MODEL-BASED PACKET LOSS CONCEALMENT FOR AMR CODERS

Prof. C.S.Xydeas and F.Zafeiropoulos
Department of Communications Systems, Lancaster University,
Lancaster LA1 4YR, UK
Email: c.xydeas@lancaster.ac.uk

ABSTRACT

In this paper a general packet Loss Correction / Concealment signal recovery framework is proposed for parametric speech coders. Both redundancy-based Forward Error Correction (FEC) and Receiver Only (RO) techniques are considered in conjunction with the Adaptive Multi-rate Coder (AMR). The robust, low bit rate, high communications speech quality Manchester Pitch Synchronous (MPS) Coder is employed in the proposed systems as a secondary coding process.

Thus the performance of AMR/MPS-FEC/RO packetised speech transmission systems is considered. Subjective and objective computer simulation results clearly indicate the superiority of the proposed schemes over conventional AMR based systems, particularly at relatively high (> 5%) packet loss rates.

1. INTRODUCTION

The most common problem encountered in the real time transmission of voice over packet-based networks, e.g. the Internet or forthcoming 3rd Generation Mobile Networks, is packet loss. Two main methodologies can be used to combat the adverse effects of packet loss namely Forward Error Correction (FEC) and Receiver Only (RO) processing. In the first case, the output of a robust "secondary" low bit rate codec is transmitted as additional "side information" and used in the recovery of missing speech segments [1]. This of course introduces an increase in the overall transmitted bit rate and possibly an extra delay. In contrast, RO techniques exploit the relatively slow evolution characteristics of the information bearing parameters of certain model based speech coders and attempt to extrapolate missing segments of information from adjacent parts of the signal [5]. Both approaches can be viewed as complementary to traditional source coding/channel coding arrangements.

This paper examines the performance of the recently established AMR speech coding standard, while operating in a packetised transmission mode and under a range of packet loss rates. Furthermore, AMR based systems are proposed incorporating FEC and/or RO packet error concealment strategies which in turn are based on the effective and robust Manchester Pitch Synchronous (MPS) [9,10] secondary coder.

Section 2 outlines the MPS coding approach and MPS based packet loss concealment. Section 3 provides a brief outline of the AMR system and its inherent RO error concealment mechanism. Two new AMR/MPS based packet loss correction schemes are also described in this section. Experimental results obtained from the proposed AMR/MPS-FEC and AMR/MPS-RO schemes, using informal subjective and objective PESQ measurements, are presented in section 4. Section 5 provides conclusions and closing remarks.

2. MPS CODER.

The Manchester Pitch Synchronous speech coder (MPS) employs an advanced mixed excitation model within a hybrid Sinusoidal/Prototype Waveform Interpolation coding framework. The system is capable of high communications quality speech (in excess of 3.6 MOS and 3.4 PESQ at 2.4 Kbs) while operating in the region of 2.5 to 1.5 Kbs. A detailed system description can be found in Ref. [10].

The encoding process operates on successive 20ms frames which are classified as Voiced ($V_n=1$) or Unvoiced ($V_n=0$), and produces the following parameters for transmission:

When the nth frame is Voiced:

- Voiced/Unvoiced decision V_n
- Pitch Period P_n
- Quantized LPC filter coefficients.
- Residual signal magnitude information MG_j^n
- Five Hybrid Excitation binary flags hv_j^n

When the nth frame is Unvoiced:

- Voiced/Unvoiced decision V_n
- Quantized LPC filter coefficients.
- Quantized RMS value of the Residual signal energy $\sqrt{\hat{E}_n}$.

Note that the MG_j^n values are quantized using either:

- a single value representation and an adaptive μ-law logarithmic 5 bit quantiser (Modified Single Value Spectral Amplitude Representation MSVSAR [9]), as in the 1.5Kbs scheme or

- a high quality Variable Size Spectral Vector Quantiser (VS/SVQ) [10] at 19 bits/frame, as in the 2.4Kbs system.

2.1 MPS Packet Loss Concealment

A Packet Loss Correction and Concealment framework for the MPS coder is presented in detail in [11]. This includes both redundancy-based FEC techniques and RO error concealment techniques. MPS-FEC can be designed for maximum redundancy, where all the MPS parameters are transmitted in addition to the MPS primary coder bit stream. Alternatively, only a subset of MPS parameters can be selected for transmission as "side information", in which case the amount of added redundancy is reduced. The values of those system parameters which are not transmitted are effectively "estimated" at the receiver. Note that the extreme case of zero redundancy corresponds to RO concealment.

