# Using Dynamic Codebook Re-ordering to Exploit Inter-frame Correlation in MELP Coders

*Venkatesh Krishnan, Thomas P. Barnwell III, and David V. Anderson*

Center for Signal and Image Processing
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
United States of America

## Abstract

Model based speech coders such as the mixed–excitation linear prediction (MELP) coder encode parameters of the autoregressive model for short-duration frames of the speech signal. Typically, parameters extracted from successive frames by the MELP coder exhibit strong correlation. Reduction in the transmitted data-rates can be achieved if the encoders for these parameters effectively exploit this inter-frame correlation. In this paper, we apply a procedure, called dynamic codebook re-ordering (DCR) to reduce the entropy in the distribution of the symbols generated by the vector quantization encoders used in coding the MELP parameters. The entropy reduction is achieved by exploiting the correlation between the vectors of MELP parameters derived from successive speech frames. The advantages of the DCR procedure over other techniques that exploit inter-frame correlation stem from the fact that it significantly reduces the data-rates without introducing any additional coding delays or increasing the distortion and it is simple and elegant.

## 1. Introduction

Low bitrate speech coders such as the Mixed Excitation Linear Prediction (MELP) vocoder [1] [2] [3] represent short duration segments of a speech signal by a source–autoregressive system model and encode the parameters of this model using suitable source coding techniques such as vector quantization (VQ) [4]. It is well known that parameters of the MELP derived from consecutive segments of a speech signal are strongly correlated. Traditionally, several techniques including structured vector quantization (VQ) [4], super-frame VQ [5] [6] and VQ with memory [4] have been proposed for reduced bit-rate coding of the speech coder parameters. However, many of these techniques render the coding process sub-optimal, or require buffering and thus introduce coding delay, or they introduce distortions in the reconstructed vector.

In this paper, we employ a simple and elegant procedure, called dynamic codebook re-ordering (DCR) [7] [8] [9], to effectively exploit the correlation between consecutive source vectors without introducing any delay, distortion or sub-optimality to the standard VQ system. In the DCR procedure, the codebook of the standard VQ is re-ordered for every encoded symbol based on a suitably chosen dissimilarity measure. The dissimilarity measure chosen for this re-ordering procedure depends only on the codevectors selected in the past. Therefore, the DCR procedure can be replicated at the receiver without requiring any additional transfer of information. Since the proposed DCR procedure does not introduce any sub-optimality to the VQ system, it is capable of achieving a significantly better rate-distortion performance than the standard VQ. Further, no buffering of past source vectors is required. Therefore, incorporation of the DCR in the encoder and decoder of a standard VQ system introduces no coding delays.

In this paper, the DCR procedure is incorporated in the encoding and the decoding algorithms employed in the MELP vocoder. We describe the application of the DCR procedure to structurally constrained VQ systems such as multistage VQ and split VQ that are often employed to encode the linear prediction (LP) model parameters used by speech coders. Also, the incorporation of the DCR procedure in encoding the other MELP coder parameters, including the pitch, gain, bandpass voicing flags and the Fourier magnitudes are described. Results are presented to demonstrate a reduction of approximately 40% in the entropy of the symbols of the encoders that are used to encode the parameters of the MELP vocoder.

In the following section, the DCR algorithm is described. In Section 3, the DCR procedure is incorporated in the encoders and decoders of the MELP parameters and empirical entropy measures are provided. Conclusions are presented in Section 4

