BANDWIDTH EFFICIENT AMR OPERATION FOR VOIP

Ingemar Johansson, Tomas Frankkila, Per Synnergren

Multimedia technologies, Ericsson Research, Ericsson AB,
Po Box 920, SE-97128, Luleå, Sweden
{ingemar.johansson, tomas.frankkila, per.synnergren}@epl.ericsson.se

ABSTRACT

An example of a bandwidth efficient Adaptive Multi Rate (AMR) system for Voice over IP (VoIP) is presented. In VoIP, packet losses cause degradation of the synthesized speech. The distortions may propagate over several consecutive frames, since predictors in the codec exploit inter-frame correlations to gain coding efficiency. To reduce the effects of packet loss, Forward Error Correction (FEC) that adds redundant information to voice packets can be used. However, while FEC can reduce the effects of packet loss, it will increase the amount of bandwidth used by the voice stream, which is not desirable. In this paper we propose FEC methods like partial redundancy, selective redundancy for the most sensitive frames and parameter interpolation in conjunction with AMR codec mode adaptation, which secure the speech quality when using AMR for VoIP without increasing the bandwidth substantially.

1. INTRODUCTION

Real-time voice transmission over IP is today used to a limited extent. But, as there is a trend to move to all-IP networks, voice transport over IP will become important in the near future.

The Adaptive Multi Rate (AMR) codec, [1], is standardized for mobile systems and a payload format for IP-transport has also been standardized in IETF, [2]. The major cause for speech quality degradation in IP-networks is packet loss. Packet loss usually occurs in routers due to congestion. Packets may also be dropped in the application, if they are received too late to be useful. While voice traffic can tolerate some amount of packet loss, a loss rate of a few percent may be harmful to the speech quality. The amount of packet loss that can be tolerated depends on the robustness of the used coding algorithm. The predictors in modern CELP codecs, such as AMR, exploit inter-frame correlations in order to achieve high coding efficiency. This causes the errors to become even more severe as they propagate over several frames.

In mobile systems with fixed bandwidth, AMR allows for varying the strength of the error protection depending on the channel condition, [3]. It will be shown that AMR codec mode adaptation can be used for packet switched networks to achieve both high speech quality for good network conditions and robustness against packet loss, without increasing the bit rate significantly.

In this paper, we present and evaluate an example of a transmission bandwidth efficient AMR system for VoIP. An overview of the system is shown in figure 1. The system utilizes four methods to improve the robustness: redundant packets, selective redundancy, parameter interpolation and partial redundancy. These methods are combined into various configurations to improve the robustness of AMR under varying IP-network conditions. The methods have been designed with the objective to maintain the low complexity of the codec.

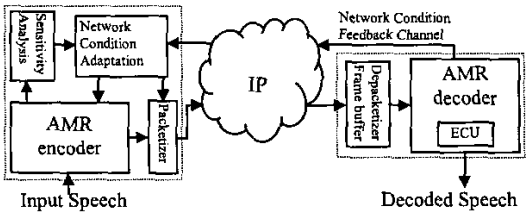

Figure 1 Adaptive Multi Rate system for Voice over IP

2. REDUNDANT PACKETS

Sender based mechanisms for recovering from packet loss can be classified as retransmission-based techniques and Forward Error Correction (FEC) techniques. For delay sensitive real-time applications, such as telephony, FEC-techniques are dominant because packet losses can be recovered without time-consuming retransmissions. FEC-techniques transmit the speech parameters in two or more consecutive packets. The FEC scheme may also use exclusive-or (XOR) to pack the redundancy information from several frames together, [4]. Two examples are shown in the figure below.

a) Single frame redundancy, 100% overhead

Orig.	P_n	P_{n+1}	P_{n+2}	P_{n+3}	P_{n+4}
Red.	P_{n-1}	P_n	P_{n+1}	P_{n+2}	P_{n+3}

b) XOR redundancy (2 frames), 100% overhead

Orig.	P_n	P_{n+1}	P_{n+2}	P_{n+3}	P_{n+4}
Red.	$P_{n-2} \oplus P_{n-1}$	$P_{n-1} \oplus P_n$	$P_n \oplus P_{n+1}$	$P_{n+1} \oplus P_{n+2}$	$P_{n+2} \oplus P_{n+3}$

Figure 2 Two examples of redundant packets. Orig. is the original payload and Red. is the redundant payload.

