QUANTISATION FOR MULTIPLE DESCRIPTION CODING FOR VOICE OVER IP

S. Kakouros, S. Villette and A.M. Kondoz

Centre for Communication Systems Research, University of Surrey
Guildford, Surrey, GU2 7XH, United Kingdom
s.villette@surrey.ac.uk

ABSTRACT

The transmission of voice over IP networks is heavily affected by packet losses. An increasingly popular method to increase the error resilience of these systems is the use of Multiple Description Coding (MDC). However, the MDC techniques commonly used tend to add a significant amount of redundancies, which are not always easy to use optimally. In this paper, we propose a simple vector quantisation scheme to maximise MDC performance, and study several factors affecting its performance under various error conditions. The results show that it is possible to obtain good performance under packet loss conditions, while using only limited amounts of redundancy.

Index Terms— Speech coding, quantization

1. INTRODUCTION

Voice transmission over packet-based data communication systems suffers from packet losses. These packet losses result in degraded speech quality that can render the signal unintelligible for the receiver.

Retransmission of the lost or dropped packets in the network is not acceptable since it would cause overlong delays. The initial approach towards combating packet losses was set off with the introduction of Packet Loss Concealment (PLC) [1] [2]. PLC offers a mechanism to replace or fill in missing speech fragments that were lost during transmission over a packet-switched network. However, the resulting signal is not always of satisfactory quality [3]. MDC solves some of these shortcomings, and allows a more graceful degradation of the received quality in the presence of packet losses, by introducing extra redundancies in the system.

The most common implementation of MDC is a system which generates two equal rate descriptions so that each description provides low but acceptable quality and both descriptions together produce higher quality. The two descriptions are packetised independently and sent through two different channels (or the same channel in different packets). More generally, there can be more than two descriptions, which can be at different rates.

In the course of its development, MDC has followed two main directions: processing of the coded speech at the source coding level [4]-[6] and processing of the speech at parameter level [7]. In this paper, the work we present is a continuation of MDC at parameter level as presented in [7]. The focus is in optimally exploiting the redundancy incorporated in the MDC system in order to achieve better quality at the receiver end.

2. QUANTISATION SCHEME

2.1. Theory

In [7] we introduced a scheme in which MDC was performed at parameter level. The various speech parameters produced by a sinusoidal speech coder were quantised at different rates (namely 5, 2 and 1 kb/s), and sent across different IP packets. Each speech frame was then decoded using the highest quality description received, or using extrapolated parameters in case all descriptions were lost in transmission. One limitation of this scheme is that it is arguably not a true MDC scheme, in that the descriptions are treated independently at the decoder, instead of being combined to produce a higher quality description than any of the individual descriptions. This is illustrated in Figure 1.

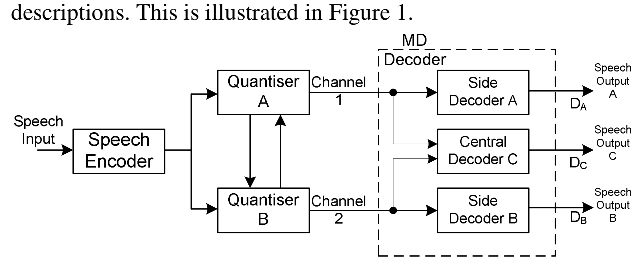

Fig. 1. Overview of MDC quantisation process

In this paper we aim to investigate general techniques to achieve this goal. Simulations have been carried out using two-dimensional Gaussian variables of zero mean and unity variance. For simplicity we are considering the case of two descriptions only, with a total bit allocation of 10 bits. This is expected to represent well the parameters typically used in speech coders,

We use the following notations:

- x and $\hat{x}$ are the original and quantised input
- $C_A(i)$ is the ith codeword from quantiser A
- $C_B(j)$ is the jth codeword from quantiser B

In the classic quantisation case, where x is quantised by quantisers A and B independently, the quantisation indexes i^* and j^* are selected as:

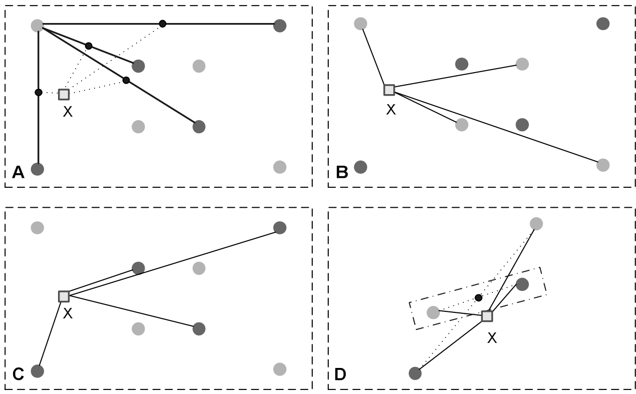

Fig. 2. Overview of proposed codeword selection process and comparison with simple quantisation. A. Proposed quantisation scheme, B. Employing only quantiser A, C. Employing only quantiser B, D. Condition for optimal selection.

$$i^* = \arg\min_i [(C_A(i) - x)^2],\ j^* = \arg\min_j [(C_B(j) - x)^2]$$

In this paper, we propose to study the case where when both descriptions are received, a linear combination of the corresponding codewords is used, i.e.:

$$\hat{x} = \alpha . C_A(i^*) + (1-\alpha) . C_B(j^*)$$

where α is an adjustable weighting factor accounting for the fact that the descriptions may be of different quality.

Figure 2 shows the different cases, and illustrates the fact that the quantisation indexes i^* and j^* which minimise the quantisation errors for the individual descriptions may not minimise the error for the combined description. In order to take this into account, we propose to select the quantisation indexes as:

$$(i^*, j^*) = \arg\min_{i,j} [\gamma . [(\alpha . C_A(i) + (1-\alpha) . C_B(j)) - x]^2 + (1-\gamma) . [(C_A(i) - x)^2 + (C_B(j) - x)^2]]$$

Where γ is a factor ranging between 0 and 1 weighting the quantisation error for individual descriptions against that when combining the descriptions.

2.2. Experimental results

Simulations have been carried out to illustrate the influence of the parameters α and γ on the performance. The results presented here are not exhaustive, but show how these parameters can be optimised according to the operating conditions.

Figure 3 shows the impact of parameter α under different bit allocations for 0% and 15% Frame Error Rate (FER) and γ=0.5 (equal importance given to individual and combined descriptions). Bit allocations of 7-3, 6-4 and 5-5 bits were used. The interesting points are the minima of the MSE, and the value of α for which they are obtained. These optimum α values are α=0.5, α=0.3, α=0.1 for the 5-5, 6-4 and 7-3 schemes respectively. As expected, the optimal value of α follows the imbalance between the descriptions.

In Figure 3A, assuming the optimal value of α is used, the best performance under 0% FER is obtained by the configuration 7-3, which significantly outperforms the other cases. In Figure 3B however, it is clear that the best configuration under 15% FER is obtained for the 5-5 configuration, again assuming optimal α.

This clearly shows that the optimal bit allocation is dependent on the FER, and therefore the system should be designed based on the expected FER. It can also be noted that the optimal value of the parameter α seems fairly independent of the FER, and seems mostly linked to the chosen bit allocation.

Table 1. MSE when varying parameter γ for different FER

Parameter γ	Frame Error Rate (%)			
	0	10	20	30
0	0,0613	0,0937	0,1634	0,2777
0,5	0,050	0,0853	0,1572	0,2739
1	0,0173	0,3235	0,6152	0,9023

For the evaluation of the impact of parameter γ we decided to use a set of system parameters which gives an overall good performance: a balanced bit allocation of 5-5 bits and a value of 0.5 for the parameter α. Table 1 shows the results for a range of values of γ. The trade-off in performance between the combined (γ=1) and uncombined (γ=0) descriptions is evident for each distinct FER. The best quality is achieved for γ=1 at 0% FER, giving a quality equal to that of the 8-bit direct quantiser. However, keeping γ=1 for increasing FER results in a heavily degraded signal. For FER greater than 0 intermediate values of γ give good performance (e.g. FER=10%, γ=0.5, MSE=0.0853). The parameter γ controls the trade-off in performance between the quality of the individual descriptions against that of the combined.