## 2. Dynamic codebook re-ordering

In this section, the DCR algorithm is presented. Let the source vector $\mathbf{x}$ be $N$-dimensional and belong to a space $\mathbf{V}$ and let $\zeta$ be a set of digital symbols. The VQ encoder can be thought of as a mapping $\Gamma$

$$\Gamma : \mathbf{x} \in \mathbf{V} \rightarrow i \in \zeta, \tag{1}$$

that maps the vector $\mathbf{x}$ to a digital symbol $i$. In a traditional vector quantizer, the symbol $i$ is selected to be the index $k$ of the $\mathbf{C}_k$ in the codebook $\mathcal{C}$ that minimizes a distortion measure $d(\mathbf{x}; C_k)$. Thus, in a traditional VQ system, if

$$k = \underset{j}{\operatorname{argmin}} \quad d\left(\mathbf{x}; C_j\right), \tag{2}$$

then $i \equiv k$. The VQ decoder reconstructs $\mathbf{x}$ using the prototype vector $C_k$. The VQ of $\mathbf{x}$ to $C_k$ can be represented as a function $Q$,

$$Q[\mathbf{x}] = C_k. \tag{3}$$

and the output of the encoder for the input $\mathbf{x}$ is $k$. If we consider the source vectors derived from the signal as random variables, then we can associate a probability $p(i)$ with every $i \in \zeta$. Let $p$ denote the probability mass function of the symbols, then the entropy of these symbols for a given signal source is

$$H = -\sum_{\text{all symbols } i} p(i).\log_2(p(i)) \quad (4)$$

Let a sequence of source vectors, $\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(t), \ldots$ be encoded using VQ and let $Q[\mathbf{x}(0)], Q[\mathbf{x}(1)], \ldots, Q[\mathbf{x}(t)], \ldots$ be the corresponding prototype vectors chosen by the vector quantizer. The DCR procedure exploits the correlation between consecutive source vectors to skew the PMF of the symbols in $\zeta$, thus resulting in a reduction in the entropy $H$. Thus at each time instance $t$, the proposed DCR algorithm re–orders the prototype vectors in the codebook in the increasing order of a suitably chosen dissimilarity measure between $Q[\mathbf{x}(t)]$ and all other prototype vectors in $\mathcal{C}$. The dissimilarity measure can be designed as a valid distance measure $D(Q[\mathbf{x}(t)], C_k)$, for $k = 0, 1, \ldots, K$.

Implementation of the DCR by physically reorganizing the codebook, which is typically stored in a memory, will require interchange of the contents of the memory and therefore may be prohibitively expensive. A much more efficient implementation of the DCR procedure can be achieved by employing a simple dynamic index map, $\Psi(k, t)$, that relates the physical address, $k$, of a prototype vector $C_k$ in the codebook $\mathcal{C}$ to its corresponding re–ordered index at each time instance $t$. Thus, $\Psi(k, t)$ can be thought of as the index of $C_k$ in the re–ordered codebook at time $t$. Unlike in the standard VQ where the digital symbol $i$ at instance $t$ (denoted $i(t)$) is set to the index $k$, in the VQ scheme employing DCR $i(t)$ is set to $\Psi(k, t)$. In the following two subsections, an algorithmic description of the VQ encoder and decoder with DCR is provided.

### 2.1. VQ Encoder with DCR

In this subsection, the VQ encoder algorithm that employs the proposed DCR is described. At $t = 0$, the dynamic index map, $\Psi(l, 0)$ is initialized as

$$\Psi(l, 0) = l, \text{ for } l = 0, 1, 2, \ldots, K-1. \quad (5)$$

For $t = 0, 1, 2, \ldots$ the encoding algorithm is given by

1) **Codebook Search:** Given the source vector $\mathbf{x}(t)$, the codebook $\mathcal{C}$ is searched according to (2) to determine the "best match" prototype vector $C_k$. Thus, $Q[\mathbf{x}(t)] = C_k$.

2) **Dynamic Index Map:** The physical index $k$ corresponding to $\mathbf{x}(t)$ is mapped to the reordered index using $\Psi(k, t)$ and the VQ encoder symbol $i(t) = \Psi(k, t)$ is made available to the decoder.

