Fifth IFSA World Congress (1993), 1321-1325

The Optimum Fuzzy Vector Quantizer for Speech Synthesis

Jin-Rhee Lee, Hyung-Seuk Kim, Nam-kon Ko, and Kwang-Hyung Lee

Department of Electronic Engineering
Soong Sil University
Sang Do-1dong Dong Jak Ku,Seoul 156-743,Korea

Key Words : Speech Synthesis, Fuzzy Vector Quantizer, Optimal FuzzinessValue, Spectorgram, Pitch Frequency, FCM algorithim, Fuzzy-Neural ADPCM

Abstract

This paper investigates the use of Fuzzy vector quantizer(FVQ) in speech synthesis. To compress speech data, we employ K-means algorithm to design codebook and then FVQ technique is used to analysize input speech vectors based on the codebook in an analysis part. In FVQ synthesis part, analysis data vectors generated in FVQ analysis is used to synthesize the speech. We have fined that synthesized speech quality depends on Fuzziness values in FVQ, and the optimum fuzziness values maximized synthesized speech SQNR are related with variance values of input speech vectors. This approach is tested on a sentence, and we compare synthesized speech by a convensional VQ with synthesized speech by a FVQ with optimum Fuzziness values.

1. Introduction

Applications of fuzzy sets theory can be found in artificial intelligence, computer science, control engineering, decision theory, expert system, speech processing, logic, management science, operation research, pattern recognition, robotics, and others. Too theoretical advances have been made in many directions. More specially, it is well known that the use of fuzzy set theory in speech recoginition results in higher recognition rate than conventional methods.

In this paper, unlike by this time applications, we present speech synthesis using fuzzy set theory, and show that the proposed method is superior to conventional method in synthesized speech quality. Present speech synthesis methods can be divided into three types:

First, synthesis based on waveform coding, in which speech waveforms are sampled by Nyquist rate, and are encoded by PCM, ADPCM, APC, and other coding techniques to reduce requirements for memory size. Then encoded speech data are stored in computer memory. In order to synthesize speech, appropriate units are extracted in memory and connected them to produce speech waveforms.

This method is relatively simple, but have the problem of information reduction and synthesis units controllibility.

Secound, in synthesis based on the analysis-synthesis methods, words or phrased of human speech are analysized based on the speech producton model ,and stored as time sequence of feature parameters. Parameter sequences of approprite units are connected and supplied to speech synthesizer to produce the desired spoken message. LPC analysis methods and PARCOR methods are used for this purpose. This method is advantageous in that changing the speaking rate and smoothing the pitch and spectral change at connections can be performed by controlling the parameters.

So that, this methods have used in text-to-speech synthesis with small size speech units. Third, compound methods are combined waveform synthesis with analysis by synthesis methods. Recently the methods become popular in information reduction and speech quality. This methods can be classified into time domain methods and frequency domain methods. MPLPC (Multi pulse LPC), RELP (Residual Excited LPC vocoder), and CELP(Code Excited LPC vocoder) are typical in time domain, and frequency domain methods have ATC (Adaptive Transform coding), SBC (SubBand Coding).

As basic scalar quantization for data compression have its limits, the vector quantization technique which process vectors whose components are samples speech data, was proposed and extensively discussed by Linde, Buzo, and Gray[1]. Science 1980 this technique has been used in speech processing and image processing, as it is more excellent than traditional scalar quantization in data compression and its performance in information theory. Speech synthesis based on waveform codings require large memory size, so that we employ VQ technique to reduce this large memory. Unlike a vector quantizer that generates the index of a single codevector that best matches an input speech vector, a Fuzzy VQ that we apply to speech synthesis generates a vector whose components represent the degree to which each codevector matches input speech vector. The synthesis part reproduces synthesized speech vectors, using the output vector of analysis part and Fuzzzy C-Means algorithim. As a results, synthesized speech quality by FVQ-waveform coding depends on fuzziness value. In this study ,random search was done in need of optimum fuzziness maximizing SQNR

per frames, and we see that its fuzziness is related to the variance per frames in original speech.

