Efficiency of Ordered Codebook Learning Vector Quantization for Speech Compression

Kreangsak Pattanaburi and Jakkree Srinonchat

***Abstract*—Combined compression and classification problems are becoming increasingly important in many applications with large amounts of data and large sets of classes. This article presents the efficiency of ordered codebook learning vector quantization (OC-LVQ) for speech compression. The algorithm is based on competitive networks. It is developed and analyzed a learning vector quantization based algorithm for combined speech compression and classification. The Peak Signal to Noise Ratio (PSNR), Signal to Noise Ratio (SNR), and Normalized Root Mean Square Error (NEMSE) are used to measure the quality of speech signal. It provides the maximum quality at 28.9432 dB and 15.0333 dB for SNR and PSNR respectively. Also the minimum error of NEMSE is 0.1578.**

***Index Terms*—Speech compression, ordered codebook, learning vector quantization.**

I. Introduction

The requirements of a speech compression signal have been sought in mainly speech coding research centers. As a result many different strategies for the suitable speech compression applications have been developed. The exploitation of bit rate speech coders have been standardized in many international and national communication systems [1].

In speech signal processing, the amount of data analyzed require a long time process. To process audio faster, speech coding or speech compression is to reduce size of the speech signal input [2], [3]. It is essentially technique for communication system which obviously uses in many researches such as, presents [4] the Kmean-LBG algorithm and KSOFM algorithm, which were investigated to use in speech coding system. The experimental results show the comparison of the performance of ordered and disordered codebooks which employ to measure and classify the repetition of the speech signal coefficients. Both ordered and disordered codebooks can reduce the number of bit rate transmission around 20% in speech coding system. Also presents [5] a good quality speech data at a low bit rate. In order to accomplish this, the most powerful speech analysis and compression techniques such as Linear Predictive Coding (LPC), Discrete Cosine Transformation (DCT) and Discrete Wavelet Transformation (DWT) are adopted for Tamil speech database. The adopted techniques are evaluated based on Compression ratio, Peak Signal to Noise Ratio (PSNR).

Normalized Root Mean Square Error (NRMSE) and Mean Opinion Score (MOS). The results show that the DWT achieves greater performance than other two techniques employed in this research.

This article presents an exploitation of Learning Vector Quantization (LVQ), in ordered codebook for speech compression. The adopted techniques are evaluated based on Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE). It is organized as follows. Section II describes the linear predictive coefficients while Section III details the LVQ neural network. Section IV details experiment while V shows its simulations results. Finally, Section VI concludes this work.

Manuscript received June 29, 2012; revised August 21, 2012.

The authors are with the Department of Electronic and Telecommunication, Rajamangala University of Technology Thanyaburi, Thailand (e-mail: kreangsak_p@hotmail.com, jakkree.s@hotmail.com).

II. Linear Predictive Coefficients

The basic idea of the linear prediction parameters is that the next sample speech signals can be predicted by a linear combination of the past values of the sample signal at time (n).This is shown in the following equation:

$$S_n = e_n + \sum_{k=1}^{p} a_k S_{(n-k)} \tag{1}$$

where S_n the value of sample is signal at time (n)

a_k is the predictor parameters

e_n is the prediction error

From the (1), it can be defined that

$$\hat{S}_n = \sum_{k=1}^{p} a_k S_{(n-k)} \tag{2}$$

Taking z-transforms gives

$$S(z) = E(z) + \left[\sum_{k=1}^{p} a_k z^{-k}\right] S(z) \tag{3}$$

$$S(z) = \frac{E(z)}{\left(1 - \sum_{k=1}^{p} a_z z^{-k}\right)} + E(z).H(z) \tag{4}$$

where $S(z)$ and $E(z)$ are the z transform of $S_{(n)}$ and $e_{(n)}$ respectively. Thus $H(z)$ can be defined as

$$H(z) = \frac{1}{\left(1 - \sum_{k=1}^{p} a_z z^{-k}\right)} \tag{5}$$

Which $H(z)$ the transfer function of a digital filter is as refers to all-pole system. Thus (5) can be rewrite as

$$H(z) = \frac{1}{A(z)} = \frac{1}{\left(1 - \sum_{k=1}^{p} a_z z^{-k}\right)} \tag{6}$$

DOI: 10.7763/IJIEE.2012.V2.236

However, a general transfer function of a real vocal has both poles and zeros.

