefficients (MFCC) or Multi-Scale Fractal Dimension (MFD) were presented in [11],[12]. Petry, and Barone obtained a speaker identification system based on Bhattacharya distance, which combines LP derived cepstral coefficients, with fractal dimension [13]. Fractal dimension and fractal complexity are presented as feature space effective compensations for speaker gender identification [14]. To improve the low recognition rate of speaker-independent recognition in intelligent robot, a kind of endpoint detection algorithm with double threshold is planned and the speech endpoint can be detected accurately using the mixed parameter of (MFCC) and fractal dimension as feature parameters, and the intelligent robot command-word recognition system is realized based on Hidden Markov Models [15]. In [16] MC fractal dimensions applied to represent the stress features in speech signals and a stressed syllable labeling approach based on it. Also fractal-based approach for speech segmentation is introduced by [2]-[4].

In this paper we introduce a new way of application of fractal dimension to classify phonemes. The method

(The first line before the Abstract was intentionally left blank. Do not delete it.)

This research is done by support of Iran communication research center under contract number 18133/500 T by Identification code: 90-01-03. The authors gratefully acknowledge that organization for its support. Also the authors gratefully acknowledge the National Elite Foundation for its support.

S. Abolfazl Hosseini is with the Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Chamran Blv., Gisha Bridgje, Tehran, Iran.(phone: 98-21-82884330; e-mail: abolfazl.hosseini@modares.ac.ir).

Corresponding Author Hassan Ghassemian is with the Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Chamran Blv., Gisha Bridgje, Tehran, Iran.(phone: 98-21-82884330; e-mail: ghassemi@modares.ac.ir).

Roya Alizadeh,, is now with K.N.Toosi University of Technology, Resalat Blv, seyedkhandan Bridge, Tehran, Iran; (phone: 98-912-5385967, e-mail: alizadeh.roya@gmail.com).

uses a vector whose elements are fractal dimensions of adjacent time segments of each phoneme waveform. This feature vector is given to a Bayesian classifier. Fig.1 shows the block diagram, which is self-explanatory. So this algorithm is used not only for purpose of dimension reduction, but also for discriminating and classifying phonemes.

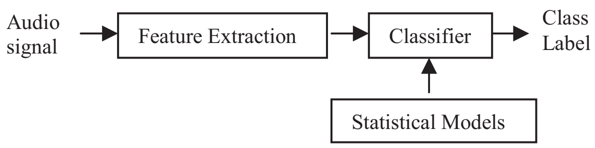

Fig. 1. Block Diagram of an Audio Signal Classification System

In section II, we introduce the definition of fractal dimension and different methods to calculate it. In section III, we explain our audio signal features, classification method based on Bayes classifier and the experimental results of the speech recognition. And section IV finalizes this paper with conclusion and future works.

II. Fractal Dimension

Fractal Dimension (FD) provides a way to describe and analyze the geometry of linear figures. There are different definitions such as Box dimension, Entropy dimension, Similarity dimension and Hausdorff dimension for dimension of a fractal. Taylor's definition says fractal is a set that its Box dimension is equal to its Hausdorff dimension and is greater than its topologic dimension. For such set all above dimensions are equal [17]. FD could be calculated by a number of methods like Variance method, spectral method and morphological covering method. In this study, FD of speech waveforms (especially phonemes) has been obtained making use of Hausdorff metric. The Hausdorff dimension (D_h) of a set in a metric space is given by:

$$D_h = \lim_{\varepsilon \to 0} \frac{-\ln[N(\varepsilon)]}{\ln(\varepsilon)} \tag{1}$$