3) **Dynamic Codebook Reordering:** This step updates $\Psi(k, t)$. For $l = 0, 1, 2, \ldots, K-1$, the dissimilarity measure $D(Q[\mathbf{x}(t)], C_l)$ is calculated. Let us denote

$$\delta(l, t) = D(Q[\mathbf{x}(t)], C_l) \text{ for } l = 0, 1, 2, \ldots, K-1. \quad (6)$$

$\delta$ is then arranged in an increasing order. Let

$$\delta(l_0, t) \leq \delta(l_1, t) \leq \delta(l_2, t) \leq \ldots \leq \delta(l_K, t) \quad (7)$$

where $l_0, l_1, \ldots, l_K \in \{0, 1, 2, \ldots, K-1\}$. The dynamic index map $\Psi(j, t+1)$ is determined according to

$$\Psi(j, t+1) = l_j \text{ for } j = 0, 1, 2, \ldots, K-1. \quad (8)$$

It may be noted that since $\mathbf{x}(t)$ was vector quantized to $C_k$, $\Psi(k, t+1) = 0$ and the prototype vectors most similar to $C_k$ have a corresponding dynamic index map that is close to 0.

Since correlated source vectors can be expected to be vector quantized to similar prototype vectors, the VQ encoder symbol $i(t) \equiv \Psi(k, t)$ frequently assumes values close to 0. The PMF of the VQ encoder symbol is largely skewed towards values closer to 0.

### 2.2. VQ Decoder with DCR

Similar to the decoder, the encoder initializes its dynamic index map according to (5). For $t = 0, 1, 2, \ldots$

1) **Inverse Dynamic Index Map:** The encoder makes the symbol $i(t)$ available to the decoder. Since $\Psi(k, t)$ at $t$ is injective, the physical address (index), $k$ can be obtained from $i(t)$ through inverse dynamic index map.

$$k = \Psi^{-1}(i(t), t) \quad (9)$$

2) **Reconstruction** The decoder then reconstructs $\mathbf{x}(t)$ as $C_k$.

3) **Update Dynamic Index Map** Since $Q[\mathbf{x}(t)]$ is known at the decoder, $\Psi(j, t+1)$ for $j = 0, 1, 2, \ldots K-1$ is determined similar to the encoder (Dynamic codebook re–ordering step in the encoder description).

## 3. Entropy reduction with DCR for MELP

In the MELP speech vocoder specified in [2], the following parameters are extracted from every 22.5 msec frame of the speech signal: 10 line spectral frequencies (LSFs), 2 gain values, 1 pitch, 5 bit bandpass voicing flags, 1 bit aperiodic flag and 10 Fourier magnitudes. The encoders for the LSFs, the gain parameters and the Fourier magnitudes employ VQ. The training databases for the design of these vector quantizers were obtained from the MELP parameters derived from 200000 frames (each of 22.5 msec) of speech. These frames were obtained from speech records in the TIMIT training corpus [10]. In the testing mode, speech files from the TIMIT testing database are concatenated to form a 37.5 minutes long record and MELP parameters are derived from consecutive frames of this record. The incorporation of the DCR procedure in the encoders of these parameters is discussed below.

### 3.1. DCR in VQ of LSFs

In most model based speech coders, structurally constrained VQ architectures such as multistage vector quantizers (MSVQ) or split vector quantizers (SVQ) are employed to code the LSFs. In [2], a 25 bit, 4 stage MSVQ is employed. In general, structurally constrained vector quantizers encode a source vector using a set of encoders and decoders that are arranged in a predetermined architecture. Associated with each VQ encoder–decoder pair, $\mathbf{Q}_i, \widehat{\mathbf{Q}}_i$, is a codebook $\mathbf{C}^{(i)}$. In a multistage VQ (MSVQ) system, these encoders are arranged in a cascade, while in a split VQ system, the $N$ dimensional input vectors are split into $K$ lower dimensional sub–vectors and the encoder–decoder pairs operate in parallel and independently on each of these sub-vectors. If no DCR is employed in any of the encoder decoder pairs, each $\mathbf{x}[t]$ will be encoded at a rate of $\sum_{i=1}^{K} \log_2 n^{(i)}$ bits, where $n^{(i)}$ is the number of codevectors in the codebook $\mathbf{C}^{(i)}$.

