國立宜蘭大學工程學刊(2006)
第二期第 37～48 頁

Combination of switched predictive network and multi-stage VQ to efficiently encode LSF parameters

Hwai-Tsu Hu[1] Chu Yu[2]

1. Professor, Department of Electronic Engineering, National Ilan University
2. Associate Professor, Department of Electronic Engineering, National Ilan University

Abstract

In this paper a switched predictive vector quantizer is employed to efficiently encode the line spectrum frequencies (LSF) extracted from speech signals. To exploit the interframe and intraframe correlation of LSF parameters, we utilize a switched predictive network (SPN) to predict the upcoming LSF parameters. The prediction residual is then coded using 3-stage vector quantization (3SVQ). Separate sets of codebooks are particularly trained to cope with individual predictive networks. Improvement over the non-predictive case is demonstrated using spectral distortion measures based on a database of 60312 LSF vectors. During training and encoding, a regulation rule has also been developed to ensure the ordering property. This rule allows a full usage of codebook space and consequently leads to the reduction of spectral distortion by 0.028 dB. If computation is of great concern, the complexity can be reduced by half at the cost of slight degradation. Experiments show that the combination of a half-searching 8-SPN 3SVQ with a 4-best searching scheme requires only 18 bits to achieve transparent quantization with 20 ms frame separation.

Key words：line spectrum frequency, vector quantization, switched predictive network, predictor-quantizer system

結合轉轍式預測網路與多級向量量化以達成線頻譜對參數之高效率編碼

胡懷祖 [1] 游 竹 [2]

1. 國立宜蘭大學電子工程學系教授
2. 國立宜蘭大學電子工程學系副教授

摘 要

本文針對語音信號之線頻譜對參數，提出一套預測量化機制俾使編碼效能提昇，我們藉由轉轍式預測網路來利用鄰近音框的關係，使現下之線頻譜對參數能從過往音框的相關參數估算，而預測之殘餘再採三級向量量化加以編碼，至於性能改善則是以總計 60312 筆之線頻譜對向量的頻譜失真量測來表現。在訓練和編碼期間，所發展的調整規則可保證參數維持其序列性，而這種規則更可活用全部代碼，進一步使頻譜失真下降 0.028 分貝。假若計算需求值得高度關切，或可將代碼簿搜尋的範圍減半，惟這樣的作法也伴隨著效能稍降的代價。實驗顯示，一旦將八轉轍預測網路、三級向量量化與前四個最佳的搜尋策略予以結合，僅需 18 個位元即可達成 20ms 音框間隔之透明編碼。

關鍵詞：線頻譜頻率，向量量化，轉轍式預測網路，預測量化器之系統

1. Introduction

Efficient quantization of spectral parameters is essential for linear prediction (LP) based speech coders operating at low bit rates. Among recently developed coding systems, the most prevailing approach to encoding the spectral information is the vector quantization (VQ) of LSF parameters. The popularity of LSF parameters results from its superiority in stability check, excellent interpolation properties, and relative insensitivity to quantization errors [1,2]. Methods for quantizing LSF parameters such as split and multi-stage VQ have been reported to achieve satisfactory performance at 22-24 bits per 20 ms frame [3,4]. By exploiting the correlation of the LSF parameters across adjacent frames, more efficient quantization is achievable by using prediction [5-11]. This implies that the LSF parameters in the current frame are predictable by using the LSF parameters in previous frames. The VQ is then applied to the prediction residual with a lower dynamic range. Simulation results show that, with the participation of prediction, fewer bits are sufficient to achieve comparable performance. In light of such a concept, we proposed a switched predictive network (SPN) in combination with 3-stage vector quantization (3SVQ) to achieve low-bit rate coding of LSF parameters. Since the codebook content in 3SVQ also affects the final result of quantization, our investigation includes how to organize these predictive networks and multistage codebooks, and how to determine the best predictor-quantizer combination. In the following we present in details the design procedure for the predictor-quantizer system as well as the performance evaluation with respect to different bit allocations. Some related factors such as computational complexity and storage requirement are taken into consideration as well.

2. Network structure

The proposed network has only one layer of neurons with inputs connected through matrix **W** and biased by vector **b**. The output of each neuron corresponds to a LSF and the employed transfer function is purely linear. These neurons are trained to predict the LSF parameters of the current frame by using the LSF parameters gathered from previous frames. Our first encountered problem is to determine the number of past frames that should take part in the prediction. We begin with a test based on a data set consisting of 60312 speech frames extracted from Mandarin sentences uttered by 10 speakers (5 males and 5 females). The sampling rate is 8 KHz and the size of analysis frames is 20 ms. For each frame the LSF parameters are obtained by converting 10th-order linear prediction (LP) coefficients. Fig. 1 presents the results in terms of mean square error (MSE) between the actual and predicted LSF parameters in the training set. Results of linear predictors derived from individual LSF contours with different orders are also provided for comparison. It is shown that the proposed network consistently outperforms the linear predictor for the orders under investigation. For both cases, the improvement becomes sluggish as the order exceeds two. Hence in the following discussion the predictive network only take the information of two previous frames as input.