Where N(ε) is total number of radius ε open balls we need to cover the set. In case of a line of length L consisting of segments of length 2ε each, there will be N(ε) =L/(2ε) segments in the line. Thus, (1) will be changed to:

$$\begin{aligned} D_h &= \lim_{\varepsilon \to 0} [\frac{-\ln(L)+\ln(2\varepsilon)}{\ln(\varepsilon)}] \\ &= \lim_{\varepsilon \to 0} [1-\frac{\ln(L)-\ln(2)}{\ln(\varepsilon)}] \\ &= \lim_{\varepsilon \to 0} [1-\frac{\ln(L)}{\ln(\varepsilon)}] \end{aligned} \tag{2}$$

In this study, the waveform of each phoneme, a planar curve in a space, was considered as the metric space. Then the waveform was scaled linearly into a normalized space, to map the original waveform into an equivalent metric space. The first scaling normalizes sample number of every point in the abscissa as:

$$x_i^* = x_i / x_{\max} \tag{3}$$

The second scaling normalizes the ordinate as:

$$y_i^* = (y_i - y_{\min})/(y_{\max} - y_{\min}) \tag{4}$$

Where yi is value of the i'th time sample of waveform and y_{min} and y_{max} are the minimum and maximum values of the sample values respectively, of the considered phoneme over the entire time duration. These two linear scaling map the N values of the waveform into another space that belongs to a unit square. This unit square may be visualized as covered by a net of N(sample number) × N (sample value) cells. Each of them contains one point of the scaled waveform. Calculating L of the scaled waveform and taking ε = 1/(2 × N′) [where N′ (the number of segments) = N – 1], Equation(2) becomes:

$$\begin{aligned} D_h &= \lim_{N' \to \infty} [1-\frac{\ln(L)}{\ln(1/2(2N'))}] \\ &= \lim_{N' \to \infty} [1-\frac{\ln(L)}{\ln(1)-\ln(2N')}] \\ &= \lim_{N' \to \infty} [1+\frac{\ln(L)}{\ln(2N')}] \end{aligned} \tag{5}$$

Therefore, the fractal feature:

$$D_h \approx 1 + \frac{\ln(L)}{\ln(2N')} \tag{6}$$

Where:

$$L = \sum_{i=1}^{N} dist(i, i+1) \tag{7}$$

The approximation to D expressed in (5), improves as N→ ∞.

In TABLE 1 you can observe some statistical parameters (mean, variance, maximum and minimum values) of Hausdorff fractal dimension of some phonemes. These phonemes are selected by chance just for avoiding long size of this paper. Fig.2 demonstrates frequency polygons of fractal dimension of same phonemes. Difference between fractal dimension statistics of phonemes gave us a motivation to expand and improve the way of FD utilization in speech recognition. In the next section we propose a technique which use a number of FD's corresponding to adjacent temporal parts of phoneme waveforms to distinguish them better.

TABLE 1: STATISTICAL SPECIFICATIONS FOR FD OF SOME PHONEMES

Phoneme	Mean	Variance	Maximum	Minimum
ʌ	1.5926	0.0003	1.6608	1.5568
l	1.6253	0.0007	1.6737	1.5247
ʒ	1.6103	0.0005	1.6403	1.5157
ʊ	1.5749	0.0014	1.6434	1.4532
ʃ	1.5741	0.0049	1.6436	1.4109
t	1.5260	0.0006	1.5683	1.4506

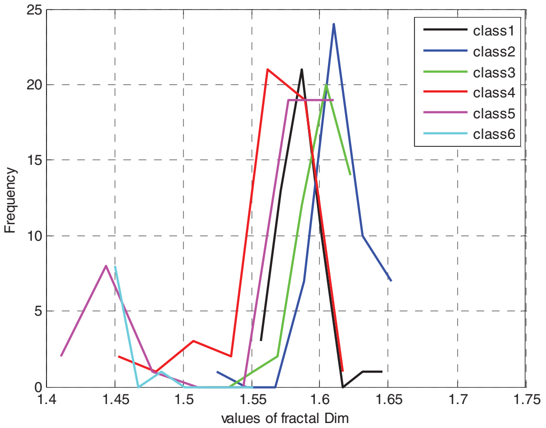

Fig. 2. Frequency Polygon of FD for Phonemes in TABLE 1

III. DATA BASE AND EXPERIMENT RESULTS

Based on reference [18] we selected 50 different pronunciations of every 46 English phonemes produced by a specific person as our initial waveforms. Then we sampled these analog waveforms using 8kb/s sample rate. Fig. 3 demonstrates sampled waveform corresponding to phoneme 'a' as "ago" as an example. All signals were down sampled to 1200 samples by keeping samples greater