III. LVQ Neural Network

Learning vector quantization (LVQ) was developed by Kohonen network. LVQ network structure is different from unsupervised training structure. LVQ algorithm is a learning algorithm to train the Kohonen layer with teacher's guide. The architecture of the LVQ network used in this paper is shown in Fig. 1.when x is input network. Output neurons are the nearest to x. It is selected as the "winning neuron".

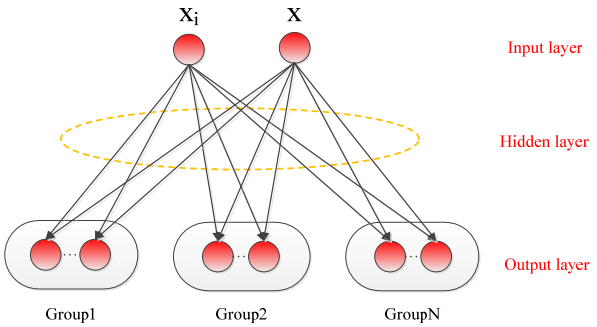

Fig. 1. Structure of LVQ network

The proposed learning algorithm of the LVQ networks is as follow:

Prepare the training data. In the input layer, there are C neurons. The continuous are input vectors are

$$X = (x_1, x_2 ..., x_C) \tag{7}$$

Connection weights vectors between input layer and Kohonen layer are

$$W^1 = (w_1^1, w_2^1, ..., w_D^1) w_i^1 = (w_{i1}^1, w_{i2}^1, ..., w_C) \tag{8}$$

where $i = 1, 2, ..., D$.

Connection weights vectors between Kohonen layer and output layer are

$$W^2 = (w_1^2, w_2^2, ..., w_k^2) w_k^2 = (w_{k1}^2, w_{k2}^2, ..., w_{kC}^2) \tag{9}$$

where $k = 1, 2, ..., E$.

Every Kohonen neuron is assigned to an output neuron and corresponding connection weights vector is 1 and other connection weights vector is 0. W^2 is fixed during training process.

$$W_{kr}^2 = \begin{cases} 1 & r \in k \\ 0 & r \in k \end{cases} \tag{10}$$

Suppose training mode as follow

$$\{x_1, t_1\}, \{x_2, t_2\}, ..., \{x_F, t_F\} \tag{11}$$

where 1j (j = 1, 2, ···, Q) is object output vector. The output of Kohonen layer is calculated as follow

$$V = W^1 X \tag{12}$$

Then the output vector is

$$T = W^2 V \tag{13}$$

W^1 Can be determined as follow:

For every input vector, the network will give aclassification result. If the result of classification is correct, the connection weights value can be corrected by (13).

$$i^{*w^1}(t+1) = i^{*w^1}(t) + \eta(p(t+1) - i^{*w^1}(t)) \tag{14}$$

If the result of classification is false, the connection weights value can be correct by (14)

$$i^{*w^1}(t+1) = i^{*w^1}(t) - \eta(p(t+1) - i^{*w^1}(t)) \tag{15}$$

where $\eta \in (0,1), i^{*w^1}(t)$ is the connection weights value of the i^*th neurons at t time.

Repeat this step until the achieved classification rate is satisfied or the maximum number of epochs is reached as show in Fig. 2.