In this paper, the MPS FEC redundancy used in conjunction with the AMR system includes, on top of all MPS information, an extra 7 bits coarsely VQ Quantized (per 20ms) LSP index. When packets are missing, LSP coefficients are recovered according to a first order Markov model. Missing Voiced/Unvoiced flags are predicted using a set of rules derived through statistical observations of MPS behaviour, whereas the Pitch Period P_n and Spectral Magnitude MG_j^n information is estimated using a repetition/attenuation process. Missing hybrid excitation flags hv_j^n are also assumed to be those of a fixed predetermined hv_j^n pattern. In the case of Unvoiced frames, the RMS of the residual signal energy $\sqrt{E_n}$ is recovered using repetition/attenuation.

3. ADAPTIVE MULTIRATE CODER (AMR)

The AMR [3] speech coder is based on the CELP coding paradigm and is in principle able to switch its bit-rate, every 20 ms, to one of eight source coding rates within the 4.75 Kbs to 12.2 Kbs range. In addition, no-speech-activity frames are represented by a low bandwidth Silence Description (SID) mode.

3.1 AMR Error Concealment framework

AMR error concealment [2] includes "Active" and "Passive" concealment in the form of parameter repetition/substitution, and attenuation/muting. Thus AMR error concealment strategy is based on gradual attenuation and a "graceful" degradation towards low level comfort noise, within a period of approximately 80 ms after the occurrence of packet loss. This approach can be damaging to speech quality /intelligibility, particularly when several successive packets are not available at the decoder and thus large segments are missing from the recovered speech signal. The use of MPS-FEC or RO, to enhance AMR's error concealment characteristics, was motivated by the observed robust performance of MPS-FEC/RO schemes operating at high packet loss rates (up to 20%).

3.2 AMR/MPS-FEC

The proposed AMR/MPS-FEC scheme is illustrated in Figure 1. AMR is used as the primary coder (in any of its 8 modes) and MPS as the secondary coder. In the case of packet loss, at the receiver, packet loss concealment is performed by MPS and the reconstructed signal is fed into an additional AMR encoder. The resulting AMR-encoded stream is then fed into the AMR decoder for final speech reconstruction.

In the absence of packet loss, the AMR and MPS decoders operate in parallel, with the former in its "normal" mode using the received AMR bit stream, and the latter using the MPS stream in order that MPS model parameters are updated. There is no interaction between the two speech coders at the parameter level. A Bad Frame Indicator (BFI) effectively selects one of the two processing paths, according to the availability of a packet.

Notice that at the transmitter, the two speech encoders work "independently" on the same 20 ms frame basis. The resulted compressed data streams can either be accommodated into the same packet (single flow) or

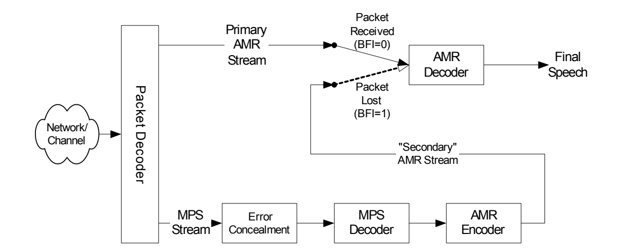

Figure 1: General block diagram of the AMR/MPS-FEC decoder

transmitted via separate packets (twin flow). The single flow approach offers simplicity and bandwidth savings in terms of network overheads. There is already a proposed framework for AMR packet transmission over IP [8] which involves the Real Time Protocol (RTP) packet format [6] and includes FEC redundancy payload. The AMR's ability of rate adaptation every 20 ms can be used to vary the size of the primary speech payload and allow more redundancy (i.e. extra MPS parameters) to be transmitted when poor channel or network overload conditions are detected.

Also, the secondary coder frame size may differ from that of the primary coder, (e.g. AMR 20 ms and MPS 80 ms). This scenario offers flexibility since it allows the secondary coder to transmit every "m"th primary frame. The bit rate of the primary coder can be reduced for those packages which carry information from both coders, thus maintaining, in single flow configurations, a constant bit rate per packet. However in this case an extra delay is introduced at the decoder.

The twin flow approach on the other hand can be advantageous in a DiffServ [4] environment, where some flows enjoy better Quality of Service (QoS) than others. The MPS FEC stream may travel in a higher class of service flow than the primary payload thus providing reliable packet loss concealment.