than 5% of maximum waveform amplitude. These vectors which have dimensions equal to 1200 were considered as initial feature vectors in Fig.1. Our proposed feature extraction method concludes dividing these samples to 'n', almost, equal length adjacent segments and calculating Hausdorff FD of each segment by using the algorithm mentioned in previous section. After that, we formed new feature vectors containing 'n' calculated FD. To have better performance for lower dimensions we calculated principle components of these new features and applied them to our classifiers and then we compared it to results gotten from applying principle components of original data straightly.

This new feature vector is considered as the input of a two different classifier; a Bayesian classifier and a nearest neighborhood classifier. For Bayesian classifier, statistical models of each 46 class were considered as Gaussian probability density functions. Their parameters, mean and covariance matrix, were estimated by using 40 training samples for each class. The other 10 data points dependent of each phoneme were considered as test points.

For statistical classifier we used Mahalanobis distance as shown in Fig.4. For both plans, performance of classifiers using these fractal features is compared to result of using Principle Component Analysis (PCA) as a basic feature reduction technique which removes linear dependence of data dimensions. Regarding to initial dimension of 1200, the effective dimension of principle components of data which contains 95% energy of signal was about 470, although due to limitation of training data to 40 data points for each class and to avoid singular covariance matrix in Mahalanobis distance, we selected the first 'n' principle components of data points, 'n' varies from 1 to 39 In statistical classifier for PCA features. The number of fractal features varied from 1 to 61, where the effective dimension reached to 39.

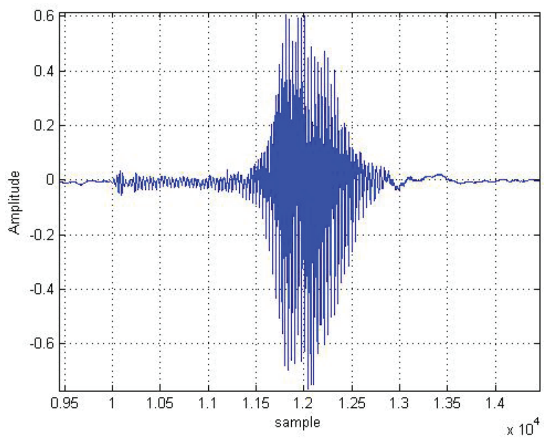

Fig.3. Speech Waveform of Phoneme 'a' (as ago) Sampled by 8 kb/s Sample Rate

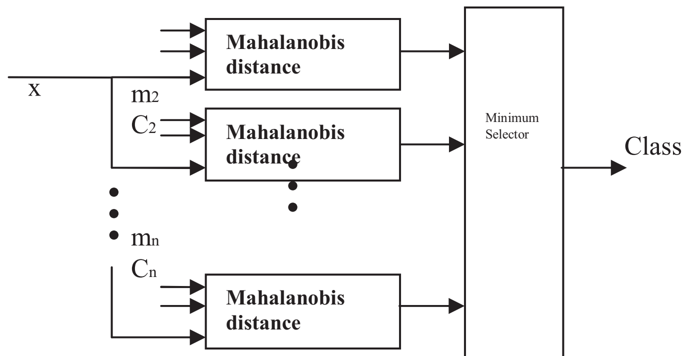

Fig. 4. Block Diagram of a Bayesian Classification System Based

Fig.5 illustrates Correct Classification Rate (CCR) for various executions of the proposed algorithm. It could be seen that performance of FD features are so much better than PCA performance for equal or some higher dimensions. The main reason for such better performance is that in classical feature reduction methods, the method is not sensitive to location of elements in original feature vector. In other word methods like PCA do not regard to ordinance of features and if you for example exchange 2nd and 4th elements of all data points, principle components don't change. Where in our application, the samples in measured vector of phoneme are members of a time sequel and by the proposed method, the effect of reposition would be considered.