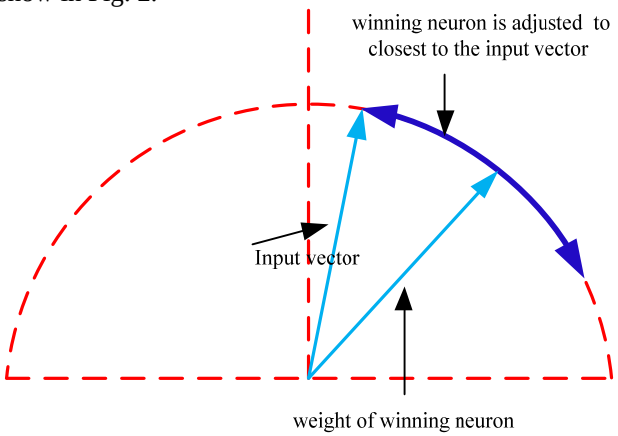

Fig. 2. Learning of LVQ neural network

After learning, the LVQ network can serve to recognize the unknown gas data [6].

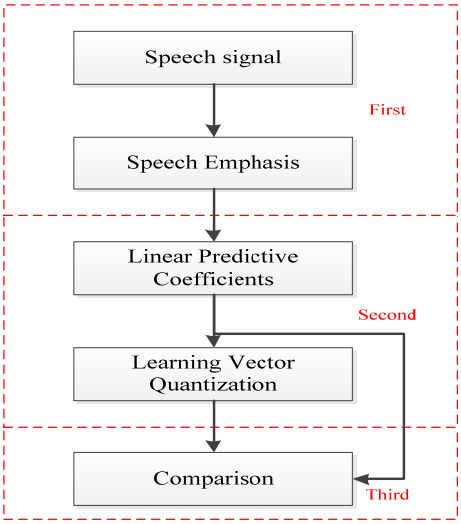

Fig. 3. Experiment processes

IV. Experiment

There are 20 speech input data which are generated from males and females. The speech signal is sampled at 8 kHz and the frame size is 200 sampling per second. This process

research can be classified into three steps as show in Fig. 3.

Firstly speech signal is passing through the speech emphasis technique to filter the back ground noise. Then the speech signals were calculated the LPC coefficients, where only 10 LPC coefficients represented a speech frame. LPC coefficients were generated in order to compare the effectiveness in the ordered codebook.

Secondly, the LPC coefficients are calculated to become the new LPC coefficients, namely codebooks, using the technique of standard deviation technique of V/UV classification [7]. The Learning vector quantization (LVQ) is used to classify LPC coefficients into groups for each particular speaker which each address of the codebook contained a set of code vectors, which represents each group. The size of the codebook has an effect on the bit rate which represents the speech coefficients. This experiment uses the sizes of codebooks ranging from 1024 to 64 addressed.

Finally, the comparison on of the performance between the different ordered codebook using PSNR, SNR and NRMSE.

A. Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\frac{NP^2}{\|p-r\|^2} \tag{16}$$

where N is the length of the reconstructed signal.

P is the maximum absolute square value of the signal p.

$\|p-r\|^2$ is the energy of the difference between. the original and reconstructed signals.

B. Signal to Noise Ratio (SNR)

$$SNR = 10\log_{10}\left|\frac{\sigma_x^2}{\sigma_e^2}\right| \tag{17}$$

where σ_x^2 is the mean square of speech signal and

σ_e^2 is mean square difference between the original and reconstructed signal [4].

C. Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \sqrt{\frac{(p(n)-r(n))^2}{(p(n)-\mu p(n))^2}} \tag{18}$$

where $p(x)$ is the speech signal.

$r(n)$ is reconstructed signal

$\mu p(n)$ is mean of the speech signal.

V. Results

The results show the comparison of different ordered performance codebook based on the frequency domain, PSNR, SNR and NRMSE.

In Fig. 4, the 1024 address ordered codebook provides the best performer of male when is compared to all of speech signal in term of the loss frequency domain. It can be notice noise in high of 512,256,128 and 64 address ordered codebook.