Redundancy transmission according to scheme a) can recover single packet losses and scheme b) can recover two consecutive packet losses. The cost of the increased robustness is increased payload size. Since AMR allows for switching to a lower rate mode, it is possible to avoid increasing the total bit rate by switching from the 12.2 kbit/s mode to, for example, the 6.7 kbit/s mode. The delay is also increased since future packets are needed to recover lost packets. In scheme a) the packet with original payload P_{n+1} is needed to recover packet P_n if it was lost. The XOR redundancy requires sequential decoding to recover from double packet losses. In scheme b), both packets with original payloads P_{n+2} and P_{n+3} are needed to recover packets P_n and P_{n+1} if they were lost.

0-7803-7549-1/02/$17.00 ©2002 IEEE.

3. SELECTIVE REDUNDANCY

Using redundant packets at all times is, of course, desirable for networks with high packet loss rates. However, for good network conditions, this would limit the speech quality, since the codec rate is reduced to avoid a large increase in bit rate. This quality loss is even more noticeable for complex signals, such as speech with background noise and music.

For low packet loss rates (<2%), packet losses during stationary speech segments are normally concealed well with a conventional Error Concealment Unit (ECU). Problems with intelligibility occur when onset frames or non-stationary frames are lost. To maintain the intelligibility, it is possible to enable redundancy only for these sensitive frames, i.e. important frames are transmitted twice while the remaining frames are only transmitted once. In cases where it is not possible to increase the bit rate by a large amount, this selective redundancy method can use the 6.7 kbit/s mode together with redundancy for sensitive frames, and the 12.2 kbit/s mode for the remaining frames. Among the methods that can be used to analyze if a frame is sensitive are:

- Using stationarity measurements within the encoder, e.g. in the Voice Activity Detector (VAD)
- Frames that are judged by the encoder to be difficult to conceal with a normal ECU, [5].

The concept of using a lower bit rate mode for onset frames is in contrast to the general principle used in variable rate speech codecs where the bit rate is increased for onsets and non-stationary segments, [6].

4. PARAMETER INTERPOLATION

Conventional ECUs extrapolate speech codec parameters from the previous frame in case of packet loss. If future speech frames are available in the receiver frame buffer, they can be exploited to interpolate the speech parameters to conceal a packet loss. The parameters in the AMR codec that may be interpolated are LSF parameters, pitch lag and gain factors.

Interpolation of LSF parameter is studied in [7]. However, we conducted a pre-study that indicated only minor improvements. For pitch and codebook gains it is found that they are less suitable for interpolation as they show an irregular behavior, independent of whether the speech signal is voiced or not.

Pitch lags on the other hand are well behaved during steady-state voiced segments. The decoding error that propagate over several frame due to the long history in the pitch predictor, may be reduced if pitch lags are interpolated between the preceding and subsequent frame instead of extrapolated from the preceding frame. This is shown in figure 3, where a packet is lost during a steady-state voiced segment. For the normal ECU, figure 3c, where the pitch lag is repeated from the preceding frame, the erroneous phase between the excitation from the pitch predictor and the codebook gives significant error propagation. With lag interpolation, the pitch lag used in the synthesis is closer to the original and the waveform error is reduced, see figure 3d.

5. PARTIAL REDUNDANCY

As figure 3d indicates pitch lag interpolation gives an improvement. The benefit could be even higher if a more correct pitch gain was available at the receiver. Unfortunately, as indicated above, pitch gain is not suitable for interpolation. The conventional ECU repeats a pitch gain based on the median of the previous pitch gains with an additional attenuation. By enabling redundancy transmission for the pitch gain parameters,

