Multiple Description Coding for Voice over IP using Sinusoidal Speech Coding

E. Orozco, S. Villette and A.M. Kondoz

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

ABSTRACT

CELP coders, such as G.729, are often used in VoIP systems as they offer good speech quality in the absence of packet losses. However, their reliance on long-term prediction causes propagation of errors across speech frames, and therefore makes CELP coders more sensitive to packet losses. Sinusoidal coders on the other hand do not rely on long-term prediction, and may be a good alternative for VoIP due to their higher resilience to packet losses. In this paper a comparison is made between CELP and sinusoidal coders in a VoIP application. A packetisation scheme based on Multiple Description Coding (MDC) applied to the sinusoidal coder is presented. The results show that under typical VoIP operating conditions, the sinusoidal coder based systems can outperform CELP based systems at equal bit rate, especially for high packet loss rates.

1. INTRODUCTION

In real-time transmissions, IP networks are unreliable and offer a best-effort delivery service with no QoS guarantees. Thus, packets may be lost. When speech coders are used in VoIP applications, loss of packets can lead to considerable degradation of speech quality. Missing packets usually cannot be retransmitted as this would imply excessive delay and more network congestion. Error concealment techniques are necessary to cope with packet losses and improve the speech quality [1].

MDC based schemes are promising when used to combat the effects of packet loss. MDC is a technique where redundant information, called descriptions, is carried in each packet and from which packet losses are reconstructed. These descriptions contain a coarsely quantised version of neighbouring frames. The main drawback of most MDC schemes is an increase in delay and bit rate. The higher the number of descriptions, the better the performance of the MDC technique in terms of speech quality, but the higher the delay and bit rate.

CELP coders are commonly used in VoIP applications. They rely heavily on long-term prediction (LTP), which help provide good speech quality, but allows propagation of errors across frames, making them sensitive to packet losses. Sinusoidal coders on the other hand do not rely on LTP, making them potentially superior for VoIP applications, especially for high packet loss rates.

For comparison purposes, a realistic CELP-based reference system was implemented, based on typical VoIP implementations (Cisco and Nortel Networks) [2] [3]. The chosen system uses the speech coder G.729 [3] [4] operating at a bit rate of 8 kbps with a frame size of 10 ms. Two 10 ms frames are sent in each packet resulting in a payload size of 20 bytes (160 bits).

This is compared to a sinusoidal speech coder based system. The commonly available sinusoidal coders, such as MELP 1.2/2.4 kbps, operate at bit rates too low to offer speech quality comparable to G.729 at 8 kbps. Therefore the SB-LPC coder (Split-Band Linear Predictive Coding) [5], which can offer high quality speech at 4 kbps, was found to be more appropriate to carry out the comparison. It is expected that the overall trend of the results obtained will hold true for any classic sinusoidal coder of similar bit rate and quality to that used in this paper.

2. NETWORK SIMULATION

A simulation of a VoIP network was carried out using NS-2 [6] to obtain realistic patterns of packet losses, since the loss statistics affect the performance of MDC systems.

Figure 1 shows the topology used in the simulation. It is a common configuration where traffic is generated in some nodes of an Ethernet network and forwarded to a shared output link. This output link is represented by an E1 link between the edge node of the Ethernet network and an edge node belonging to another network. Packet losses occur at the bottleneck of this link.

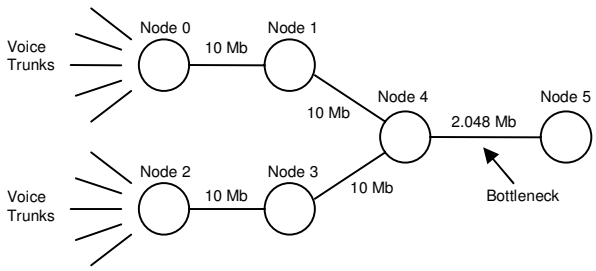

Figure 1: Topology used in the simulation

The simulation was carried out with two traffic sources: one for voice traffic, modelled by a constant bit rate source, and one for background traffic, which was modelled by a Pareto On-Off

source. It was configured to generate traffic occupying approximately 90% of the bandwidth of the output link, which represents the proportion of the TCP controlled Internet traffic [7].