In another experiment we applied Nearest Neighborhood (NN) classifier to examine and compare the power of FD features to PCA features, using Euclidean distance. As you see in Fig. 6 the performance of PCA features is not absolutely comparable to FD features. It should be mentioned that CCR of this classifier dramatically decrease as the number of features increase more than 15.

IV. CONCLUSION

Different parameters like pitch, cepstral coefficients, and formant energies are used for speech recognition. Fractal nature of speech waveforms gives us a motivation for

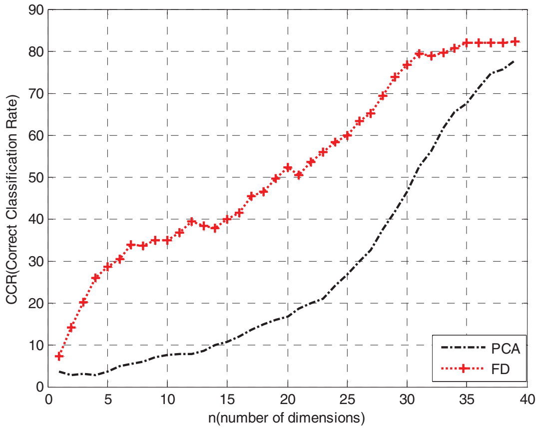

Fig. 5. Correct Classification Rate (CCR) vs. Number of Dimensions for Statistical Classifier-(original data size is 1200)

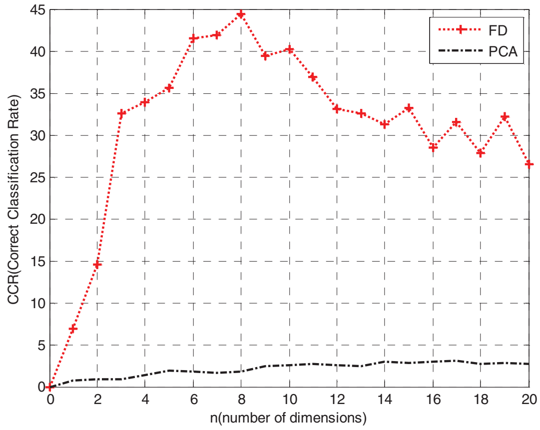

Fig 6. Correct Classification Rate (CCR) vs. Number of Dimensions for Nearest Neighborhood Classifier- (original data size is 1200)

applying fractal dimension properties of them for this purpose. A fast and simple algorithm was suggested that exploit a vector whose elements are Hausdorff FD of adjacent temporal parts of phonemes waveforms as input of a statistical classifier based on Bayesian theory using Gaussian probability density function model and Mahalanobis distance, or a NN classifier to classify phonemes. The advantages of algorithm are in consideration of order of samples in a feature vector and in low computation cost of FD calculation. The results show that the method has an acceptable performance specially for reducing dimension of data from 1200(or higher) to near 40(or less) especially for low training data number. So the algorithm is used not only for purpose of dimension reduction, but also for producing good discriminative features to classify phonemes. Further research is ongoing to use a Gaussian mixture density model for classes statistics and also merging FD to classic features like pitch, cepstral coefficients, and formant energies in feature level or in decision level.

ACKNOWLEDGMENT

This research is done by support of Iran communication research center under contract number 18133/500 T by Identification code: 90-01-03. The authors gratefully acknowledge that organization for its support. Also the authors gratefully acknowledge the National Elite Foundation for its support.