Too this method can smooth at each boundary between adjacent synthesis units, so that synthesized speech is clearer to original speech than that using VQ in synthesized speech quality.

2. K-means algorithm

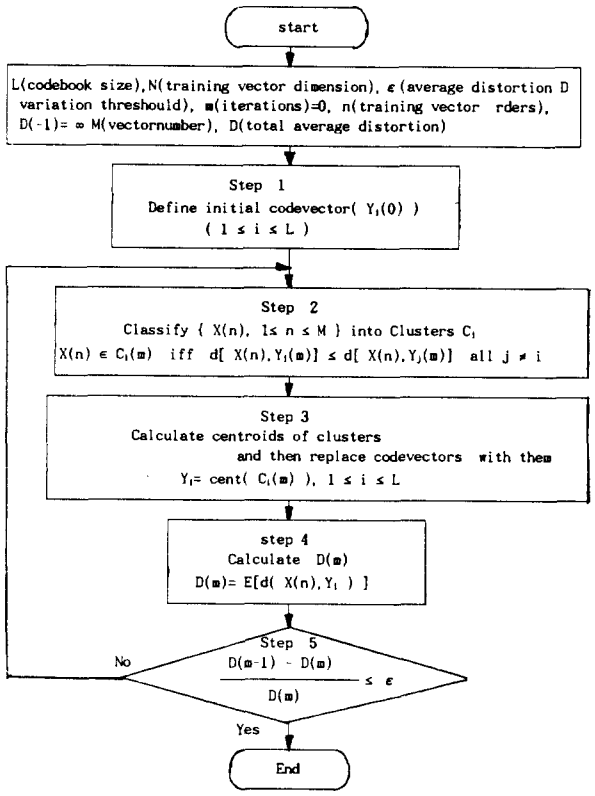


Fig.1. K-means algorithm

3. Fuzzy-VQ Speech analysis and synthesis

Unlike a vector quantizer that generates the index of a single codevector that best matches an input speech vector, a Fuzzy VQ that we apply to speech synthesis generates a vector whose components represent the degree to which each codevector matches input speech vector. The synthesis part reproduces synthesized speech vectors, using the output vector of analysis part and Fuzzzy C-Means algorithm.

3.1 FVQ speech analysis

In FVQ Speech analysis, Membership values which represent the degree to which each codevector matches the input vectors are defined according to the rule[2]:

$$m_{ij} = \left[\sum_{k=1}^{L} [\, d(X_i, Y_j) / d(X_i, Y_k) \,]^{\frac{1}{F-1}} \right]^{-1} \quad (1)$$

FVQ speech analysis part generates output vbector O_i:

$$O_i = (m_{i1}\ m_{i2}\ \cdots\ m_{iL}) \quad (2)$$

Note that the components of O_i are positive and sum to 1. Let $d(X_i, Y_j)$ represents Euclidean distance input vector X_i and codevector Y_j ,where is mean square error distance measure.

F(Fuzziness) > 1 is a constant called the degree of fuzziness vector O_i is chosen in this way because it minimizes the fuzzy objective function.

$$\sum_{i=1}^{T} \sum_{j=1}^{L} m_{ij}^{F}\, d(X_i, Y_j) \quad (3)$$

Notice that as F tends to infinity, each component of O_i tends to 1/L; as F tends to 1, then the component corresponding to mininmum value of $d(X_i, Y_j)$ tends to 1 and all other components tends to 0. Thus F can be selected to make the FVQ's decision very hard (F->1) or very soft(F->∞).

Hard decision corresponds to VQ decision. While this hard decision discards information about the degree to which incomming vector matches other codevectors, soft decision makes use of the information about degree.

