FAST ALGORITHM FOR VECTOR QUANTIZATION OF SPEECH

By

M. Elshafei Ahmed

HOSTEK, House Of Systems Technology
P.O.Box 24, Special District,
Madinat 6 October, Giza 12568

M. I. Al-Suwaiyel

King Abdulaziz City for Science and Technology
P.O. 6086, Riyadh 11442,
Saudi Arabia

Yilmaz Akyildiz

College of computer Science and Engineering
King Fahd University of Petroleum and Minerals,
Dhahran, Saudi Arabia

ABSTRACT

This paper proposes a new algorithm for Vector Quantization of speech using the Space Filling Curve heuristic (SFC). The Space Filling Curve method scans an N-dimensional space and reduces it to a one-dimensional line. Points close to each other on the resulting line are the mapping of points close to each other on the N-dimensional space. The paper presents an evaluation of the method as applied to vector quantization of a Code Excited Linear Prediction CELP speech Coder. The method achieves a dramatic saving in the complexity of Vector Quantization, and is particularly attractive for integer implementation in real time.

I. INTRODUCTION:

The objective of this paper is to present a new approach for Vector Quantization (VQ) of speech [1] based on the Space-Filling Curve (SFC) technique [2]. The SFC represents a powerful method for mapping a high dimensional signal vector into a reduced dimensionality feature vector or to a single dimensional vector. Space-filling curves were first described by the mathematicians Peano, Hilbert and Sierpinski at the beginning of the century as "topological monsters" since they seem contrary to intuition that a lower-dimensional space can be mapped continuously onto a space of higher dimension.

An example of a space-filling curve on the plane is illustrated in Figure 1. It has been first presented by David Hilbert in 1891, [1]. One partitions the interval [0,1] into 4 congruent intervals and the square [0,1] x [0,1] into 4 congruent squares which are to be enumerated so that each two consecutively numbered squares have an edge in common(Fig.1(a).) If the entire interval can be mapped onto the entire square, then [0,1/4] can be mapped onto square 1,[1/4, 1/2] can be mapped onto square 2, etc. Again, partition every subinterval into 4 congruent intervals, each sub-square into 4 congruent squares, enumerate the squares so that the preceding mapping is preserved and that each two consecutive squares have an edge in common as depicted in Fig.1(b), and repeat the previous argument. The next step is illustrated in Fig. 1(c). Proceeding in this manner at infinitum, one obtains a mapping that is, by construction, continuous. It is shown in [3] that the Space-Filling Heuristic requires O(nlog n) arithmetic operations and O(n) memory registers, where n is the number of data points.

II. VQ OF SPEECH USING SFC :

The SFC technique is applied to the problem of Vector Quantization of the LPC filter parameters of a CELP speech coder [4,5]. VQ is performed by selecting the best filter, from a codebook of filters, to approximate a given formant filter vector X. The codebook usually contains from 1024 to 4096 fixed filters. In this case, only the 10 to 12 bit index of the best approximating filter need to be transmitted instead of transmitting all the coefficients of the formant filter.

Application of the SFC here proceeds first by applying the SFC mapping to all the vectors in the codebook, i.e.,for each vector Y_i in the codebook we associate a point q_i on the line. A sorted table is then constructed to provide the inverse mapping from a given point p on the line to the nearest vector Y_j. Now, to find the index j of the nearest vector to X, We first transfer X to a point p on the real line via the SFC mapping. Then, the nearest code vector j is the one having the nearest image q_j to the point $p$on the line.

However, to attain good quality of speech, VQ should be based on a distortion measure which correlates with the perceptual characteristics of the human ear. Unfortunately, the geometric closeness criterion of the SFC does not imply perceptual closeness. To overcome this problem the SFC is used to identify a nearest neighbor point $p$which is then considered to be the center of a set of candidate vectors consisting of its nearest-k (k= 16-128) elements of the codebook. Then a more expensive distortion measure may be applied to the selected neighborhood to find the perceptually nearest member to the input vector X. In CELP, the refined filter selection is based on an analysis-by-synthesis procedure which minimizes a perceptually weighted squared error between the synthesized speech and the original speech.

Furthermore, since the SFC is not used to pin point the optimum vector, but rather to identify a neighborhood of vectors, a reduced dimensional space, called the feature space, can be used in the SFC mapping. The dimension of the feature space may be taken as 2-D or 3-D.

III. EVALUATION RESULTS :

In our experiment, a trained codebook of 4096 LPC filters is used. The filters are represented by the PARCOR parameters. About a 2 second utterance, "one ..two ..three", of a male voice was used as a test sample. The speech sample was encoded using the CELP method. The S/N ratio of the synthesized speech is used for performance evaluation of the methods.