To quantize the input LSF parameters, a SPN is adopted as the front-end processor of 3-stage VQ. Such a SPN aims at reducing the dynamic range of input parameters for efficient quantization. The procedure for generating the required number of predictive networks is similar to that for conventional VQ. For explanatory convenience, we regard the sequence of 10th-order LSF parameters as a vector. Initially all LSF vectors are grouped into one cluster. A primitive network is trained by using the quasi-Newton backpropagation method with all LSF vectors involved. The LSF vectors are split into two clusters of equal size, i.e., vectors with smaller prediction errors are gathered as the first subgroup while those with larger errors are categorized into the second subgroup. We then derive two predictive networks according to two subgroups of LSF vectors. Subsequent to the generation of two new networks, an iterative approach is brought in to adapt the networks to minimize the sum-squared prediction error. Within each iteration, we examine the predictability of these two networks for LSF vectors in the cluster. The LSF vector is reassigned to an alternative subgroup if the underlying predictive network attains a larger prediction error. Finally, we retrain the two networks according to the rearranged subgroups of LSF vectors. Such an iterative process continues until a convergence criterion is satisfied or a certain number of iterations are completed. At the end of iteration we tag the values of sum-squared-error (SSE) with the predictive networks and treat each subgroup of LSF vectors as a new cluster. In case the expected number of predictive networks is not reached, the next cluster for splitting is chosen as the one with a maximum SSE among existing clusters. In Table 1, we present experimental results of the 1-, 2-, 4- and 8- SPN using all training samples. The obtained results surely illustrate the efficiency of the switched network in predicting the LSF parameters.

3. Vector quantization of LSF parameters

As the switched network offers an initial estimate of the LSF vector, the object for quantization thus becomes the prediction residual that is obtained by subtracting the predicted LSF vector from its actual value. We employ an *M*-best 3-stage VQ scheme [2] to encode the LSF residual. The codebooks are trained using the well-known generalized Lloyd algorithm. During codebook training, the distance measure *D* between the original and quantized LSF vectors is defined as

$$D=\sum_{i=1}^{10} w_i\left(l_i-\tilde{l}_i\right)^2 \qquad (1)$$

where l_i represents the *i*th original LSF parameter and $\tilde{l}_i$ denotes its quantized version. w_i is the weight derived from a Mel-frequency warping-based function given by [12].

$$w_i=\left(\frac{1}{l_i-l_{i-1}}+\frac{1}{l_{i+1}-l_i}\right)\left(1+\frac{2}{l_i}\tan^{-1}\frac{0.45\sin l_i}{1-0.45\cos l_i}\right)^2,$$
$$1\le i\le 10;\quad l_0=0,\ l_{11}=1. \qquad (2)$$

This weighting function emulates the human auditory system that emphasizes spectral peaks more than spectral valleys on a Mel-frequency scale. For each residual vector to be encoded, the 3-stage codebooks are searched using a tree search procedure. At the first stage, *M* codevectors that achieve the lowest weighted distortion are selected and the error vectors are obtained by subtracting the residual vector from the codevector. The second-stage codebook is then searched using the *M* error vectors; each leading to *M* possible paths with lowest weighted distance. Again, *M* out of M^2 paths that achieve the overall lowest distance measures are selected. This procedure is continued for all stages of codebooks. Once *M* paths are resolved for all stages, the best out of *M* paths is determined by minimizing the spectral distortion (SD) between the given and quantized vectors. Eventually, the attributes used to characterize the quantized LSF vectors include the index of the chosen predictive network along with the indices of the selected codevectors from all stages. The quantized LSF vector is reconstructed by summing up the chosen network output and the codevectors from the 3-stage codebooks, i.e.,

$$\tilde{\mathbf{l}}_t=\mathbf{W}\begin{bmatrix}\tilde{\mathbf{l}}_{t-1}\\ \tilde{\mathbf{l}}_{t-2}\end{bmatrix}+\mathbf{b}+\sum_{k=1}^{3}\mathbf{c}_{k,i_k}, \qquad (3)$$

where $\tilde{\mathbf{l}}_t$ represents the quantized LSF vector at the *t*th frame. $\mathbf{W}$ and $\mathbf{b}$ denote the weighting matrix and bias vector, respectively. The codevector at the *k*th stage is denoted as $\mathbf{c}_{k,i_k}$, where the subscript i_k indicates the codebook entry. Fig. 2 illustrates the performance gain due to the *M*-best search scheme for various bit allocations while *M*=1,2, and 4 respectively. The SD measure in our computer simulation is defined as

$$d\left(S(f),S_q(f)\right)=\left(\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}\left[10\log_{10}S(f)-10\log_{10}S_q(f)\right]^2 df\right)^{1/2}, \qquad (4)$$

where $S(f)$ and $S_q(f)$ are the original and quantized power LP-spectra, respectively. α and β correspond to 125 Hz and 3.7 KHz, respectively. The allocation of bits in Fig. 2 is arranged as follows. We begin with 6 bits per stage. Each additional bit is added to the stages in a sequential manner such that the number of bits in the subsequent stage is not larger than that of the preceding stage. We increment the number of bits by one each time until 24 bits are reached. The increase of *M* improves quantization performance but inevitably raises computational burden to a great extent. Although *M*=8 is experimentally sufficient and suggested by many researchers, we choose *M*=4 as a tradeoff between quantization performance and computational load.