In both MSVQ and SVQ, the DCR algorithm described in Section 2 is applied to each codebook $\mathbf{C}^{(i)}$, $i = 1, 2, \ldots, K$, at each instance $t$, depending on the prototype vector selected from that codebook. The symbol outputs of the $K$ constituent encoders can be jointly encoded. If $\varsigma^{(k)}$ is the output symbol of the $k^{\text{th}}$ encoder, the joint entropy $H_J$ can be defined in terms of the joint PMF of the $K$ stage symbols denoted by, $p(\varsigma^{(1)}, \varsigma^{(2)}, \ldots, \varsigma^{(K)})$.

$$H_J = - \sum_{\substack{\text{all possible} \\ \{\varsigma^{(1)},\ldots,\varsigma^{(K)}\}}} p(\varsigma^{(1)}, \ldots, \varsigma^{(K)}) \log \left( p(\varsigma^{(1)}, \ldots, \varsigma^{(K)}) \right) \tag{10}$$

The MELP standard uses a 4 stage 25 bit MSVQ for encoding the LSFs. In MSVQ, the vectors quantized by the stages following the first stage are typically decorrelated and consequently the DCR does not yield significant reductions in datarates. On the other hand, significant reduction in the joint symbol entropy can be obtained with a split VQ. We implement a $K = 2$ split VQ for the 10 LSFs obtained every frame, the first encoder encoding the first 4 LSFs and the second encoder encoding the remaining 6. The codebooks were trained using 200000 LSF vectors obtained from the TIMIT [10] training database. Each of the 2 encoders of the Split VQ uses a codebook with 4096 codevectors. The empirical probability mass function (PMF) of the VQ symbols, derived from 100000 LSF vectors obtained from the speech files in the TIMIT testing database [10], with and without the incorporation of DCR for the split VQ are shown in Fig. 1.

When DCR is not employed, the PMFs of the output symbols of the two encoders are approximately flat. Thus, to encode these symbols, approximately 24 bits are required. With the incorporation of DCR, it is observed that the symbols close to 0 occur more frequently than symbols further away from 0, thus dramatically skewing the PMFs. The empirical entropy corresponding to the empirical joint PMF of the symbols of the two encoders of the split VQ was found to be 16.63.

### 3.2. DCR in coding the pitch parameter

The pitch parameter, is quantized on a logarithmic scale with a 99-level uniform quantizer ranging from 20 to 160 (samples) in the MELP coder [11]. This uniform quantizer can be thought of as a 1 dimensional VQ, with the reconstruction levels representing the prototype vectors. The DCR procedure is applied to this quantizer and the empirical PMF of the output symbols of the encoder obtained from 100000 consecutive pitch values and the resultant empirical entropy was found to be 3.67 bits.

### 3.3. DCR in coding the MELP gain parameter

The two gain values, $G_1$ and $G_2$ derived every frame in the MELP coder, are quantized as follows in the standard MELP coder [11]: $G_2$ is quantized with a 5-bit uniform quantizer ranging from 10 to 77 dB. $G_1$ is quantized to 3 bits. We replaced the above mentioned encoders with a 2 dimensional vector quantizer with 256 prototype vectors. The quality of the reconstructed speech with the vector quantizer for gain parameters was indistinguishable from the standard MELP reconstruction. The DCR procedure is applied to the encoder and the decoder of the 2 dimensional VQ and the empirical PMF of the output symbols. The empirical entropy in this case is 6.51 bits.

### 3.4. DCR in coding the MELP bandpass voicing constants

The 5 bit bandpass voicing decision corresponding to consecutive MELP frames tend to be similar. To exploit this inter-frame correlation, a dynamic codeword re-ordering procedure is described below that is similar to the DCR. Thirty-two possible codewords can formed from the 5 bandpass voicing decision bits and these can be stored in a lookup table. This lookup table is similar to the codebook used in the DCR procedure. Instead of using the Euclidian distance to re-order the lookup table, the dissimilarity measure used is the Hamming distance between the codewords. Since 5 codewords exist that are equal Hamming distance away from a given 5 bit codeword, codewords that vary in their upper band voicing decisions are assigned lower values in the dynamic index map. As in DCR, the output symbol of the encoder is the mapped codeword.