REFERENCES

[1] H. Harb, and L .Chen, "Voice-based gender identification in multimedia applications," *Journal of Intelligent Information Systems*, vol. 24, no. 3, pp. 179-198, 2005.

[2] S. Fekkai, M. Al-Akaidi, and J.Blackledge, "*Fractal dimension segmentation: isolated speech recognition,*'' proc. ICASPAT 2000, Dallas, USA, October 2000.

[3] W. Kinsner and W. Grieder, "Speech segmentation using multifractal measurs and amplification of signal features'', proc. 2008 IEEE International Conference Series :pp. 351-357.

[4] Z.Armein , R. Langitt K.Soemintapurat, and W. Kinsners, "Multifractal Processing of Speech Signals'', in Proc. of 1997 International Conference on Information,Communications and Signal Processing.

[5] H. F.V. Boshoff and M. Grotepass, "*The Fractal Dimension of Fricative Speech Sounds'',* In proc. 1991 South African Symposium on Communications and Signal Processing, pp.12 - 16

[6] E. L. J. Bohez, T. R. Senevirathne and J. A. Van Winden, "*Fractal dimension and iterated function systems(IFS)for speech recognition*'', Electronics Letters. Vol. 8;pp.:1382-1384, 1992.

[7] J.Keller, R.Crownover, and R.Yu Chen, "*Characteristics of natural scenes related to the fractal dimension*," IEEE Trans. Pat. Anal. & Mach. Intel.,Vol. PAMI-9, No. 5, September 1987, pp 621-627.

[8] E. L. J. Bohez, "Two level cluster analysis based on fractal dimension and iteratedfunction systems (ifs) for speech signal recognition," IEEEAsia-Pacific Conference on Circuits and Systems, pp. 291–294, 1998.

[9] F. Albregtsen, B.t. Nielsen and K. Yogesan, "Fractal dimension, only a fraction of the truth ?" Proc.1992 International Conference on pattern Recognition, Vol.III. Conference C: Image, Speech and Signal Analysis,pp.733-736

[10] R. Senevirathne, E. L. J. Bohez and I.A . Van Winden, "*Amplitude scale method: new and efficient approach to measurefractal dimension of speech waveforms''* Electronics Letters, Vol. 28, , pp. 420 – 422,1992

[11] Chen,J.F.Huang, and P.Nontou "Speaker recognition using spectral dimension features'', proc. Fourth International Multi-Conference on Computing in the Global Information Technology, pp.: 132 – 137,2009

[12] F.V. Nelwamondo, U.Mahola and T. Marwala, "*Improving Speaker Identification Rate Using Fractals'', proc.* 2006 International Joint Conference on Neural Networks, pp.3231 – 3236

[13] A.Petry, and D.A. C Barone, "Fractal dimension applied tospeaker identification'', proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001.
F. Yingle, Y. Li and T. Qinye, *"Speaker Gender IdentificationnBased on Combining Linear and Nonlinear Features'',* proc. 7th World Congress on Intelligent Control and Automation, pp. 6745 – 6749,2008.

[14] Y. Zhang, Y. Li, L. Zeng, and Q. Liu , "Research on intelligent robot command-word recognition system basedon feature extraction.'', in proc. 2009 Chinese Conference on Pattern Recognition, pp. 1 - 4

[15] Q. Chen, D. Wang, and X. Wang, "*A stressed syllable labeling approach using fractal dimensions.''* In proc. Eighth International conference on Intelligent Systems Design and Applications, pp.353 – 357,2008

[16] J.K. Ghosh, A. Somvanshi and R. C. Mittal, "*Fractal feature for classification of hyperspectral images of Moffit Field,USA''*, Current Science 94 (2008) pp 356-358.

[17] Oxford dictionary 2009 ,Oxford university press available :http://www.oupchina.com.hk/dict/phonetic/home.html