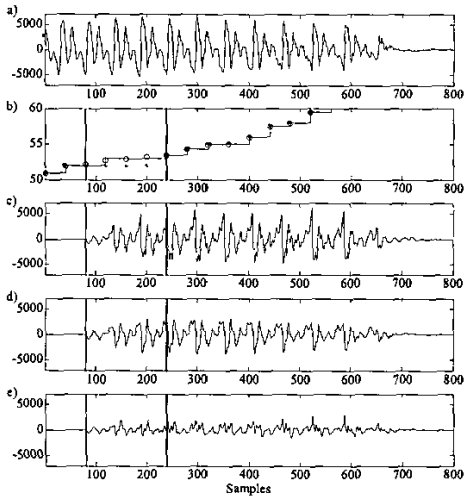

Figure 3 One frame (160 samples) is lost in the marked region, a) original signal, b) correct pitch lag (solid line), pitch lag extrapolated by the normal ECU (dots), interpolated pitch lag (circles), c) residual error using normal ECU (SNR = 3.2 dB), d) error using pitch lag interpolation (SNR = 5.9 dB), e) error using partial redundancy in combination with pitch lag interpolation (SNR = 12.7 dB).

it becomes possible to improve the synthesis considerably, as shown in figure 3e. As an example, the pitch gains in the AMR 12.2 kbit/s mode are encoded with 16 bits per frame. Partial redundancy transmission of this parameter will increase the bit rate by just 0.8 kbit/s.

The principle of partial redundancy can be generalized in such a way to transmit as redundant information a given amount of the most important coded speech bits, which have the highest potential to improve the quality of the decoded speech after error concealment. These bits may be derived in the speech encoder either ignoring or pre-empting the actions of the ECU. The amount of partial redundancy may depend on the potential gain for the decoded speech as well as on the channel quality.

6. EXPERIMENTS

A set of objective tests, using PESQ, and subjective tests, using paired comparison tests, were performed.

Description	A	B	C	D	E	F
AMR12.2	X	X	X	X		
Pitch lag interpolation		X	X	X	X	X
Partial redundancy (AMR12.2 frames only)			X	X		
Selective redundancy (AMR6.7 + redundancy for sensitive frames)				X		
AMR6.7 Single frame redundancy					X	
AMR6.7 XOR (2 frames) redundancy						X
Bit rate [kbit/s]	12.2	12.2	13.0	<13.4[1]	13.4	13.4
Future packets needed	0	1	1	1	1	3

Table 1 Codec configurations. [1]Important frames use a bit rate of 13.4 kbit/s while the remaining frames use 13.0 kbit/s.

The codec configurations used are shown in table 1, and table 2 shows the packet loss distribution for the test cases. A Gilbert

model, [8], was used to simulate the degraded network conditions. An ideal network test case (Clean) was included as a reference. The results of the objective tests are shown in table 3. To confirm the objective results, a blind paired-comparison test was performed with 8 expert listeners. The speech consisted of 10 Swedish sentences with both male and female speakers. Speech without background noise was used. Figure 4 shows the average scores and the 95% confidence interval.

Condition	Single	Double	Triple	>	Total
I	1.7	0.6	-	-	2.3
II	3.7	1.8	0.9	0.4	6.8
III	5.7	3.6	2.1	0.7	12.1
IV	8.7	8.2	5.4	3.4	25.7

Table 2 Packet loss distribution (%), '>' means that more than three consecutive packets are lost.

Configuration	Clean	I	II	III	IV
A	3.98	3.31	2.81	2.51	1.70
B	3.98	3.34	2.85	2.56	1.72
C	3.98	3.41	2.94	2.67	1.80
D	3.72	3.50	3.17	2.90	2.15
E	3.62	3.54	3.33	3.13	2.47
F	3.62	3.62	3.53	3.41	2.87

Table 3 PESQ results

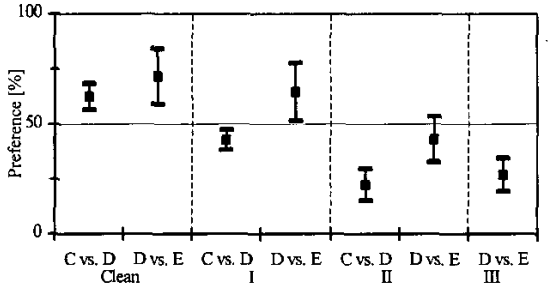

Figure 4 Result from the blind paired-comparison test

The experiments evaluate the quality improvements the various configurations give under different network conditions. Based on the results, the adaptation between configuration C, D, E and F should be done as follows:

- For clean conditions, the best choice is to use the AMR12.2 mode. Partial redundancy can be used to handle occasional packet losses, with only a small increase in bit rate.
- For low packet loss rates (~2%), the subjective test showed that selective redundancy was the best choice, although the objective results showed a small preference for single frame redundancy.
- Selective redundancy performs well up to packet loss rates of about 5%. For higher packet loss rates, the lower rate mode and single frame or XOR redundancy transmission has to be used.

Configurations A and B were omitted from the blind paired-comparison test since informal listening tests showed that configuration C in all conditions performs the same or slightly better than A or B with only a minor increase in bit-rate. Configuration F was omitted since the delay is much longer which would affect the conversational quality. In a real system, F would be used for situations when the loss rate is high since a longer delay is preferred over loosing intelligibility.

7. NETWORK ADAPTATION

A straightforward network condition adaptation scheme for the VoIP system is presented in figure 5. The horizontal bars are proposed working ranges for the different configurations. The arrow is a suggested adaptation path that shows which configuration to use depending on the packet loss rate. The packet loss rate is reported to the encoder via a feedback channel from the decoder (see fig. 1). This may be done either using dedicated receiver reports or by in-band signaling within the return speech path.

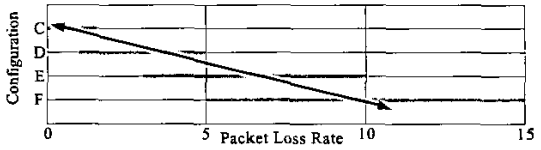

Figure 5 A system following the suggested adaptation path will maintain the network load, while still striving towards maximizing the subjective speech quality for the present network condition.

8. CONCLUSIONS

In this paper, it has been shown that the AMR codec is a suitable choice for voice over IP, and bandwidth efficient systems may be developed without modifying the speech codec specification or increasing the complexity of the codec considerably. The adaptive capabilities allows for maximizing the quality of the service for all network conditions, without increasing the bit rate significantly.

For packet loss rates below 1%, a high rate mode should be selected to maximize the basic speech coder quality. For high packet loss rates (~5%), a lower rate should be selected and redundancy transmission should be enabled. To cope with occasional packet loss bursts and low packet loss rates (1-5%), selective redundancy and partial redundancy is used together with an enhanced Error Concealment Unit.

9. REFERENCES

[1] 3GPP TS. 26.090, "Adaptive Multi-Rate (AMR) speech transcoding," 3rd Generation Partnership Project (3GPP).

[2] IETF RFC 3267, J. Sjöberg et al., "RTP payload format and file storage format for the Adaptive Multi Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) audio codecs," 2002.

[3] A. Uvliden, S. Bruhn and R. Hagen, "Adaptive multi-rate. A speech service adapted to cellular radio network quality," *Proceedings 32nd Asilomar conf. on signals, systems and computers*, vol. 1, pp. 343--347, Nov. 1998.

[4] N. Shacham and P. McKenney, "Packet recovery in high-speed networks using coding and buffer management," *Proceedings IEEE Infocom 90*, pp. 124-131, May 1990.

[5] J. C. De Martin, "Source-Driven Packet Marking for Speech Transmission over Differentiated-Services Networks," Proceedings *ICASSP 2001*, pp.753-756, May 2001.

[6] A. Das et al., "Multimode variable bit rate speech coding: an efficient paradigm for high-quality low-rate representation of speech signal," *Proceedings ICASSP 1999*, pp. 2307-2310, Mar. 1999.

[7] J. Wang and J. D. Gibson, "Parameter interpolation to enhance the frame erasure to CELP coders in packet networks," *Proceedings ICASSP 2001*, pp. 745-748, May 2001.

[8] H. Sanneck and G. Carle, "A framework model for packet loss metrics based on loss runlengths," *Proceeding. SPIE/ACM SIGMM Multimedia Computing and Networking Conference*, pp 177-187, Jan. 2000.