### 3.5. DCR in coding the MELP Fourier magnitudes

In [11], the Fourier magnitudes are encoded using an 8-bit full-search vector quantizer with bark-scale weighting. The DCR algorithm is incorporated in this VQ procedure. The empirical entropy corresponding to the empirical PMF of the encoder output symbol obtained from 100000 consecutive Fourier magnitude vectors obtained from 37.5 minutes of speech from the TIMIT testing database was found to be 6.06 bits.

### 3.6. Reduced entropy coding of MELP parameters

In Table 1, the entropy in the the encoding the parameters of MELP coder when the DCR employed are summarized. The number of bits used in the fixed rate coding of the MELP parameters specified in [11] are also presented for comparison.

| Parameter | DOD standard MELP | | Entropy: MELP with DCR | |
|---|---|---|---|---|
| | Voiced | Unvoiced | Voiced | Unvoiced |
| LSFs | 25 | 25 | 16.63 | 16.63 |
| Gain | 8 | 8 | 6.51 | 6.51 |
| Pitch | 7 | 7 | 3.67 | 0 |
| Bandpass Voicing | 5 | 0 | 2.60 | 0 |
| Fourier Magnitudes | 8 | 0 | 6.06 | 0 |
| Aperiodic Flag | 1 | 0 | 1 | 0 |
| Error Protection | 0 | 13 | 0 | 0 |
| Total | 54 | 54 | 36.47 | 23.14 |

Table 1: Bit allocation for MELP coding and empirical entropy in symbols when encoders with DCR are used for MELP parameter coding.

In the 100000 frames used in these experiments, 76.94% were voiced and 23.06 % were unvoiced. Therefore, if encoder–decoder pairs with DCR are designed that are capable of representing the parameters using, on an average, the same number of bits as the empirical entropy shown in Table 1, we obtain a coding rate of approximately 1484 bits per second.

## 4. Conclusions

In this paper, we demonstrated the application of the dynamic codebook re-ordering algorithm to the encoders and decoders employed in coding the parameters of the MELP speech coder. It was demonstrated that significant reduction in the entropy of the output symbols of these coders can be obtained by the incorporation of the DCR procedure, without any degradation in quality or additional encoding delays compared to the traditional MELP coder. A lossless encoders such as a Huffman coder can be employed to exploit this reduction in entropy. Furthermore, the DCR procedure itself can be designed such that the resultant distribution of the VQ symbols is best suited for a given lossless compression scheme.

