

ENTROPY CODED VECTOR QUANTIZATION WITH HIDDEN MARKOV MODELS

Tadashi Yonezaki[†] and Kiyohiro Shikano[‡]

[†] Telecom Research Lab., Matsushita Communication Ind.Co.,Ltd.
600 Saedo-cho,Tsuzuki-ku,Yokohama,224,JAPAN

[‡] Graduate School of Informatoin Science,
Nara Institute of Science and Technology,
8916-5 Takayama-cho,Ikoma-shi,Nara,630-01,JAPAN

ABSTRACT

We propose a new vector quantization approach, which consists of Hidden Markov Models(HMMs) and entropy coding scheme. The entropy coding system is determined depending on the speech status modeled by HMMs, so the proposing approach can adaptively allocate suitable numbers of bits to the codewords. This approach realizes about 0.3[dB] coding gain in cepstrum distance(8 states HMMs). In other words, 8 bit-codebook is represented by about 6.5 bits for average code length. We also research for robustness to the channel error. HMMs and the entropy coding system, which seem to be weak to the channel error, are augmented to be robust, so that the influence of the channel error is decreased into one-third.

1. INTRODUCTION

Finite state vector quantization(FSVQ), which describes a source vector sequence as Markov model, has been studied in speech coding for its potential to efficient bit-rate reduction. Although many kinds of applications with FSVQ have been proposed, such as speech coding system[1], image coding system[2][3], it seldom uses in practical system because of its sensitivity to channel errors.

On the other hand, entropy coding, which is most efficient one as lossless coding, is also expected to be utilized in speech coding. Entropy-constrained vector quantization(ECVQ)[4], whose vector quantizer is designed for having minimum distortion subject to an entropy constraint, was proposed for waveform coding. It shows one example for increasing coding gain by applying entropy coding scheme to conventional vector quantization.

Futhermore, Chou and Lookabough modified ECVQ to incorporate some inter-block information like sequence of vectors[5]. It is same as FSVQ with entropy coded state codebooks and realizes efficient bit rate reduction. This result shows vector quantizer with combining of Markov model and entropy coding could realize a good performance.

In this paper, we use a hidden Markov model(HMM) as

model of source sequence. The HMM is popular in automatic speech recognition because of the simple and effective time sequence representation of speech spectrum features. Thus the HMM could be superior to Markov model, that is it could realize highly efficient speech coding system. Sections 2 and 3 provide details of a novel method for vector quantization and its performance. In section 4, we argue robust structure against channel errors.

2. ENTROPY CODING WITH HMMS

A hidden Markov models(HMMs) is a stochastic model with two kinds of probability, which are state transition probability and observation probability. In the Markov model, a speech status is described with only one state of the model. On the contrary, the HMMs describes a speech status with distribution of probabilities over the states (state distribution). Figure 1 shows the state transition of source models. The probabilities can take arbitrary value, thus the HMMs can express an extraordinary number of speech status with limited number of states.

At discrete HMMs, the features of an observed sequence are vector-quantized with a codebook, and the occurrence probability of each codevector are calculated by summing up over all the HMM states weighted with state distribution. From the other point of view, every codevector has an HMM predicted occurrence probability at each time. Figure 2 shows transition of output probability of codevectors. For speech coding, these HMM predicted occurrence probabilities for codevectors are used in order to set up codes by an entropy coding system such as Huffman coding.

Figure 3 shows encoder and decoder block diagrams of an entropy coded VQ with HMMs, where Huffman coding is used as an entropy coding. LPC spectrum envelopes derived from input data are encoded by Huffman coding. The codes for codevectors are updated adaptively according to the HMM state existence probabilities and the occurrence probabilities of codevectors in HMM states, as follows;