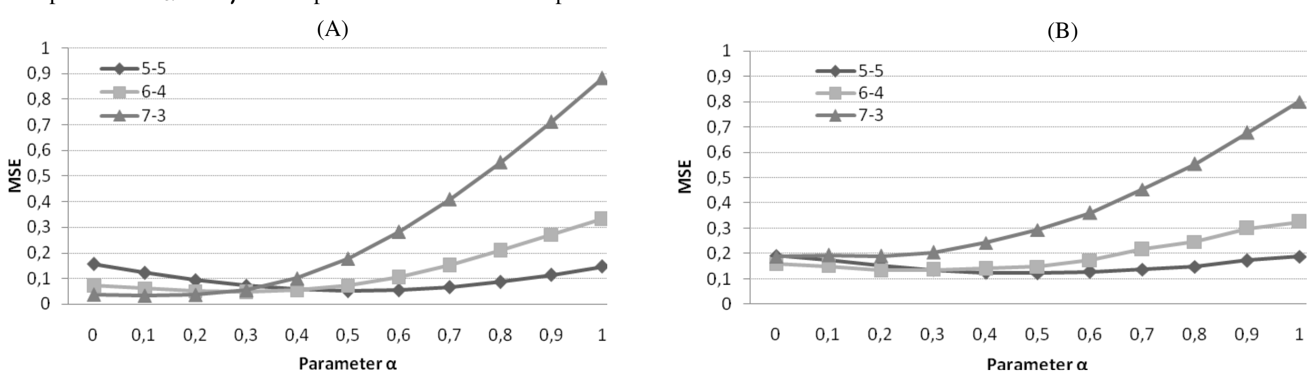

Fig. 3. Effect of varying parameter α for different bit allocations for (A) 0% FER and γ=0.5 (left) and (B) 15% FER and γ=0.5 (right)

3. CODEBOOK DESIGN

3.1. Theory

The codebooks for the two descriptions can be trained independently, using classic training algorithms such as LBG [8] [9]. However, the resulting codebooks are unlikely to be optimal for the case when both descriptions are combined. The codebooks have to be designed jointly for maximum performance in the combined case.

To achieve this, we propose a modified version of the LBG algorithm. In a standard LBG algorithm, each codeword is updated using the centroid of the cluster of training vectors it represents. In the modified algorithm, the codewords C_A and C_B have to be updated so as to minimise:

$$E = \sum_x \{\gamma_t \cdot [\alpha_t \cdot C_A(i) + (1-\alpha_t)\cdot C_B(j) - x]^2 + (1-\gamma_t)\cdot[(C_A(i)-x)^2 + (C_B(j)-x)^2]\}$$

where α_t and γ_t denote the weighting parameters used during training.

As a result, the elements x in the training database are clustered according to which codewords are used to represent them, and they can then be updated as:

$$C_A^{'}(i) = \gamma_t \cdot \underset{x\in S_i^{AB}}{E}\left\{\frac{1}{\alpha_t}\cdot[x-(1-\alpha_t)\cdot C_B(j_x)]\right\} + (1-\gamma_t)\cdot \underset{x\in S_i^{A}}{E}\{x\}$$

$$C_B^{'}(j) = \gamma_t \cdot \underset{x\in S_j^{AB}}{E}\left\{\frac{1}{1-\alpha_t}\cdot[x-\alpha_t\cdot C_A(i_x)]\right\} + (1-\gamma_t)\cdot \underset{x\in S_j^{B}}{E}\{x\}$$

Where S_i^A and S_j^B are the clusters of training vectors represented by and respectively for the case of individual descriptions, and S_i^{AB} and S_j^{AB} are the clusters corresponding to the combined descriptions.

The codewords are updated as a weighted average of the updated codewords according to the combined description, and that according to the individual descriptions, in accordance with the chosen quantisation criteria.

The parameter α_t in the distance minimisation criterion affects the balance between the weighted codewords from the two quantisers. The codebook training equation implies interdependencies and previous knowledge for the codewords of both quantisers. In other words, for a given quantiser A (resp. B), in order to optimise the codebook entries, for each input sample x there must be one matching codeword from quantiser B (resp. A) that minimises the parametric quantisation selection.