During the codebook implementation, we also examine the effect of different error criteria on quantization performance. Given that the mean-squared-error (MSE) criterion is used, the resultant *i*th codevector at the *m*th codebook, termed $\mathbf{c}_{m,i}$, is derived as

$$\mathbf{c}_{m,i}=\frac{1}{N_i}\sum_{k=1}^{N_i}\mathbf{e}_{m,k}, \qquad (5)$$

where $\mathbf{e}_{m,k}$ denote the error vector, and N_i is the number of error vectors in the *i*th cluster. When a weighted mean-squared-error (WMSE) criterion is adopted, the codevector $\mathbf{c}_{m,i}$ becomes

$$\mathbf{c}_{m,i}=\sum_{k=1}^{N_i}\left(\mathbf{e}_{m,k}\right)^T\mathbf{W}_k\left(\mathbf{e}_{m,k}\right)\Big/\left(\sum_{k=1}^{N_i}\mathbf{W}_k\right)^{-1}, \qquad (6)$$

where $\mathbf{W}_k$ is a diagonal matrix with w_i's on the main diagonal. We demonstrate the efficacy of the two criteria in Table 2 by showing the WMSE's and SD's derived from the 3SVQ using a 4-best searching scheme. As revealed by the simulation results, it is to our surprise that the employment of the WMSE criterion in codebook generation does not always yield better quantization performance. However, we find no fault with the WMSE criterion since it outperforms the MSE criterion in terms of the average WMSE especially obtained at the first stage. We believe that such results are due to the following two reasons. First, it appears that certain mismatch exists between the average WMSE and SD. Second, the *M*-best searching scheme may end up with an inferior codevector at the preceding stage as long as the final reconstructed vector is optimal. Since the use of the WMSE criterion can only increase computational burden without showing any advantage in SD, Eq. (5) is employed to derive the codevector throughout the rest of this paper.

Notice that in the above simulation we use the original LSF vector instead of the quantized one to carry out network prediction. Nevertheless, this impractical condition will be amended later when we re-optimize the multi-stage codebooks. A disadvantage of the 3SVQ is that the reconstructed LSP parameters are not guaranteed to satisfy the ordering property. As a result, the codevectors that violate the ordering property are discarded during the codevector search. It has been reported that efficient usage of codebook space can improve quantization performance [13]. Thus, we proposed a rectification rule to ensure the ordering property by letting

$$\tilde{l}_i = \max\left\{\tilde{l}_{i-1} + 0.003, \hat{l}_i\right\}, \tag{7}$$

where $\hat{l}_i$ represents an intermediary value obtained prior to the addition of the underlying codevector. The proposed rectification allows us to exploit the full codebook space and reduce SD by approximately 0.028 dB for the 4-best searching scheme shown in Fig. 2.

4. Switched predictive network

As mentioned earlier, the switched network is intended to improve the predictive accuracy. Though an accurate prediction of the LSF vector is beneficial to subsequent VQ, the network that yields the minimum prediction error will not necessarily lead to the best quantization result. In other words, the deficiency of network prediction may be partially compensated by manipulating the codebook content. Hence it seems plausible to cope each predictive network with distinct codebooks in order to attain the optimal performance. In this study, the 3-stage codebooks associated with a specific network are obtained by quantizing the LSF residual vectors that belong to the same cluster during network training. The motivation behind such an arrangement is to minimize the average quantization error for the residual vectors in each cluster. However, quantization in this manner can only results in a local minimum because it does not take into account the interaction between the predictive network and 3-stage codebooks. This problem is resolved by a procedure called codebook re-optimization, which is presented in the next section. Here we put emphasis on the quantization performance achieved by the SPN-3SVQ system. To encode a given LSF vector, all the network/codebook pairs are examined and the one that provides the smallest distortion is selected. Fig. 3 presents the resulting average distortion measures given that the bit allocation ranges from 18 to 22 per frame while the number of reserved bits for network switching extends from 1 to 3. Performance measures of the 8-SPN 3SVQ are particularly listed in Table 3 for reference. It appears that the 8-SPN in conjunction with 3-stage {5,5,5} VQ yields an averaged spectral distortion below 1 dB, which is a level close to transparent quantization. Thus our efforts have been made to achieve transparent quantization of spectral characteristics with 18 bits per 20 ms frame.