1. Calculate the existence probability of each state.

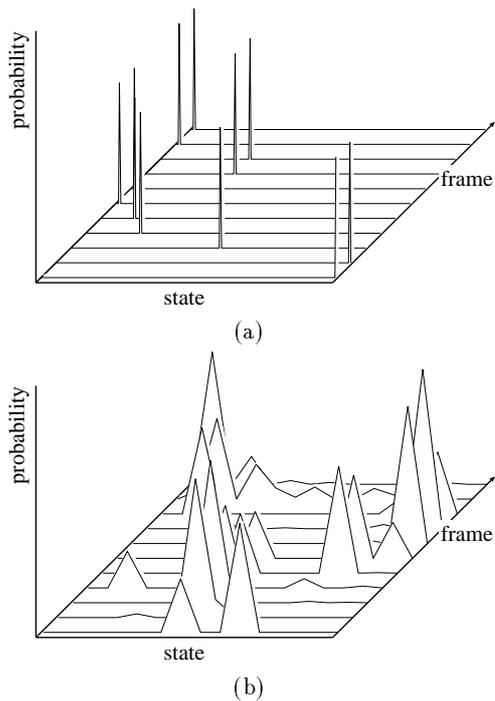


Figure 1: State transition of source models — (a) Markov Models, (b) Hidden Markov Models

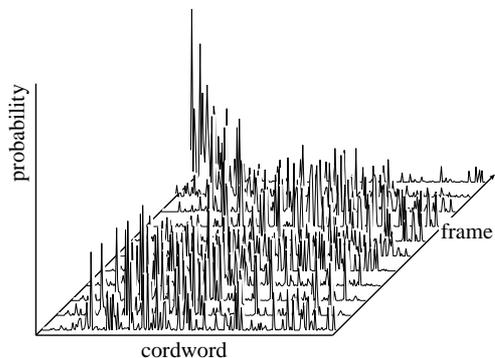


Figure 2: Transition of output probability

2. Calculate the occurrence probability of each codevector.
3. Update Huffman codes of the codevectors according to HMM predicted probabilities, which are calculated based on the state existence probabilities and the occurrence probabilities.

2.1. State Existence Probabilities

The state existence probabilities at time n are calculated from state existence probabilities at time $n - 1$ and the observation at time $n - 1$. When x_{n-1} is an observed codevector

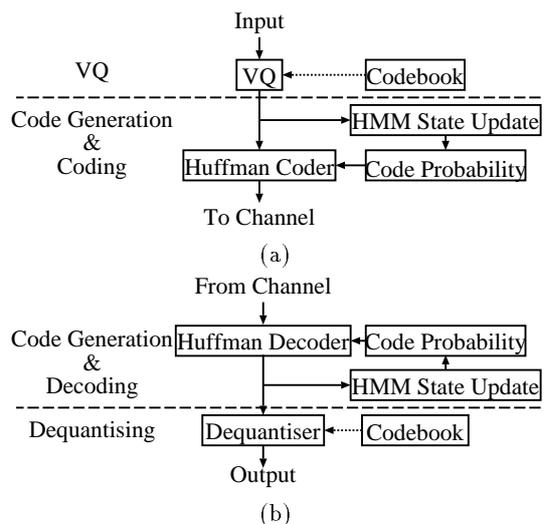


Figure 3: Block diagram of HMM-VQ — (a) coder, (b) decoder

at time $n - 1$, the existence probability for state s at time n , $P_{exist}(s; n)$, is

$$P_{exist}(s; n) = \sum_{i=1}^{N_s} P_{exist}(i; n-1) a_{is} b_{is}(x_{n-1}), \quad (1)$$

where N_s is number of HMM states, a_{is} is a transition probability from state i to s , and $b_{is}(x_{n-1})$ is an observation probability of x_{n-1} on the transition from state i to s .

2.2. Codevector Occurrence Probabilities

The occurrence probability $P_{code}(C_k)$ of a codevector C_k in a codevectors is calculated as follows,

$$P_{code}(C_k) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} P_{exist}(i; n) a_{ij} b_{ij}(C_k). \quad (2)$$

2.3. Updating Huffman codes

Using $P_{code}(C_k)$ for $k = 1, \dots, N_c$, Huffman codes of a codebook are determined, where N_c is codebook size. This Huffman code determination is carried out independently in the encoding and the decoding using the same HMMs.