3.3 AMR/MPS-RO

A general block diagram of the AMR/MPS-RO system, which employs at the receiver both an MPS encoder and decoder, is shown in Figure 2. When packet P_{n-1} is correctly received, it is processed via the primary AMR decoder whose output is the S_{n-1} speech segment i.e. $\overline{\overline{S}} = S_{n-1}$. S_{n-1} is also fed in a "feedback" loop where it is MPS encoded, decoded, and AMR encoded. In this case the resulting $\hat{P}_{n-1}$ bits are discarded. However, when

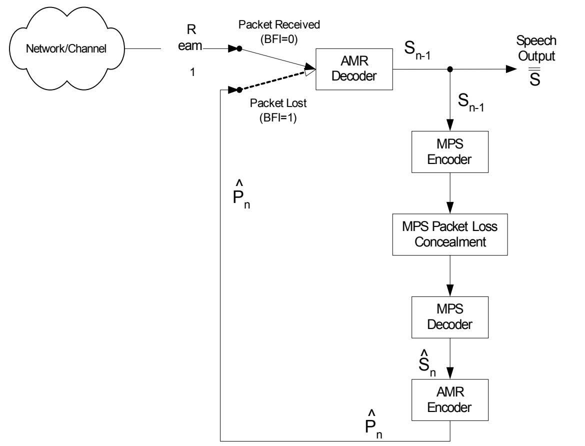

Figure 2: General block diagram of AMR/MPS-RO

packet P_n is lost, the MPS packet loss Concealment process, employed in the feedback loop of the system, is activated and provides the information needed by the following MPS decoder to produce $\hat{S}_n$. This signal is then AMR encoded and the resulting $\hat{P}_n$ packet estimate of P_n is now available to the primary AMR decoder and thus used to produce the $\overline{\overline{S}}$ final speech output. Note that $\overline{\overline{S}}$ is also fed back to the MPS encoder.

4. EXPERIMENTAL RESULTS

The proposed AMR/MPS-FEC and AMR/MPS-RO schemes were tested at three different AMR bit rates i.e. 12.2, 7.95 and 5.15 Kbs. For simplicity, a primary coder packet length of 20 ms (i.e. 1 AMR frame/packet) was used. The MPS system also operated on 20ms frames at 2.4Kbs. Longer coding frame lengths can be used, resulting to more efficient MPS configurations, at the expense of introducing further delays into the system.

In the case of AMR/MPS-FEC, both the Single and Twin Flow approaches were tested. With Single Flow, the secondary coder suffered the same packet loss rate and pattern as the primary coder. This provided a direct comparison between the existing AMR system and AMR/MPS-FEC.

The Twin Flow approach assumed that the FEC flow enjoyed higher QoS than the primary AMR bit stream. Thus a packet loss rate for the secondary flow of 5% is assumed

Packet losses are simulated using a modified Gilbert model [1] which allows for the more frequent occurrence of a large number of successive lost packets. Furthermore system performance is measured in terms of MOS scores obtained via informal subjective tests and also in terms of objective PESQ [7] scores.

AMR/MPS-FEC outperforms AMR at every packet loss rate from 1% to 20%. The difference between the two methods becomes more pronounced for packet losses over 5%, with AMR/MPS-FEC achieving a maximum advantage of 1 PESQ point or 0.7 MOS points in the case of 20% packet loss rate (see Figure 3). As expected, the twin flow system performs slightly better than the single flow version, due to the lower packet loss rate of the secondary FEC flow and the fact that flow packet loss patterns are different.

The AMR/MPS-FEC and AMR/MPS-RO results of Figure 3 were obtained using the 7.95 Kbs AMR coding mode. Note that similar system performance was also obtained with the AMR coder operating at 12.2 and 5.15 Kbs..

The AMR/MPS-RO scheme gives generally lower quality speech than AMR/MPS-FEC and the difference between schemes is rather constant across all packet loss rates This AMR/MPS-RO reduction in performance is due to MPS operating on AMR decoded rather than "clean" speech. Furthermore at low packet loss rates (i.e. 1%-2%) conventional AMR outperforms AMR/MPS-RO in terms of both subjective and objective quality metrics. However AMR/MPS-RO outperforms AMR at packet loss rates higher than 10%. Note that at these high rates, longer chains of successive packets are lost and AMR/MPS-RO offers better reconstruction and thus speech intelligibility, than the AMR system with its