The Average S/N with no VQ was about 8.26 dB. When VQ is applied by exhaustive search of the codebook using the Itakura-Saito IS measure, the S/N becomes 7.24 dB. The SFC 2D/32 (two dimension SFC with a neighborhood of +-32) achieves a S/N of 7.69 dB. However, with 2D/64 the average S/N becomes 8.06 dB; that is a loss of only 0.20 dB. Fig. 2. Shows the S/N ratio per frame of the synthesized speech with no VQ, and the signal to Noise S/N ratio obtained with SFC 2D/64.

The experiment was repeated, however, using 3D SFC Successive Compression and various neighborhood set size as summarized in Fig. 3. The 3D SFC is better than 2D for small neighborhood. However, the 2D performance tend to grow better than the 3D-SFC with a larger Neighborhood.

In Summary, the SFC can be used as a very fast method for preselection of candidate vectors for a more expensive perceptual selection method. The SFC could lead to a valuable saving of about 600 bps in the US government speech coding standard with relatively small increase in complexity. The SFC is also believed to have potential applications in image compression and pattern recognition as well.

LIST OF REFERENCES :

[1] D. Hilbert, "Uber die steitige Abbildung einer Linie auf ein Flachenstuck," Math. Ann., 38, 1891.

[2] J.J. Bartholdi,III, and L.K. Platzman, "Heuristics based on Space Filling Curves for Combinatorial Problems in Euclidean Space", Management Science, Vol. 34, No. 3, pp. 291-305, March 1988.

[3] B.S. Atal ,"Predictive coding of speech at low bit rates," IEEE Trans. Comm. vol. COM-30, pp. 600-614, April 1982.

[4] M.R. Schroeder and B.S. Atal," Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in Proc. Int. Conf. On Acoustics, Speech and Signal Processing, vol. 1, paper no. 25.1.1,March 1985.

[5] A. Gersho and R.M. Gray, Vector Quantization and Signal Compression, Kluwer Academic Publisher, Boston, 1992.

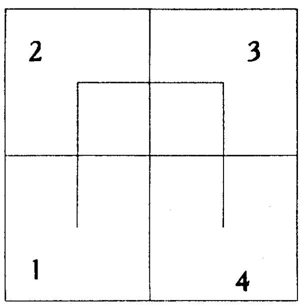

Fig. 1(a)

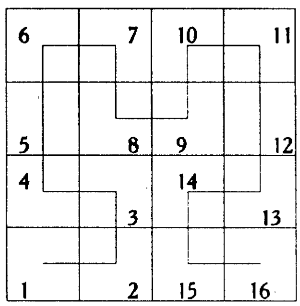

Fig.1(b)

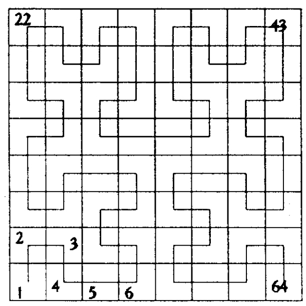

Fig.1(c)

Fig. 1. Space Filling Curve

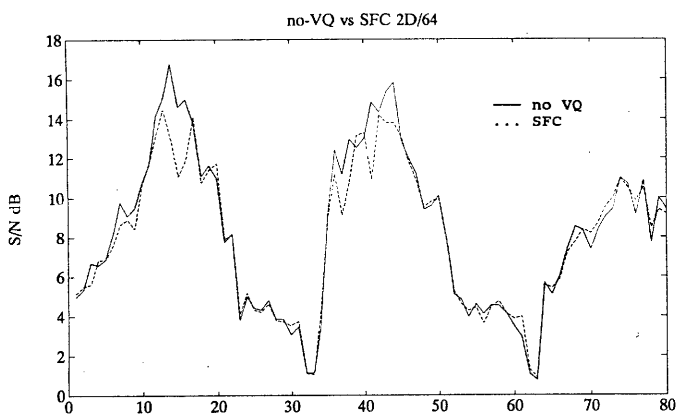

Fig.2 S/N per frame for no VQ and VQ SFC 2D/64

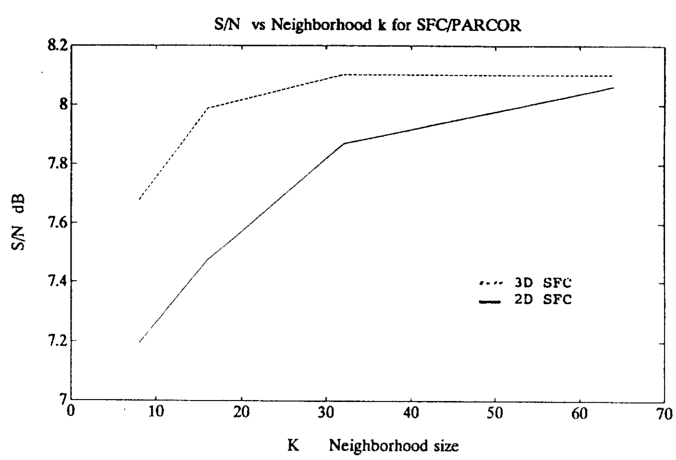

Fig.3 Ave. S/N for 2D & 3D SFC vs neighborhood size.