Despite that the use of separate codebooks improves the quantization performance, it demands more computation efforts as well as memory storage. Let us now adopt the 3-stage, $\{p,q,r\}$, VQ with a *M*-best searching scheme as the baseline for comparison. The sizes for codebooks at three stages are 2^p, 2^q, and 2^r, respectively. Suppose that the amount of computation required for assessing the suitability of a codevector is λ. If we disregard the need for drawing *M*-best paths out of possible candidates at each stage, the overall computation for determining the optimal codevector sequence can be approximated by $[2^p + M(2^q + 2^r)]\lambda$. A similar analytic strategy can be applied to the SPN-3SVQ. We now reserve some bits, e.g. $\{a,b,c\}$ which are originally allocated for codebooks at three stages, to indicate the index of predictive networks. The computation necessitated by the SPN-MSVQ sums up to $(2^{a+b+c})[2^{p-a} + M(2^{q-b} + 2^{r-c})]\lambda$ in addition to the evaluation of predictive networks. Since $(2^{a+b+c})[2^{p-a} + M(2^{q-b} + 2^{r-c})]\lambda \geq [2^p + M(2^q + 2^r)]\lambda$, we may readily conclude that the employment of the SPN surely raises the computational complexity.

The analysis of storage requirement due to the involvement of the SPN follows the same principle. Given that each codevector occupies a memory space of σ, the total allocation for codebooks in 3-stage $\{p,q,r\}$

VQ amounts to $(2^p + 2^q + 2^r)\sigma$. On the other hand, as we connect each predictive network with separate codebooks, there are in total 2^{a+b+c} 3-stage codebooks, each of which calling for a memory space of $(2^{p-a} + 2^{q-b} + 2^{r-c})\sigma$. Again, since $2^{a+b+c}(2^{p-a} + 2^{q-b} + 2^{r-c})\sigma \geq (2^p + 2^q + 2^r)\sigma$, the proposed quantization scheme thus consumes more memory than that needed by multistage VQ without prediction.

In this study, we have verified the advantage of using separate sets of multi-stage codebooks rather than sharing a single set of codebooks. If memory space becomes a serious concern in practical implementation, a better way to reduce memory would be the use of algebraic VQ [14] or split-matrix [15] methods. This paper does not intend to get involved with the other possible substitutes. Our attempt instead focuses on lowering computational efforts without causing too much degradation. As shown in Eq. (3), the quantized LSF vector comprises two parts: one is the network output and the other is the summation of the chosen codevectors from all stages. To encode a LSP vector, we need to examine all network/codebook pairs before making any selection. It is observed in our experiments that the optimal result often emerges from the codebooks with the most accurate predictive network. Fig. 4 demonstrates such a phenomena by showing the accumulated probability for the optimal codevector sequence drawn from the network/codebook pairs that are sorted according to the WMSE's of network prediction. Taking the 18-bit 8-PSN MSVQ as an example, the search for the first half of the network/codebook pairs with lesser errors has nearly 96% of chance to acquire the optimal codevectors before completing a full search of all pairs. More importantly, the corresponding average SD is only slightly degraded. If the computational efficiency is a matter of concern, it seems worthwhile to reduce the computation by half at the cost of slightly increased SD. In accordance with the principle used for complexity evaluation, the computational requirement for a half searching of this 8-SPN 3SVQ is roughly equal to that required for a 3-stage {7,7,7} VQ without prediction. For real-time implementation the demand on computation can be easily afforded by nowadays processors.

5. Codebook reoptimization

Following the initial establishment of multistage codebooks, the next step to improve the quantization performance is codebook reoptimization. Notice that the search for the optimal codevector at the present stage does not make reference to codebook contents at subsequent stages. One may therefore consider all codevectors at the following stage as null. However, since the multi-stage codebooks become available after initial setup, the codebook at each stage can be optimized subject to the others. More specifically, the object for codebook training at each stage can be replaced by the error vector between the residual vector and the reconstructed vector that consists of codevectors from all stages except the one being reutilized.

The reoptimization process is carried out in an iterative manner. Within each iteration, we perform quantization in respect of every LSF vector to render an optimal index sequence termed $(p, j_1(n), j_2(n), j_3(n))$, where p denotes the network selection, n is a counter, and $j_k(n)$ corresponds to the codevector sequence. For the codebooks connected with the pth predictive network, the i_mth codevector at stage m, denoted as $\mathbf{c}_{m,i_m}{}^{(p)}$, is updated by

$$\mathbf{c}_{m,i_m}{}^{(p)} = \frac{1}{N}\sum_{n=1}^{N}\left\{\mathbf{l}_t - \left(\mathbf{W}^{(p)}\begin{bmatrix}\tilde{\mathbf{l}}_{t-1}\\ \tilde{\mathbf{l}}_{t-2}\end{bmatrix} + \mathbf{b}^{(p)}\right) - \sum_{\substack{k=1\\k\neq m}}^{3}\mathbf{c}_{k,j_k(n)}{}^{(p)}\right\}, \tag{8}$$