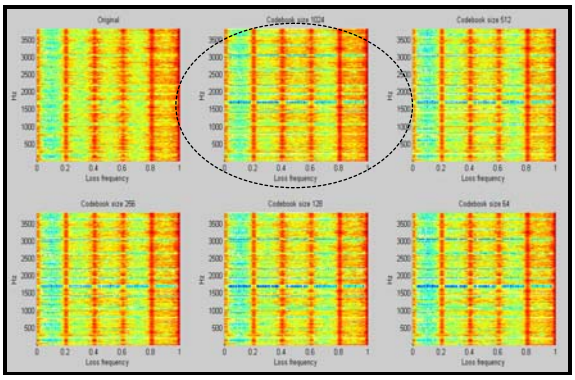

Fig. 4. Comparison of order based on the loss frequency (Male)

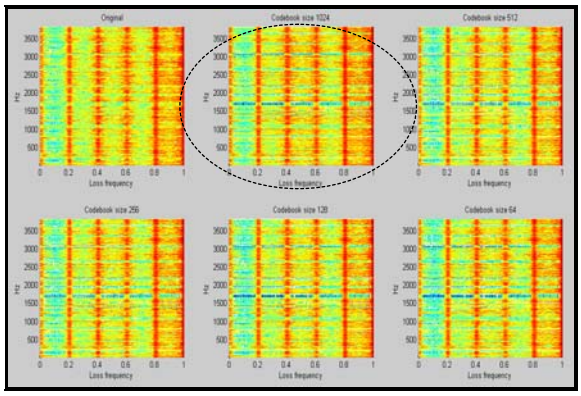

Fig. 5. Comparison of order based on the loss frequency (Female)

In Fig. 5, the 1024 address ordered codebook provides the best performer of female when is compared to all of speech signal in term of the loss frequency domain. It can be notice noise in high of 512,256,128 and 64 addresses ordered codebook.

TABLE I: Comparison of Order Based on PSNR

Peak signal to noise ratio (dB)					
Speech	Codebook size				
	1024	512	256	128	64
Male 1	28.9432	28.8432	28.5604	28.4131	28.2671
Male 2	24.3278	24.3280	24.2865	24.2376	24.1985
Female 1	28.1904	28.1891	28.1448	28.1385	28.0505
Female 2	27.8795	27.6994	27.7460	27.5003	27.3852

In the Table I, the 1024 address ordered codebook provides the best performer at male1 when is compared to all of speech signal in term of PSNR.

TABLE II: Comparison of Order Based on PSNR

Signal to noise ratio (dB)					
Speech	Codebook size				
	1024	512	256	128	64
Male 1	14.1268	14.1263	14.1263	14.0962	13.9502
Male 2	13.0010	13.0012	12.9365	12.8776	12.7829
Female 1	12.8215	12.8189	12.8070	12.7461	12.7093
Female 2	15.0333	14.8532	14.8734	14.8195	14.6889

In the Table II, the 1024 address ordered codebook provides the best performer at female2 when is compared to all of speech signal in term of SNR

TABLE III: COMPARISON OF ORDER BASED ON NRMSE

Normalized Root Mean Squared Error					
Speech	Codebook size				
	1024	512	256	128	64
Male 1	0.2000	0.2000	0.2001	0.2007	0.2041
Male 2	0.1931	0.1954	0.1945	0.1953	0.1979
Female 1	0.1952	0.1954	0.1956	0.1970	0.1979
Female 2	0.1578	0.1611	0.1609	0.1616	0.1626

In the Table III, the 1024 address ordered codebook provides the best performer at female2 when is compared to all of speech signal in term of NRMSE

VI. CONCLUSION

This article presents the efficiency of ordered codebook learning vector quantization for speech compression. In term of the speech quality using PSNR, the large ordered codebook size provides best performer than the small ordered codebook size because of large number of codebooks. It provides the maximum quality at 28.9432 dB and the minimum quality at 24.1985 dB as shown in Table I. In the term of the speech quality using SNR, the large ordered codebook size also provides better performer than the small ordered codebook size. This article provides the maximum quality at 15.0333 dB and the minimum quality at 12.7093 dB as shown in Table II. In the term of the speech quality using NRMSE, the large codebook size was better than the small codebook size because it will have a minimum error value. The article provides the maximum error at 0.2041 and the minimum error at 0.1578 as shown in Table III.