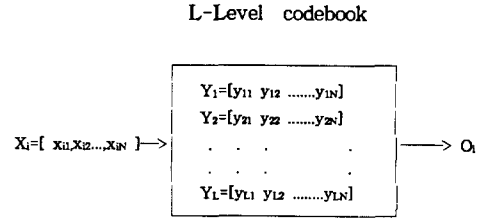


Fig.2. FVQ speech analysis

3.2 FVQ Speech Synthesis

Unlike pattern matching speech synthesis form by VQ, one new synthesis vector is obtained by using membership vector generated in FVQ analysis and FCM (Fuzzy C-means) algorithm.

$$X_{ij} = \frac{\sum m_{ij}^{F}\, Y_{ij}}{\sum m_{ij}^{F}} \quad (4)$$

where i is input speech vector number and j is codevector number in codebook.

$\hat{X}_i$ is synthesixed vector by FCM, and X_{ij} are components of $\hat{X}_i$.

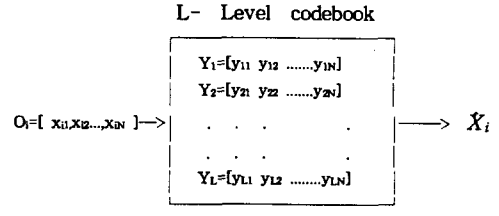


Fig.3. speech synthesis using FVQ

4. Experimental Results and Discussion.

We have experimented a VQ and FVQ speech analysis -synthesis on a sentence, "An young ha sim ni ka."(Fig 4). Used codebook size is 4 dimension 32 level, 64level. A sentence divided by 20 frames with 400 samples.

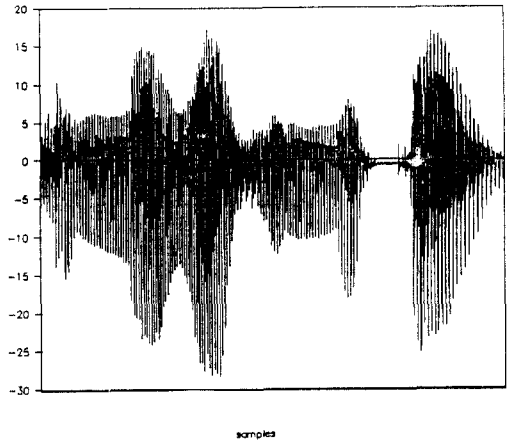
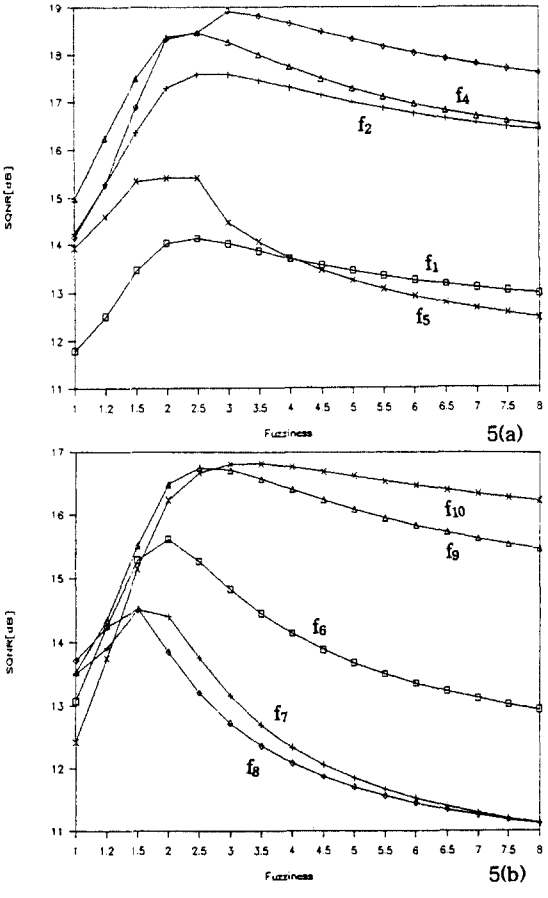


Fig.4. Original speech / An young ha shim ni ka /

Fig. 5(a),(b),(c),(d) show SQNR of synthesized speech verse Fuzziness values per frame orders.