where N represents the number of occurrences when $i_m = j_m(n)$ and the switch index is p. Eq. (8) is analogous to moving the codevector to the centroid of error vectors. The above-mentioned reoptimization iterates between codevector searching and codebook readjustment. By using the reoptimized codebooks, the average quantization error exhibits a declining tendency in a mean-squared sense, and the corresponding SD is usually decreased as well. It is particularly emphasized that during the codebook reoptimization we adopt the quantized LSP vectors of past frames to predict the current LSP vector. This is the actual situation encountered at the receiving end of a speech coder. As the predictive networks are derived from the original LSF vectors, the substitution of quantized vectors in network prediction generally leads to unexpected perturbation. However, during the iterative process we decide to keep the predictive networks intact despite that the adaptation of the predictive networks may possibly reduce the perturbation. Reasons for such a decision are: 1) the perturbation can be partially compensated by proper adjustment of multistage codebooks. 2) As the quantized vectors progressively move toward the original vectors via codebook reoptimization, the resulting network coefficients eventually come close to that obtained using the original LSP vectors.

Fig. 5 shows the average SD measures derived from the reoptimized codebooks of the 18-bit SPN-3SVQ for first 20 iterations with the performance measures tabulated underneath for every 5 iterations. It can be seen that the SD is maintained below 1 dB as the number of iterations goes beyond 6. Moreover, fewer than 2% outliers with SD in the range of 2-4 dB are observed, and outliers with SD greater than 4 dB are almost not found.

Apart from the data for training, an independent database consisting of 60145 LSF vectors is used for evaluating the out-of-training case. Table 4 lists the performance of the proposed 8-SPN 3SVQ at 18 bits/frame. The employed 3-stage codebooks are that derived from the reoptimization process with 20 iterations. The tabulated data indicate an insignificant degradation for the out-of-training test. We believe that such a result is due to similar distribution of prediction error vectors in both inside- and outside-training cases.

6. Concluding remarks

This paper presents a switched predictive network to decorrelate the LSF parameters in both temporal and spectral domains. The parameters to be quantized are the prediction residuals. As the variance of the prediction residuals is far less than that of the original LSF parameters, the switched prediction strategy makes certain the success of low bit-rate speech coding. We employ a 3-stage vector quantizer to encode the prediction residual. The combination of an 8-switched predictive network with 3-stage VQ results in transparent quantization at a rate of 18 bits per 20 ms frame. This implementation adopts a regulation rule to properly order the LSF parameters and reduce the required computation by half through selective searching of network/codebook pairs. Furthermore, the simulation with respect to the out-of-training group demonstrates the efficiency and robustness of the proposed quantizer.

Acknowledgment

The research was supported by the National Science Council under Grant NSC89-2213-E-197-002.

References

[1] SOONG, F. K., and JUANG, B.-H., “Line spectrum pair and speech data compression,” *Proceedings of ICASSP*, 1984, pp. 1.10.1-1.10.4

[2] SOONG, F. K., and JUANG, B.-H., “Optimal quantization of LSP parameters’, *IEEE Trans. Speech Audio Process.*, 1993, 1, (1), pp. 15-24

[3] PALIWAL, K., and ATAL, B., “Efficient vector quantization of LPC parameters at 24 bits/frame,” *IEEE Trans. Speech Audio Process.*, 1993, 1, (1), pp. 3-14

[4] LeBLANC, W. P., BHATTACHARYA, B., MAHMOUD, S. A., and CUPERMAN, V., “Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding,” *IEEE Trans. Speech Audio Process.*, 1993, 1, (4), pp. 373-385

[5] YONG, M., DAVIDSON, G., and GERSHO, A., “Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction,” *Proceedings of ICASSP*, 1988, pp. 402-406

[6] GRASS, J. and KABAL, P., “Methods of improving vector-scalar quantization of LPC coefficients,” *Proceedings of ICASSP*, 1991, pp. 657-660

[7] KUO, C.-C., JEAN, F.-R., and WANG, H.-C., “New LSP encoding method based on two-dimensional linear prediction,” *IEE Proc. -I*, 1993, 140, (6), pp. 415-419

[8] XYDEAS, C. S., and SO, K. K. M., “A long history quantization approach to scalar and vector quantization of LSP coefficients,” *Proceedings of ICASSP*, 1993, pp. II-1-II-4

[9] JEAN, F.-R., and WANG, H.-C., “Transparent quantization of speech LSP parameters based on KLT and 2-D-prediction,” *IEEE Trans. Speech Audio Process.*, 1996, 4, (1), pp. 60-66

[10] ZARRINKOUB, H., and MERMELSTEIN. P., “Switched prediction and quantization of LSP frequencies,” *Proceedings of ICASSP*, 1996, pp. 757-760