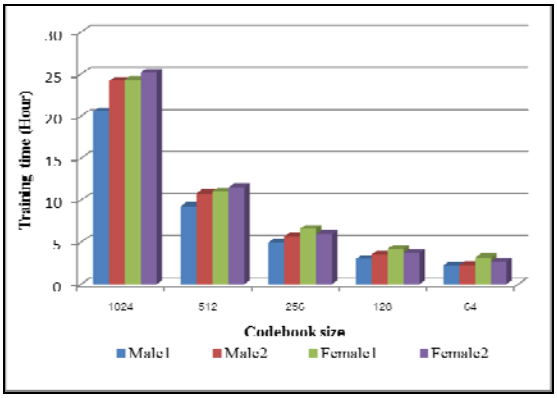

Fig. 6.Comparision of order based on training time

The experiments of using ordered codebook LVQ for low bit rate speech compression show that 1024 address ordered codebook is the best performer but it requires large amount of bit rate to storage and transmission data (10 bits). Also it needs more time to train the winning neuron as in shows in Fig. 6. Therefore this experiment can apply to those speech compression and data storage.

ACKNOWLEDGMENT

We would like to express our gratitude to all staff of Signal Processing Research Laboratory, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Thailand, for giving us the speech signal. We have furthermore to thank Prof. Dr. Sean Danaher from the University of Northumbria at Newcastle whose help stimulating suggestions. Also we would like to thank the Office of National Research Council of Thailand for financial support (2012) in this research.

REFERENCES

[1] W. T. K. Wong, et.al, "Low rate speech coding for telecommunications," *BT Tehnology*, vol. 14, 1, pp. 28-43, 1996.
[2] J. Srinonchat, "Comparison of the efficiency of ordered and disordered codebook techniques in speech coding," *IEEE-International Conference on Information and Communication Systems*, pp. 195 - 198, 2005.
[3] J. Srinonchat, "Enhancement arificial neural networks for low-bit rate speech compression system," *IEEE- International Symposium on Communications and Information Technologies*, pp. 195-198, 2006.
[4] Weerayuth khunrattanasiri and Jakkree srinonchat, "Comparison efficiency of speech compression using wavelet technique," *JICTEE-2010*, pp. 242-246, 2010.
[5] V. Radha, C. Vimala, and M. Krishnaveni, "Comparation analysis of compression techniques for Tamil speech datasets," *IEEE-International conferent on trends in Information technology*, pp. 712-716, 2011.
[6] J. Liu, Y. Liang, and X. Sun, "Application of learning quantization network in faule diagnosis of power transformer," *IEEE-International conference on mechatronics and automation*, pp. 4435-4439, 2009.
[7] K. Pattanaburi and J. Srinonchat, "Enhancement pattern analysis technique for voiced/unvoiced classification," *IEEE-International symposium on computer, consumer and control*, 2012.

Kreangsak Pattanaburi received the bachelor's degree in electronic and telecommunication engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 2010 where he is currently studying toward the master's degree in the Department of Electrical Engineering, RMUTT.

Jakkree Srinonchat received bachelor's degree in electronic and telecommunication engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 1995, and his Ph.D. in Electrical Engineering, major signal processing from University of Northumbria at Newcastle, UK, in 2005. He is currently a lecturer of Department of Electronics and Telecommunication Engineering, Faculty of Engineering, RMUTT, Thailand. His research is focus on the signal processing, especially FPGA Design, speech and image processing. He is currently the advisor of the Signal Processing Research Laboratory, which establishes to provide and services the new design and solution to industry.