3.2. Experimental results

In order to evaluate the performance of the optimised quantisers, we compared the following systems:

- Simple MDC: two independent descriptions are used. When both descriptions are received correctly, only the highest rate one is used. In our experiments a 5-5 bit allocation was used.
- Single Description (SD) Coding: a single n bit description is used. 5 bits were used in our experiments.
- Proposed MDC: the proposed MDC system with a bit allocation of 5-5 and the weighting parameters set at α=0.5, γ=0.5.
- Optimised: the proposed MDC system where the codebooks are designed jointly, for 0% FER. 5-5 bit allocation was used.

Table 2. MSE for different schemes

Method	**Frame Error Rate (%)**			
	0	**10**	**20**	**30**
Proposed MDC	0,0430	0,0876	0,1567	0,2844
Optimised	0,0128	0,3549	0,6385	0,9058
Adaptive	0,0148	0,0810	0,1533	0,2569
Simple MDC	0,1186	0,1358	0,1992	0,2812
SD	0,1214	0,3137	0,5147	0,6963

Our experiments focus on the effect of parameters γ/γ_t and the effect of different bit allocations. In Figure 4 and Table 3 the results of optimization under different conditions are presented. Varying values (0.5 and 1) of parameters γ/γ_t (Table 3), when codebooks are optimised, give better performance when compared to the ones of the proposed MDC scheme (Table 2). It is evident from Figure 4 that schemes utilising $\gamma= \gamma_t$=0 and 0.5 are more robust for packet loss rates greater than 5% while $\gamma= \gamma_t$ =1 seems to be optimum under no packet losses.

Table 3. MSE for optimised codebooks for α=0.5

γ_t/γ/Iteration/ Bit allocation	**Frame Error Rate (%)**			
	0	**10**	**20**	**30**
1,1, 3, 5-5	0,0143	0,3455	0,6555	0,9316
0.5,0.5, 3, 5-5	0,0454	0,0854	0,1526	0,2615

Compared with the adaptive method, the optimised method has slightly better performance (Table 2, 0.0148 MSE). This improvement disappears as FER goes up, as the optimised method has been optimised for 0% FER. In conclusion, for high performance at minimal packet loss rates (around 0%) schemes utilising γ close to 1 and unbalanced bit allocations (e.g. 6-4, 7-3, etc) are best. On the other hand, for optimal performance under increasing packet loss rates (FER >5%), balanced bit allocations using values of γ around 0.5 give the best results.

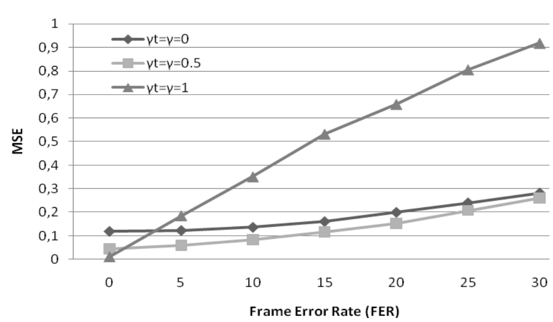

Fig 4. MSE for optimised codebooks for different γ, γ_t and $\alpha=\alpha_t$=0.5 with respect to different FER.

4. ADAPTIVE QUANTISATION

In an adaptive quantisation scheme, the parameters α and γ vary depending on the packet loss rates to minimise the distortion. This assumes that the system is able to estimate the amount of packet losses, e.g. through a feedback channel. Different sets of values have to be evaluated under different conditions, so that the best performing configurations for each condition can be selected.