In Fig.5 we show that the larger are variance values of frames, the smaller are fuzziness values maximizing SQNR. From this observation, it is deduced that in speech frames with small variance values, small fuzzines value is used to reduce correlation between codevectors, while in speech frames with large variance values, large fuzziness values are used to increase correlation between codevectors.



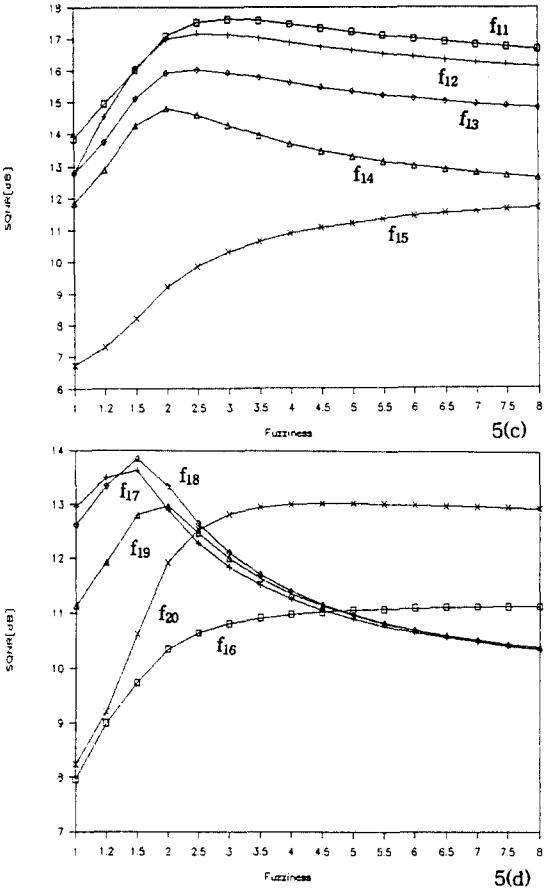


Fig.5. SQNR vs. Fuzziness values per frame orders
(a) frame 1 - 5 (b) frame 6 - 10
(c) frame 11 - 15 (d) frame 16 - 20

Fig.6 show optimum fuzziness values maximizing SQNR and variance values of frames.

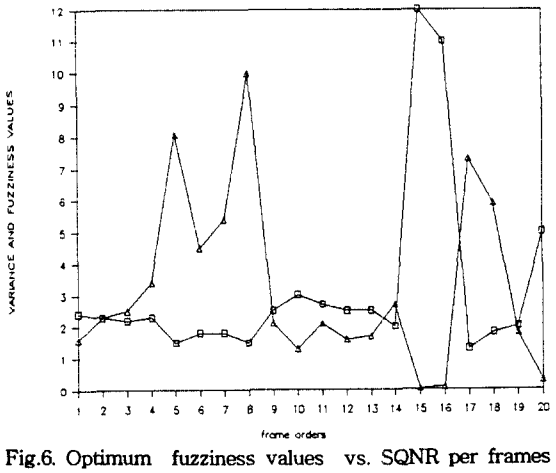


Fig.6. Optimum fuzziness values vs. SQNR per frames orders.

Fig.7 (a),(b),(c) show, respectively, synthesized speech waveform by 32 level VQ, 32 level FVQ with optimum fuzziness values, 64 level VQ.

Fig.8 compares SQNR of synthesized speech by 32 level VQ, 32 level FVQ with optimum fuzziness per frames, 64 level VQ.