[11] NANDKUMAR, S., SWAMINATHAN, K., and BHASKAR, U., “Robust speech mode based LSF vector quantization for low bit rate coder,” *Proceedings of ICASSP*, 1998, pp. 2193-2196

[12] CHOI, S. H., KIM, H. K., and LEE, H. S., “LSP weighting functions based on spectral sensitivity and mel-frequency warping for speech recognition in digital communication,” *Proceedings of ICASSP*, 1999, pp. 401-404

[13] KIM, M. Y., HA, N. K., and KIM, S. R., "Linked split-vector quantizer of LPC parameters," *Proceedings of ICASSP*, 1996, pp. 741-744

[14] XIE, M, and ADOUL, J.-P., "Algebraic vector quantization of LSF parameters with low storage and computation complexity," *IEEE Trans. Speech Audio Process.*, 1996, 4, (3), pp. 234-239

[15] XYDEAS, C. S., and PAPANASTASIOU, C., "Split matrix quantization of LPC parameters," *IEEE Trans. Speech Audio Process.*, 1999, 7, (2), pp. 113-125

94 年 01 月 20 日投稿

94 年 01 月 20 日接受

Table 1: Performance improvement due to switched predictive network in terms of MSE's.

	1-SPN	2-SPN	4-SPN	8-SPN
MSE ($\times 10^{-3}$)	0.6726	0.5193	0.4113	0.3443

Table 2: Quantization performance due to different criteria in codebook training.

	MSE criterion				**WMSE criterion**			
	Distortion measured at 1st stage		**Distortion measured at 3rd stage**		**Distortion measured at 1st stage**		**Distortion measured at 3rd stage**	
Bit allocation	SD	WMSE	SD	WMSE	SD	WMSE	SD	WMSE
18 (6,6,6)	2.1216	0.4064	1.1063	0.1286	2.1293	0.4026	1.1033	0.1279
21 (7,7,7)	1.9963	0.3540	0.9226	0.0879	1.9996	0.3503	0.9235	0.0883
24 (8,8,8)	1.8773	0.3054	0.7693	0.0599	1.8770	0.3030	0.7706	0.0606

Table 3: Performance of the 8-SPN 3SVQ at rates of 18-22 bits/frame.

Bits/frame	**SD (dB)**	**2-4 dB (%)**	**>4 dB (%)**
18 (3,5,5,5)	0.9689	2.1919	0
19 (3,6,5,5)	0.9133	1.3016	0
20 (3,6,6,5)	0.8562	0.5952	0
21 (3,6,6,6)	0.8111	0.2238	0

Table 4: Performance evaluation of the 18-bit 8-SPN 3SVQ for out-of-training data.

Bits/frame	**Search range**	**SD (dB)**	**2-4 dB (%)**	**>4 dB (%)**
18 (3,5,5,5)	Half	0.9958	2.2762	0.0033
18 (3,5,5,5)	Full	0.9837	2.0018	0.0033

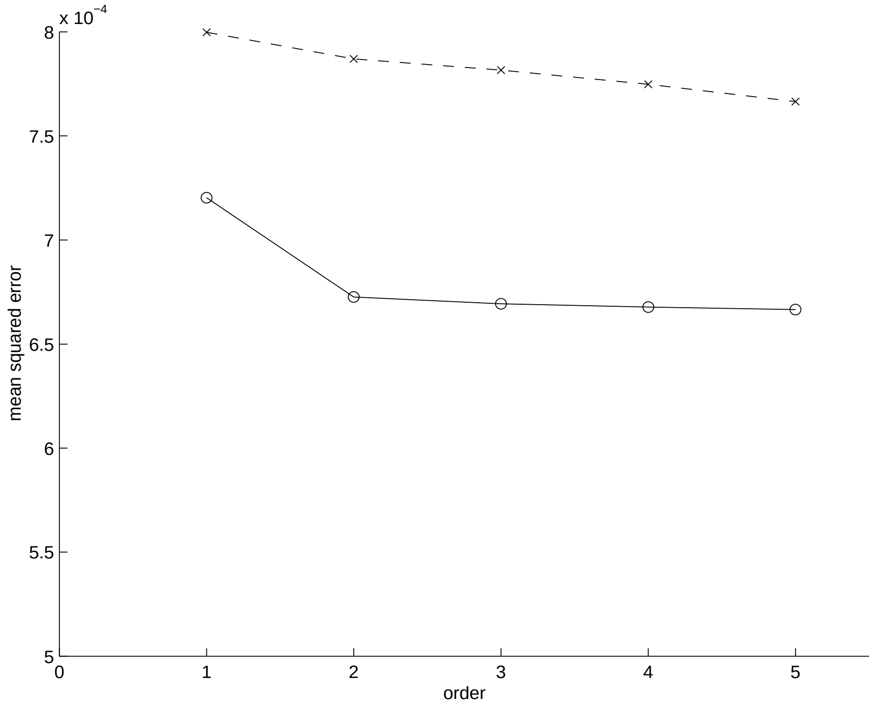


Fig. 1　MSE's between the actual and predicted LSF parameters.