From the simulation, error patterns of packet loss rates between 5% and 30% were obtained, covering the realistic range of operating conditions. A study of the histograms of burst lengths showed that even at high bit rates, bursts of three or more packet losses occur infrequently. This is used in the design of the MDC scheme.

3. PROPOSED PACKETISATION SCHEME

The proposed approach consists of an MDC based packetisation scheme at parameter level where each packet contains redundant information about future neighbouring frames. This redundant information does not have the quality of the original one as it is more coarsely quantised but it helps to reconstruct speech when a packet is lost. The total amount of payload bits in one packet is limited to 160 bits, i.e. 8 kb/s and 20 ms frames, in order to have consistency with the packet size defined in the G.729 based reference system.

The level of burstiness shown by the error patterns indicates that two coarse descriptions in addition to one fine description would cover most of the packet losses while only requiring 40 ms of extra delay. This was selected as a suitable trade-off.

Since it is intended that the whole system with SB-LPC operates at 8 kbps, this paper proposes to have one finely quantised description at 5 kbps, one coarsely quantised description at 2 kbps and another at 1 kbps. The fine description at 5 kbps is required in order to provide good speech quality in no-error conditions. The configuration (5, 2, 1) allows a coarse description at 2 kbps with reasonable quality to recover single packet losses which are the most common, and a very coarse description at 1 kbps for larger bursts. The packetisation scheme is shown in Figure 2.

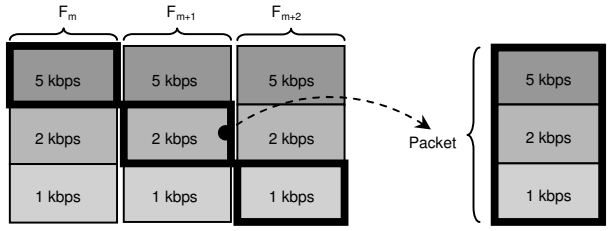

Figure 2: Packetisation of descriptions in a (5,2,1) configuration

Figure 3 shows how the proposed MDC system works. The number of packet losses that can be recovered is limited by the number of descriptions. Extrapolation techniques are used to conceal packet losses that are beyond the limits of the MDC system.

4. QUANTISATION

The SB-LPC uses five parameters to represent speech: 10^{th} order LP coefficients in LSF domain, pitch period, voicing cut-off frequency, speech energy and spectral amplitudes. The bit allocation required to quantise each parameter is given in Table 1. A frame size of 20 ms and parameter update rate of 10 ms are assumed.

Parameter	Number of bits per 20 ms		
Bit rate	5 kbps	2 kbps	1 kbps
LSF	42	24	10
Pitch	-	7	5
Voicing	-	3	2
Joint Pitch and Voicing	12	-	-
Energy	10	6	3
Spectral Amplitudes	36	0	0
Total number of bits	100	40	20