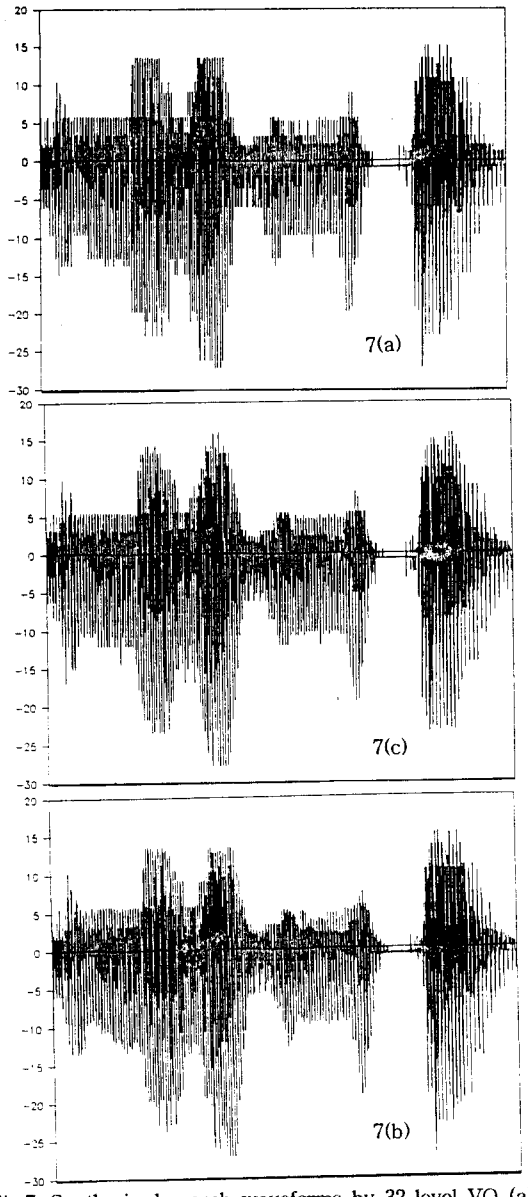


Fig.7. Synthesized speech waveforms by 32 level VQ (a), 32 level FVQ with optimum fuzziness values(b), 64 level VQ (c).

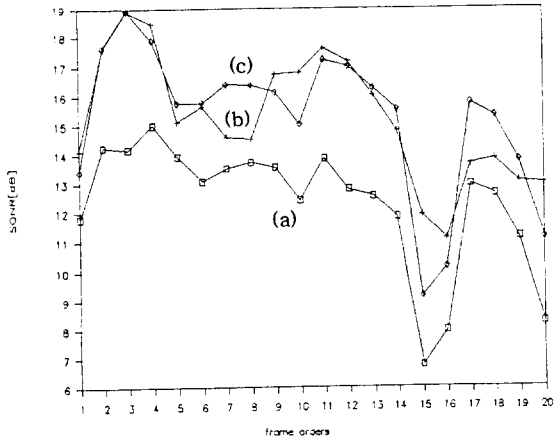


Fig.8. SQNR of synthesized speech by 32 level VQ(a), 32 level FVQ with optimum fuzziness per frames(b), and 64 level VQ(c).

This results can explain that FVQ speech synthesis performance is almost equal to its VQ, in spite of half a codebook size in FVQ.

Fig.9. indicate, rexpectively, spectrogram of original speech, synthesized by 32 level VQ, 32 level FVQ with optimum fuzzines per frames, and 64 level VQ.