3. EXPERIMENTAL RESULTS

We apply a new HMM-based vector quantization scheme (HMM-VQ) to LPC spectrum envelope coding. We use 1,2,4,8 and 16 state ergodic discrete HMMs, which are trained with 44 speakers (6639 utterances). The codebook includes 15th-LPC cepstrum coefficients, which are extracted

from input frames as spectrum envelope features, as codevector.

The proposed scheme is evaluated from two points of view, these are bit reduction rate and distortion–rate function. FSVQ with entropy coded state codebooks[5] is also evaluated as reference. The next-state function is designed by conditional histogram, thus FSVQ has label-state structure. In our experiments, all state-codebooks consist of the same codevectors, but the different occurrence probabilities.

3.1. Bit Reduction Rate

Figure 4 shows the results of bit reduction rate to the number of HMMs state. The codebook includes 256 codevectors, that is conventional VQ, whose average bit rate corresponds to the one at the 0-state HMM (without HMMs) in Fig.4, needs 8 bits for each code. In the 16-state ergodic HMMs,

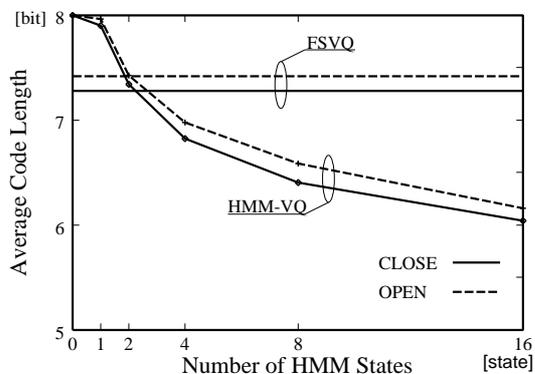


Figure 4: Bit reduction for number of HMMs state

our Huffman coded VQ with HMMs can reduce the average bit rate from 8 bits to about 6.1 bits. 2-state-HMM realises same bit reduction rate of FSVQ. 1-state HMM means conventional VQ followed by entropy coding, whose codes are determined by histogram of training data. It means the proposed scheme can adaptively allocate suitable code length depending on the characteristic modeled by HMMs.

3.2. Distortion–Rate Function

Figure 5 shows relation between distortion and bit rate. In this experiment, the system consist of 8-state HMMs. Cepstrum distance is used as distortion measure. The proposed scheme achieves coding gain of about 0.3 dB against conventional VQ.

In our experiments, FSVQ consists of the same number of states as codevectors. In fig. 5, performance of FSVQ is improved with increasing bit rate (number of codevectors) until around 9.2 bits, but depressed over that rate. At designing of the next-state function for FSVQ, conditional histogram design algorithm needs histogram of transition between certain states. Therefore, required number of training data increases

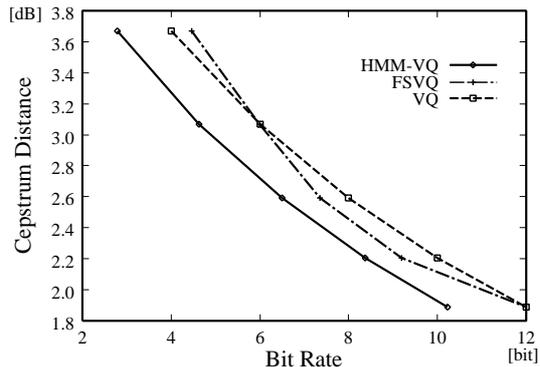


Figure 5: Distortion vs. bit rate for 8-state HMMs

exponentially according to the increasing number of states. As the result, depressing performance of FSVQ at high bit rate comes from insufficient number of training data. Consequently, the proposed scheme is superior to FSVQ in not only coding gain but training efficiency.