Table 1: Bit allocation for SB-LPC descriptions

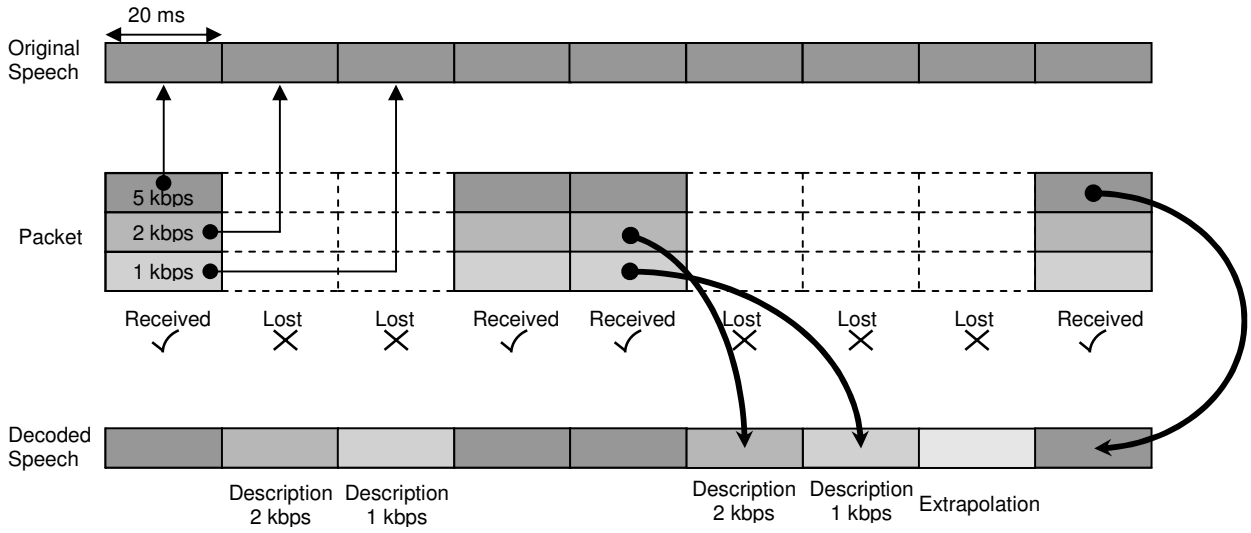

Figure 3: Process of packets recovery through the proposed MDC scheme

The configurations presented in Table 1 assume no predictors in the quantisation process, as using inter-frame prediction such as Moving Average (MA) leads to a decrease in bit rate, but also to lower packet loss resilience. The configurations at 2 kbps and 1 kbps were derived from that at 5 kbps.

The different descriptions are packetised so that a packet contains the 5 kb/s description of frame F_m , the 2 kbps description of the following frame F_{m+1} , and the 1 kb/s description of frame F_{m+2} , as illustrated in figure 3.

In order to improve quantisation performance, inter-frame prediction is commonly used in speech coders, at the risk of increased error sensitivity through error propagation. Moving Average (MA) is generally used as it limits error propagation, while the more efficient Differential Quantisation (DQ) schemes propagate them [8]. However, as a given packet contains descriptions of successive packets, it is possible to use a first order DQ to predict F_{m+1} from F_m, and a second order DQ to predict F_{m+2} from F_{m+1} and F_m. By using the descriptions present in the same packet for prediction, there is no risk of error propagation across frames.

The p^{th} order DQ predictor is given by

$$\tilde{x}_n^k = \sum_{l=1}^{p} \alpha_n^l \hat{x}_n^{k-l}$$

where $\tilde{x}_n^k$ is the prediction vector for the k^{th} frame, $\hat{x}_n^{k-l}$ is the vector containing the quantised values of the parameters corresponding to the p previous frames and α_n^l are the prediction factors.

Figures 4 and 5 show the quantisation error for the DQ and no prediction quantisers for the energy and LSF respectively. It can be seen that a large bit saving of the order of 40% is obtained through the use of the proposed DQ scheme, at no cost in terms of error sensitivity, making it very suited for use in MDC schemes.

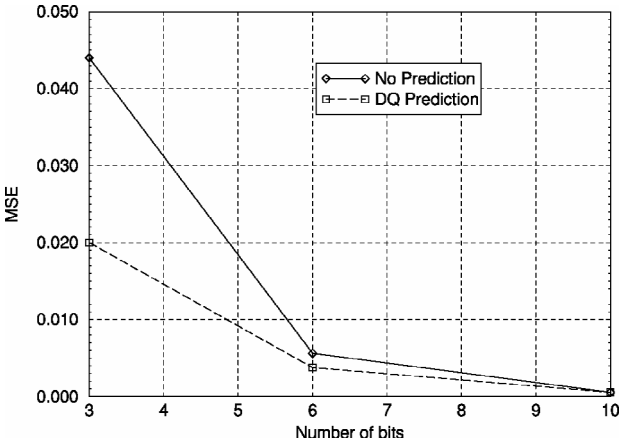

Figure 4: Energy quantisation. MSE vs number of bits.