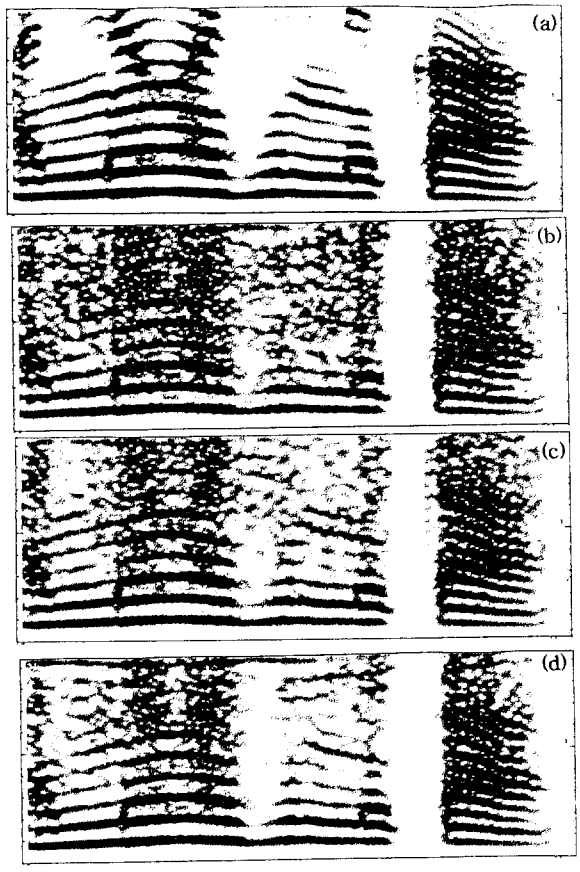
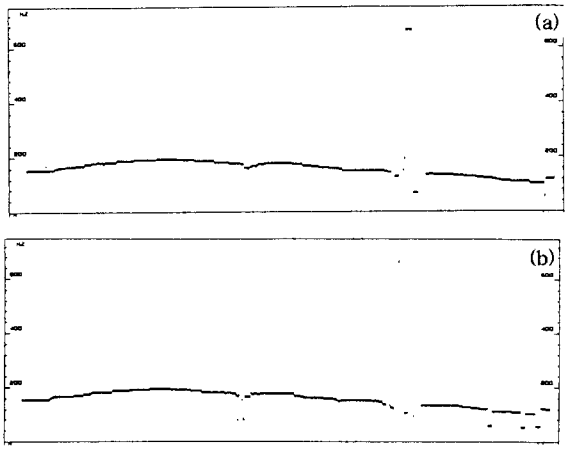


Fig.9 Spectrogram of original speech(a), synthesized speech by 32 level VQ(b), 32 level FVQ with optimum fuzzines values per frames(c),and 64 level VQ(d).

Fig.10. shows, respectively, pitch frequency of original speech(a), synthesized speech by 32 level VQ(b), 32 level FVQ with optimum fuzziness values(c), amd 64 level VQ(d).



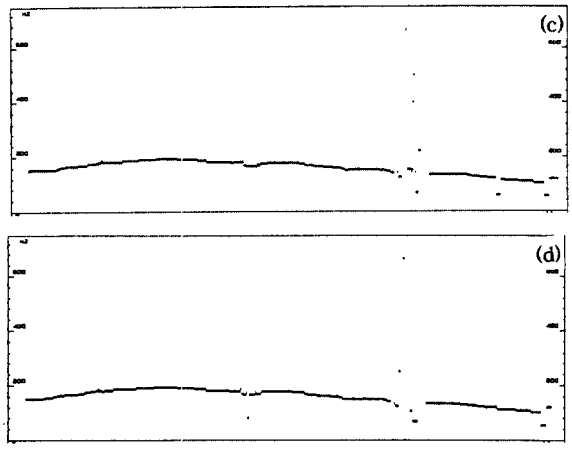


Fig.10. Pitch frequency of original speech (a), synthesized speech by 32 level VQ(b), 32 level FVQ with optimum fuzziness values(c), amd 64 level VQ(d).

5. Conclusion

We compare performance Fuzzy VQ and traditional nonfuzzy VQ in speech synthesis. From listening tests, we see that speech quality synthesized by FVQ is closer to original speech than VQ. What fuzziness value of FVQ is chosen is important in speech synthesis. Our results indicate that its value relates with the variance values of input speech vectors. Future work is about reducing membership values generated by using membership vector in FVQ. Too we are currently investigating the use of fuzzy clustering alorithm to design the coderbook used by FVQ. On the base of those results, we will complete Fuzzy Neural Nets-ADPCM speech coding.

[Reference]

[1] Y. Linde, A. Buzo, and R.Gray,"An algorithm for vector quantizer design," IEEE Tran.,commun.,vol. COM-28, pp. 84-95,Jan.1980.

[2] H. P. Tseng, M. J. Sabin,and E. A. Lee," Fuzzy vector quantization applied to hidden Markov modeling." Proc. ICASS P87,Paper 15.5.