4. ROBUST STRUCTURE AGAINST CHANNEL ERRORS

The proposed VQ structure consists of recursive quantization and entropy coding. These are known with its sensitiveness to channel errors. In this section, We argue the robustness against transmitting channel errors. The response against one bit transmission error is depicted in Figure 6. The mea-

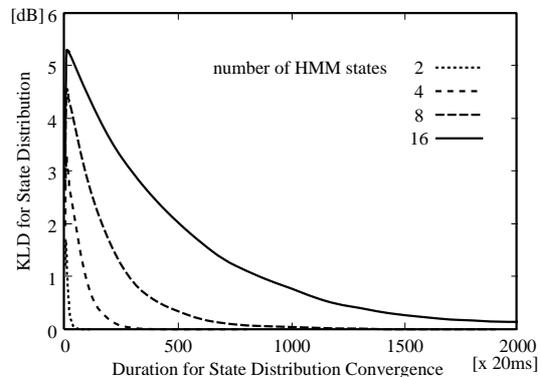


Figure 6: State convergence response against a channel error

sure for convergence is Kullback-Leibler Divergence of state distribution between coding and decoding system. Although it takes long time, Fig. 6 verifies the system's convergence property which comes from ergodicity of HMMs. This long convergence time is almost due to inconsistent of the generated Huffman codes at encoder and decoder. Avoiding this and shortening the convergence time, we propose two kinds of constraints. One for the Huffman codebooks for the transmission and the other for the distribution of HMM

state probabilities. In the first constraint, available Huffman codes are restricted in advance. During the codec process, the Huffman codebook which is used for the transmission of speech information on a frame is selected by a mean of minimizing average code length from predetermined Huffman codes by training data. In the second constraint, the distribution of state probabilities are vector-quantized.

Comparison of FSVQ and the proposed methods on convergence time and average code length is depicted on Figure 7. In Fig. 7 duration for convergence is defined by the time

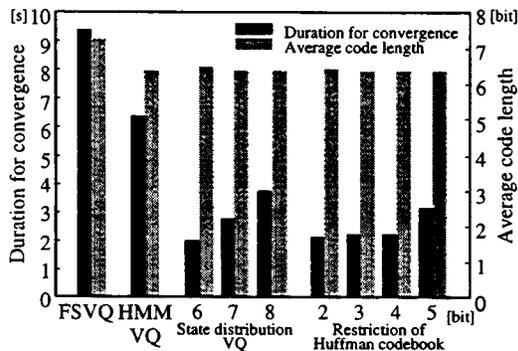


Figure 7: Comparison of convergence time and average code length for several coding structures

when the cepstrum distance between coder and decoder decreases below 1 dB. Convergence time for FSVQ and the proposed scheme are about 9.4s and 6.4s respectively. Furthermore, restricting the available Huffman code or the distribution of state probabilities shortens the convergence time to about 2.2s. Fig. 7 shows the restrictions of Huffman codebooks and state distributions have little effect on average code length but realize great improvement in duration for convergence.

5. CONCLUSION

This paper proposes a new vector quantization scheme, which consists of HMMs and entropy coding scheme. This approach realizes about 0.3[dB] coding gain in cepstrum distance (8 states HMMs). This scheme is superior to FSVQ in coding efficiency and robustness to channel errors. Furthermore, the proposed scheme can be augmented to be robust, so that the influence of the channel error is decreased from 6.4s to 2.2s.

6. REFERENCES

1. Foster, J., Gray, R. M., and Dunham, M. O. "Finite-state Vector Quantization for Waveform Coding," *IEEE Trans. IT* 31: 349-359, May, 1985.
2. Aravind, R., and Gersho, A. "Low-rate Image Coding with Finite-state Vector Quantization," *ICASSP*: 137-140, 1986.
3. Kim, T. "New Finite State Vector Quantizers for Images," *ICASSP*: 1180-1183, 1988.
4. Chou, P. A., Lookabaugh, T., and Gray, R. M. "Entropy-constrained Vector Quantization," *IEEE Trans. ASSP* 37: 31-42, January, 1989.
5. Chou, P. A., and Lookabaugh, T. "Conditional Entropy-Constrained Vector Quantization of Linear Predictive Coefficients," *ICASSP*: 197-200, 1990