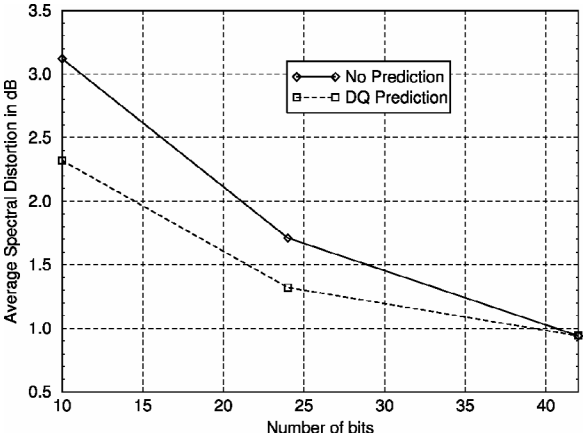

Figure 5: LSF quantisation. SD vs number of bits.

5. QUALITY EVALUATION

To evaluate the quality of the proposed method an objective test using PESQ [9] was carried out. PESQ was used as a large number of conditions need to be tested which would be impractical with a subjective MOS test. Informal listening tests indicate that the PESQ scores shown here appear to be realistic and provide a good indication of actual MOS scores.

Six speech samples (three male speakers and three female speakers) with duration of eight seconds were degraded by packet losses, and reconstructed according to various configurations and assessed. The configurations evaluated were:

- G.729 with standard error concealment (8kbps)
- G.729 with MDC (16 kbps)
- SB-LPC without MDC (5 kbps)
- SB-LPC using MDC with no prediction (8 kbps)
- SB-LPC using MDC and DQ prediction (8 kbps)

The error patterns, corresponding to 32 seconds of transmission, were obtained through the NS-2 simulation.

Figure 6 shows the results obtained from the objective tests. In error-free conditions, G.729 scores higher than SB-LPC, whereas informal listening indicated their quality to be similar. This may be due to the limitations of PESQ when using non-CELP coders, which do not attempt to match the original signal. However, in the presence of packet losses, SB-LPC performs significantly better than G.729. The quality of G.729 drops rapidly as packet losses increase, showing poor performance at rates above 15% of packet loss. It can be seen that the quality of G.729 at packet loss rates of around 10% is equivalent to the quality of SB-LPC without MDC at 15% packet loss, and also similar to the quality of SB-LPC using MDC with DQ prediction at 30% packet loss

From these results it can be concluded that performance under packet loss conditions is quite unrelated to error free performance, and that the error resilience of the speech coder can be far more relevant to the overall performance of a VoIP system in typical operating conditions than the error free speech quality.

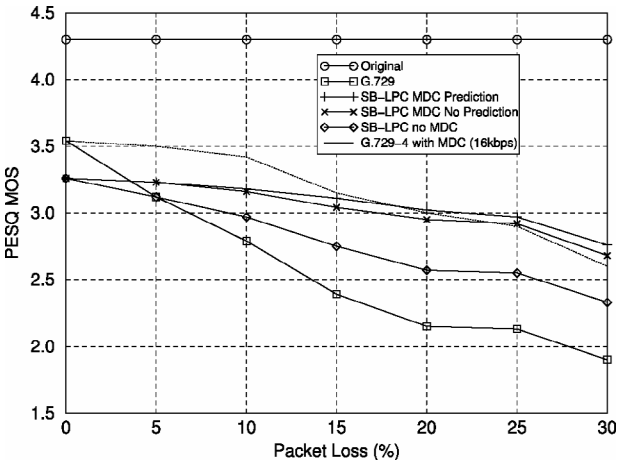

Figure 6: Objective test results

It can also be seen from Figure 6 that the use of prediction significantly improves the quality of the speech under packet loss conditions. The improvement increases with the packet losses, which is explained by the fact that only the second and third description at 2 and 1 kbps respectively make use of the prediction. The higher the frame losses, the more often these descriptions are used, hence the higher the quality improvement from the use of prediction.