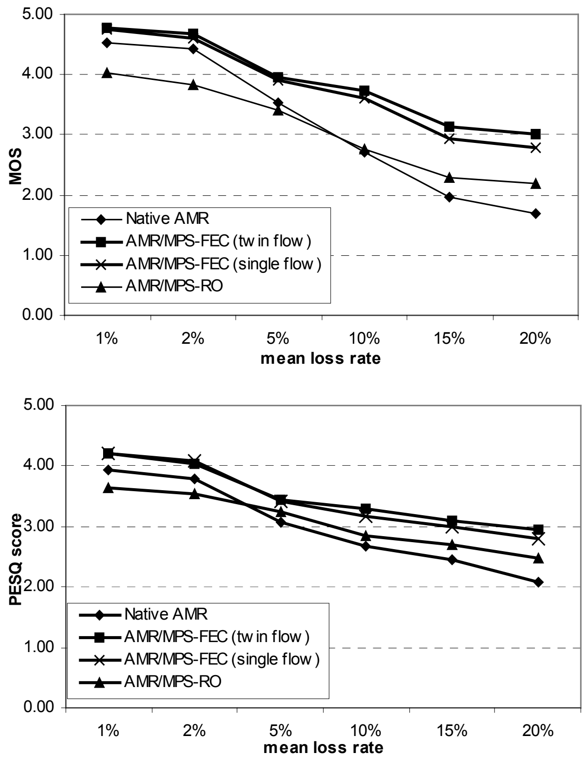

Figure 3: Packet loss concealment system performance comparison between AMR/MPS-FEC, AMR/MPS-RO and native AMR: (a) subjective quality Mean Opinion Scores (MOS); (b) Objective quality PESQ measurements. The AMR coder operates at 7.95 Kbps.

rather simple attenuation towards a comfort noise-floor strategy.

5. CONCLUSIONS

Objective PESQ measurements and informal subjective MOS results are broadly in agreement and indicate that AMR/MPS-FEC outperforms the conventional AMR system by 0.2 to 0.5 PESQ or MOS points at low to medium packet loss rates. In the case of medium to high packet loss rates (i.e. 10%-20%), AMR/MPS-FEC offers significantly better reconstruction quality (of the order of 1 PESQ and 0.7 MOS points) than AMR.

Although AMR/MPS-RO is outperformed by AMR at low to medium (1%-5%) packet loss rates, it offers significantly more intelligible speech reconstruction at medium to high packet loss rates (10% - 20%).

These results indicate the potential of the MPS speech modeling approach in providing an advanced packet loss concealment capability to conventional speech coders in general and the AMR system in particular. The benefits, in recovered speech quality, obtained from the resulting AMR/MPS hybrid systems are significant.

5. REFERENCES

[1] Bolot J.C., Fosse -Parisis S., and Towsley D., "Adaptive FEC-Based Error Control for Interactive Audio in the Internet", *Proc. IEEE INFOCOM*, New York, March 1999.

[2] 3GPP TS 26.091, Universal Mobile Telecommunications System (UMTS);: *Mandatory Speech Codec speech processing functions; AMR Speech Codec; Error concealment of lost frames*, ETSI

[3] 3GPP TS 26.090, Universal Mobile Telecommunications System (UMTS); *Mandatory Speech Codec speech processing functions; AMR Speech Codec; Transcoding functions,* ETSI

[4] Grossman D., "New Terminology and Clarifications for Diffserv", *IETF RCF 3260*, April 2002, http://www.ietf.org/rfc/rfc3260.txt

[5] Lindblom J. and Hedelin P., "Packet Loss Concealment based on Sinusoidal Extrapolation", *Proc. ICASSP 2002*, Vol. 1, pp. 173-176

[6] Perkins C., I. Kouvelas, O. Hodson, V. Hardman, M. Handley, J.C. Bolot, A. Vega-Garcia, S. Fosse-Parisis, "RTP Payload for Redundant Audio Data", *IETF RCF 2198,* September 1997, http://www.ietf.org/rfc/rfc2198.txt.

[7] Rix A. W., Beerends J. G., Hollier M. P. and Hekstra A. P., "Perceptual Evaluation Of Speech Quality (PESQ) – A New Method For Speech Quality Assessment Of Telephone Networks And Codecs", *Proc ICASSP 2001,* Vol. 2, SPEECH-P4.9

[8] Sjoberg J., Westerlund M., Lakaniemi A., Xie Q., "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", *IETF RFC 3267*, June 2002, http://www.ietf.org/rfc/rfc3267.txt

[9] Xydeas C.S., Papanastasiou C., "Efficient Mixed Excitation Models in LPC based Prototype Interpolation Speech Coders", *Proc ICASSP 97*, pp. 1555-1558

[10] "Speech synthesis System", Patent, International application number PCT/GB97/01831, International publication number WO 98/01848, 15 January 1998

[11] Zafeiropoulos F., *Packetised Low Bit Rate Speech Coding: An Investigation of Error Concealment Techniques,* PhD Thesis submitted to the School of Engineering, University of Manchester, 2002.