[3] R.M.Gray, "Vector quantization," IEEE Assp Mag., vol.1, pp.4-29,April 1984.

[4] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding,"Proc.IEEE,vol.73,pp.1551-1588,Nov.1985.

[5] H.-J. Zimmermann(1991). Fuzzy set theory and its applications, second edition, Kluwer academic publisers.

[6] Stanley C, Ahalt, Ashok K, Krishnamurthy, Prakoon Chen,and Douglas E. Melton, " Competitive Learning Algorithms for Vector Quantization," Neural Networks 4,1990

[7] Lloyd Watts and Vladimir Cuperman," A vector ADPCM Analysis-by-Synthesis configuration for 16kbits/s speech coding," pp.9.2.1- 9.2.5, GLOBECOM, 1988

[8] N.Mohsenian and Nasser M.Nasrabadi," A Neural Net A pproach to Predictive Vector Quantization," pp.476-487, V isual Communications and Image Processing 1992.

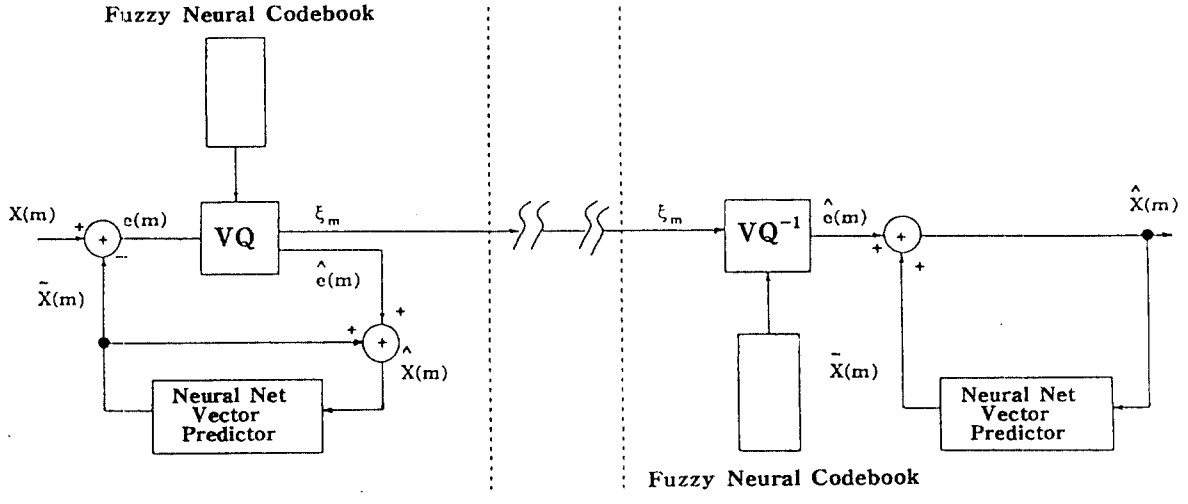
[9] J.H.Chen and A.Gersho," Vector Adaptive Predictive Coding of Speech at 9.6Kb/s," Proc. ICASSP,pp.1693-1696,Tokyo.

[10] V.Cuperman and A.Gersho," Vector Predictive Coding of Speech at 16 kbits/s", IEEE Tran.,comm.,vol.com-33, NO.7,July 1985

[11] Nasrabadi,N.M.,& Feng,Y." Vector quantization of images based upon the Kohonen self-organizing feature maps," IEEE International Conference on Neural Networks, pp 1101-1108.San Diego:IEEE

[12] Y.Shoham," Vector predictive quantization of the spectral parameters for low rate speech coding," pp.51.2.1-51.2.4, ICASSP, 1987.

[13] P.Cummiskey,N.S.Jayant and J.L.Flanagan," Adaptive Quantization in Differential PCM Coding of speech," Bell System Technical Journal, pp.1105-118, September 1973.



Fuzzy Neural ADPCM (FN-ADPCM) Coding System