×-mark: by linear predictor
○-mark: by predictive neural network

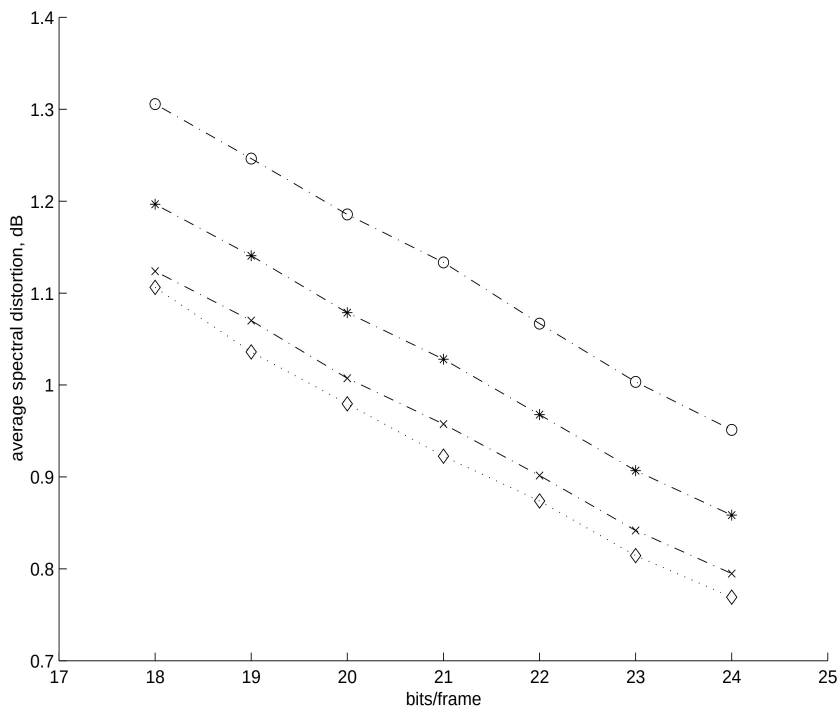


Fig. 2　Average spectral distortion between the original and quantized LSF parameters obtained from a single predictive network with 3-stage VQ using *M*-best searching.

○-mark:　*M*=1　　　×-mark: *M*=4
✶-mark:　*M*=2　　　◇-mark:　*M*=4 with regulation

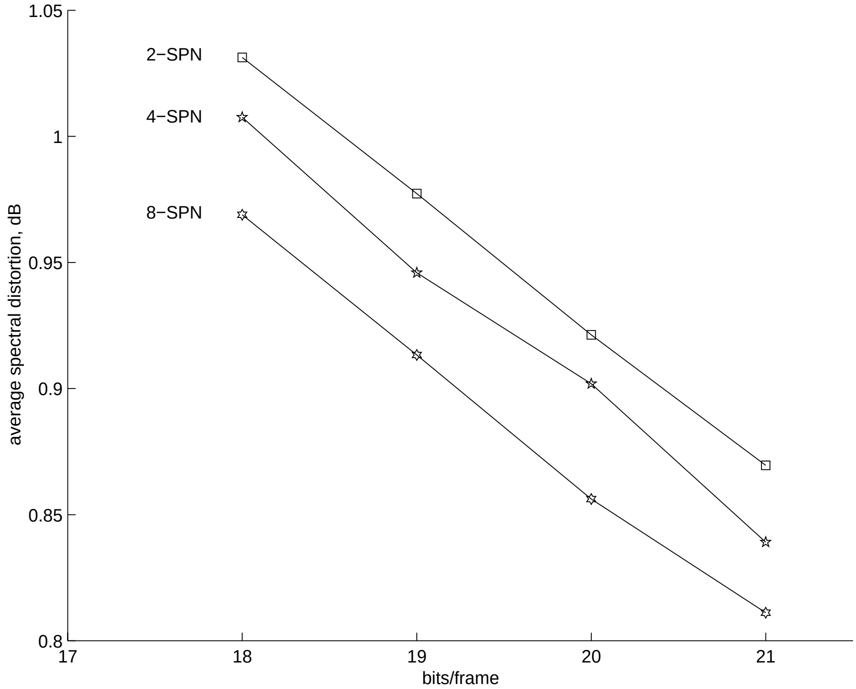


Fig. 3 Average SD performance for various SPN-3SVQ at rates of 18-22 bits/frame.