MDC methods to improve the speech quality of G.729 in presence of packet losses have been presented in [10]. The approach from [10] that offered the highest quality (G.729-4 at 16 kbps) was compared against the MDC method with prediction using SB-LPC. Figure 6 indicates that G.729-4 offered better quality at packet loss rates between 0% and 15%, while SB-LPC with MDC provided better performance at higher packet loss rates. It can be noted that G.729-4 uses an additional delay of 20 ms and operates at 16 kbps whilst SB-LPC with MDC requires additional delay of 40 ms and operates at 8 kbps. This shows that despite the use of MDC and high bit rates, the lack of error resilience of CELP coders is a clear limiting factor of quality in packet loss environments.

An alternative to G.729 for VoIP applications is the Internet Low Bit Rate Coder (iLBC) [11]. It is designed to stop errors from packet losses from propagating across frames, and therefore is expected to perform better than standard CELP under packet loss conditions. However, it operates at a significantly higher bit rate (e.g. 15.2 kbps) than the 8 kbps systems studied here, and it was shown in [10] to be very significantly outperformed by the G729 with MDC at 16 kbps used here for comparisons, which performs similarly to the proposed system for packet losses equal to or higher than 15%. Therefore iLBC was not considered in this study.

6. CONCLUSION

This paper has presented a MDC method to conceal packet losses, and a comparison between a standard CELP-based system and a proposed sinusoidal coding based system for VoIP applications. A reference CELP system based on the G.729 coder was compared to a SB-LPC coder with MDC at parameter level, for the same overall bit rate of 8 kbps.

Tests on realistic VoIP error patterns showed that the proposed system outperforms the reference CELP system when the packet losses exceeded 5%, as the higher clean speech quality of the CELP system cannot compensate for its high error sensitivity under packet loss conditions. It was also shown that it is possible to use very efficient inter-description prediction in the MDC system for sinusoidal coding, leading to a bit saving of approximately 40% for the coarse descriptions. Finally, the proposed system was shown to compare favourably at high packet loss scenarios to a reference G.729 system with MDC, despite using only half of its bandwidth.

Overall, this paper shows that the use of a sinusoidal coder can give higher performance than CELP systems for VoIP applications, especially when the available bandwidth is limited, and provide acceptable speech quality at 8 kbps with up to 20-25% packet loss rate.

7. REFERENCES

[1] B. W. Wah, X. Su, D. Lin, "A Survey of Error-Concealment Schemes for Real-Time Audio and Video Transmissions over the Internet" *in Proceedings of the International Symposium of Multimedia Software Engineering*, pp. 17-24, 2000.

[2] Cisco, "Voice over IP – Per Call Bandwidth Consumption", Technical notes 7934, 2002.

[3] Nortel Networks, "The G.729 Speech Coding Standard", Technical brief, 2000.

[4] ITU-T, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited-Linear-Prediction (CS-ACELP)", Recommendation G.729, International Telecommunications Union, Geneva, 1996.

[5] S. Villette, M. Stefanovic, A. Kondoz, "Split Band LPC Based Adaptive Multi-Rate GSM Candidate", *in Proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP 1999*, pp. 249-252, 1999.

[6] K. Fall, K. Varadhan, "The ns Manual", Technical report, UC Berkeley, 2003.

[7] S. Georgoulas, G. Pavlou, P. Flegkas, P. Trimintzios, "Buffer and Bandwidth Management for the Expedited Forwarding Traffic Class in Differentiated Services Networks" *in Proceedings of the London Communications Symposium (LCS)*, pp. 349-352, 2002.

[8] J. Skoglund, J. Linden, "Predictive VQ for noisy channel spectrum coding: AR or MA?" *in Proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP 1997*, pp. 1351 - 1354, 1997

[9] ITU-T, "Perceptual evaluation for speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", Recommendation P.862, International Telecommunications Union, 2001.

[10] R. Lefebvre, P. Gournay, R. Salami, "A Study of Design Compromises for Speech Coders in Packet Networks" *in Proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP 2004*, pp. 265-268, 2004.

[11] S. V. Andersen, et al. "ILBC – A linear predictive coder with robustness to packet losses", *in Proceedings of 2002 IEEE Speech Coding Workshop*, Tsukuba, Japan, pp. 23-25, 2002.