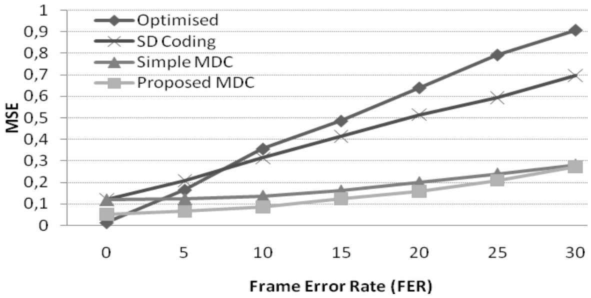

Fig. 5. Overall comparison of performance for different methods for γ=0.5 (γ_t=0.5), α=0.5 and varying FER

Our simulations showed that for different FER, different values of γ are optimum. When FER increases, γ decreases in order to compensate for the growing number of individual descriptions that arrive degraded. The results are presented in Figure 6. It is clear from the plot that an adaptive scheme offers substantial performance improvement over the basic scheme. For 0% FER this improvement translates to 2 bits per sample, while for 5% FER it is 1 bit. Increasing FER, over 5% gives performance close to that of the basic scheme. It is important to note that the decoder does not have to be aware of this adaptation, as it only affects the codeword selection process and not the codebooks, which makes the system more robust, as there is no risk of loss of synchronization.

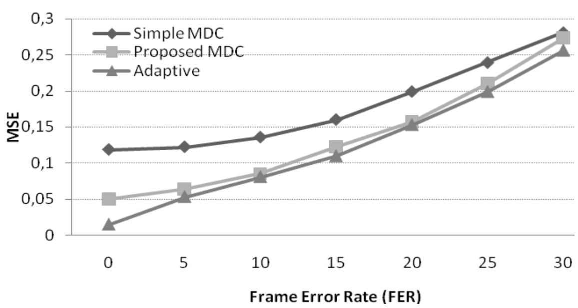

Fig. 6. Comparison between the proposed adaptive MDC, proposed MDC and simple MDC

Overall, our results for the adaptive scheme show that the performance achieved by combining two 5 bits descriptions is that of a 9 bits direct quantiser for an MDC system with no losses in the network (Figure 5). This shows the proposed technique only adds a small amount of overhead over SDC, but in return offers significantly improved performance under packet losses.

5. CONCLUSION

In this paper, we proposed a quantisation scheme for MDC and studied several factors affecting its performance. This scheme involves processing at parameter level, quantising them as a weighted combination of the codewords from the individual quantisers. The codebook selection is performed with two weighting parameters, which account for the balance between different quality descriptions, and between the quantisation error of the individual and combined descriptions, and affect the overall performance of the system in error free and error prone conditions.

The presented scheme was evaluated under different conditions and different approaches, and a specific codebook training algorithm was described. The results show that the proposed MDC system can provide significant improvements in error performance with only limited overhead. This approach offers an effective way of using the channel capacity of a communication system as well as a way for more robust speech communications with efficient use of the added redundancies. Further research will be carried out to apply these results to actual speech coding systems under realistic error conditions.

6. REFERENCES

[1] M. Lee, S. Jung, H. Kang, Y. Park and D. Youn, "A Packet Loss Concealment Algorithm Based on Time-Scale Modification for CELP-type Speech Coders," in Proc. of the International Conference on Acoustics, Speech and Signal Processing ICASSP 2003, vol. 1, pp. 116-119, 2003.

[2] S. Praestholm, S. S. Jensen, S. V. Andersen, M. N. Murthi, "On packet loss concealment artifacts and their implications for packet labelling in voice over IP," in IEEE International Conference on Multimedia and Expo (ICME '04), vol. 3, pp. 1667 – 1670, 2004.

[3] Jan Linden, "Achieving the Highest Voice Quality for VoIP Solutions," Global IP Sound Inc, White paper.

[4] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," Bell Syst. Tech. J., vol. 59, pp. 1417–1426, 1980.

[5] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in Proc. of VCIP'99, 1999.

[6] A. Ingle and V.A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," IEEE Trans. Speech Audio Processing, vol. 3, pp. 48-57, January 1995.

[7] E. Orozco, S. Villette and A. M. Kondoz, "Multiple Description Coding for Voice over IP using Sinusoidal Speech Coding," in Proc. of the International Conference on Acoustics, Speech and Signal Processing ICASSP 2006, pp. 9-12, 2006.

[8] Y. Linde, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantizer Design," IEEE Transactions on Communications, Vol. COM-28, pp. 84-95, 1980.

[9] J. Makhoul, S. Roucos, and H. Gish, "Vector quantisation in speech coding," in Proc. of IEEE, vol. 73, issue 11, pp. 1551 - 1588 , November 1985.