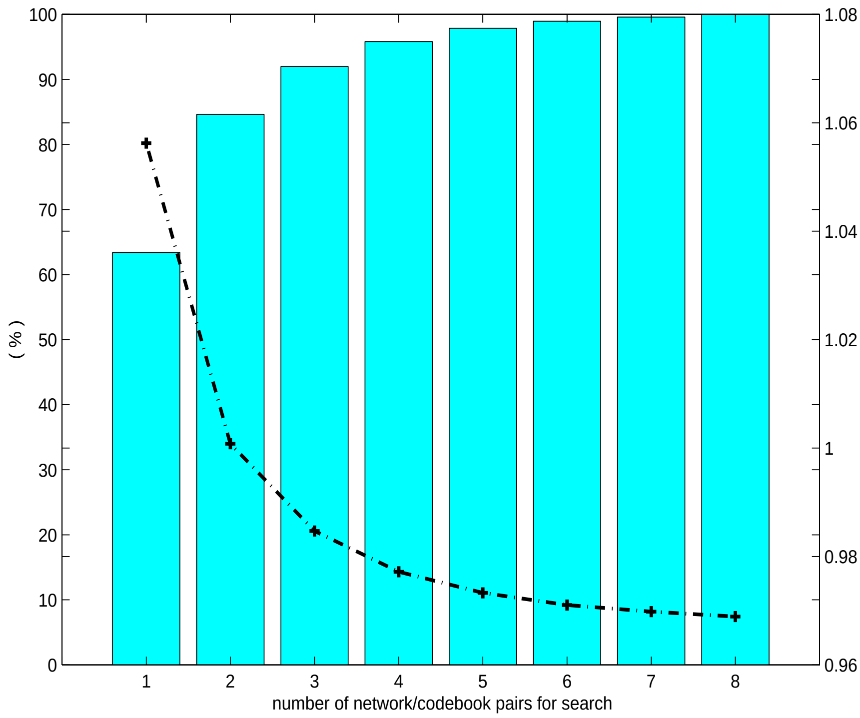
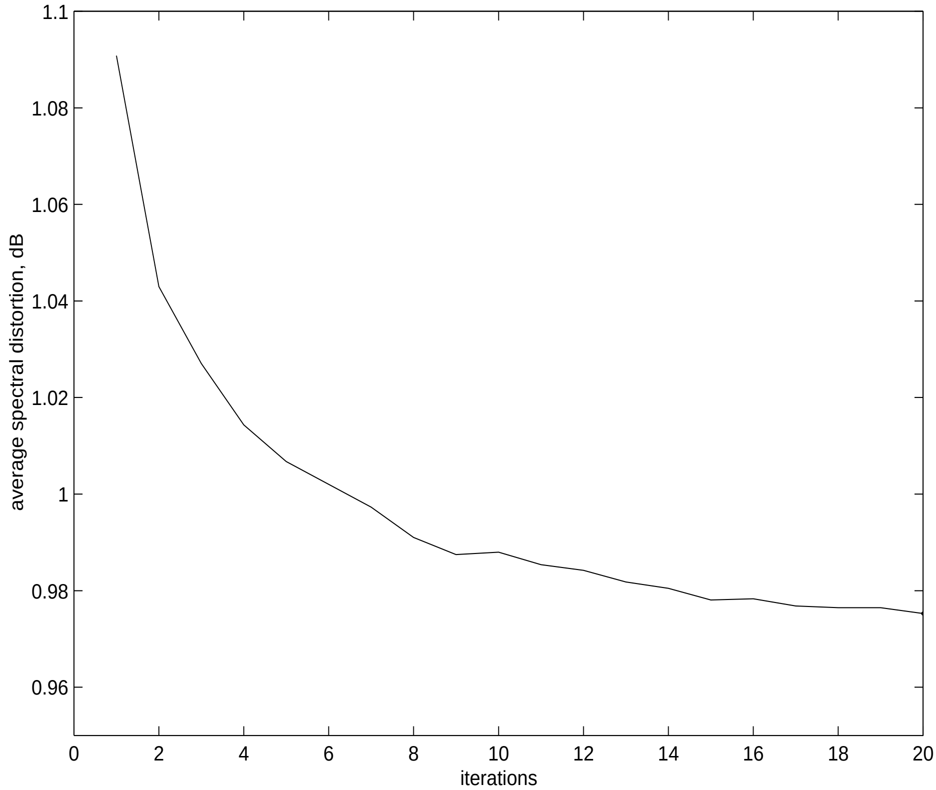


Fig. 4 Illustration of incomplete codebook search in the 18-bit 8-SPN 3SVQ. The bar chart reflects the accumulated percentage of the optimal codevector sequence acquired from a number of network/codebook pairs with better prediction; the overlaying plot indicates the average SD of the reconstructed LSF vectors.



Iterations	SD (dB)	2-4 dB (%)	>4 dB (%)
5	1.007	2.029	0.005
10	0.988	1.816	0.003
15	0.978	1.658	0.005
20	0.975	1.655	0.007

Fig. 5　Average SD versus iterations of codebook reoptimization with 8-SPN 3SVQ at 18 bits/frame.