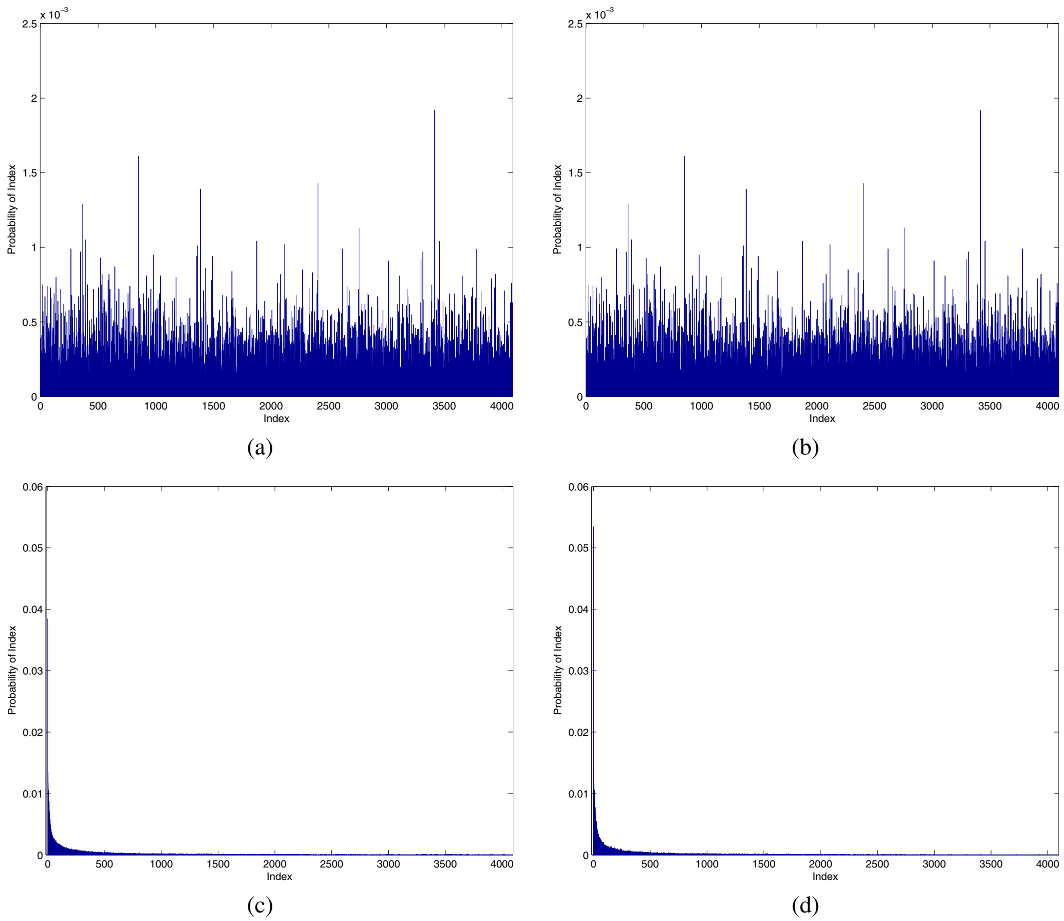


Figure 1: The empirical PMF of the symbol output of the (a) first sub-vector VQ encoder and the (b) second sub-vector VQ encoder without DCR. Correspondingly, (c) and (d) represent empirical PMFs when DCR is employed

## 5. References

[1] A. V. McCree and T. P. Barnwell-III, "A mixed excitation LPC vocoder model for low bit–rate speech coding," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 242–250, 1995.

[2] "The 1200 and 2400 bit/s NATO interoperable narrow band voice coder," NATO Standardization Agency, Den Haag, The Netherlands, STANAG No. 4591, 2003.

[3] E. B. George, A. V. McCree, and V. R. Viswanathan, "Variable frame rate parameter encoding via adaptive frame selection using dynamic programming," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 1996, pp. 271–274.

[4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1991.

[5] T. Wang, K. Koishida, V. Cuperman, A. Gersho, and J. Collura, "A 1200/ 2400 bps coding suite based on MELP," in *Proceedings of the Speech Coding Workshop*, 2002, pp. 90–92.

[6] ——, "A 1200 bps speech coder based on MELP," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 1375–1378.

[7] F. G. B. DeNatale, S. Fioravanti, and D. D. Giusto, "DCRVQ: A new strategy for efficient entropy coding of vector-quantized images," *IEEE Transactions on Communications*, vol. 44, no. 6, pp. 696–706, 1996.

[8] S. Sridharan and J. Leis, "Two novel lossless algorithms to exploit index redundancy in VQ speech compression," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 1998, pp. 57–60.

[9] V. Krishnan, "A framework for low bit–rate speech coding in noisy environment," Ph.D. dissertation, Georgia Institute of Technology, April 2005.

[10] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "The DARPA TIMIT acoustic-phonetic continuous speech corpus," Feb. 1993, cDROM: NTIS order number PB91-100354.

[11] "MIL-STD-3005: Specifications for the analog to digital conversion of voice by 2,400 bit/second mixed excitation linear prediction," Military Standard, US Department of Defense, 1999.