Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 54 (2012) 957–974

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Comparison and optimization of packet loss recovery methods based on AMR-WB for VoIP

Zhongbo Li [a,b], Shenghui Zhao [a,*], Stefan Bruhn [c], Jing Wang [a], Jingming Kuang [a]

[a] *School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[b] *Institute of the Chinese Electronic Equipment System Engineering Company, Beijing 100141, China*
[c] *Multimedia Technologies, Ericsson Research, Sweden*

Received 19 January 2011; received in revised form 21 January 2012; accepted 7 April 2012
Available online 20 April 2012

Abstract

AMR-WB codec, which has been standardized for wideband speech conversational applications, has a broad range of potential applications in the migration of wireless and wireline networks towards a single converged IP network. Forward error control (FEC) and multiple description coding (MDC) are two promising techniques to make the transmission robust against packet loss in Voice over IP (VoIP). However, how to achieve the optimal reconstructed speech quality with these methods for AMR-WB under different packet loss rate conditions is still an open problem. In this paper, we compare the performance of various FEC and MDC schemes for the AMR-WB codec both analytically and experimentally. Based on the comparison results, some advantageous configurations of FEC and MDC for the AMR-WB codec are obtained, and hence an optimization system is proposed by selecting the optimal packet loss recovery scheme in accordance with the variable network conditions. Subjective AB test results show that the optimization can lead to obvious improvements of the perceived speech quality in the IP environment.

Keywords: VoIP; AMR-WB; Packet loss; FEC; MDC

1. Introduction

IP network is evolving into a universal communication network that will accommodate all types of traffic, including voice, video, and data. An elementary and challenging component among them is the transmission of voice packets.

Speech coding is usually used in VoIP to reduce the transmission bit rate. The narrowband speech coding standards, e.g., G.711 (ITU-T, 1988), G.726 (ITU-T, 1990), G.728 (ITU-T, 1992), G.729 (ITU-T, 2007), G.723.1 (ITU-T, 2006) and AMR (Adaptive Multi-Rate) (AMR, 2001) are widely used in narrowband speech communication. Compared to narrowband speech coding (limited to about 200–3400 Hz and sampled at a rate of 8 kHz), the wideband speech coding with high-frequency extension from 3400 to 7000 Hz not only improves the intelligibility and naturalness of speech, but adds also a feeling of transparent communication. However, the formerly proposed wideband speech coding standards G.722 (ITU-T, 1988) at 48, 56, and 64 kbps and G.722.1 (ITU-T, 2005) at 24 and 32 kbps are not suitable for VoIP due to their high rates.2 kbps are not suitable for VoIP due to their high rates.

The AMR-WB (Adaptive Multi-Rate Wideband) codec has been standardized by 3GPP (3GPP, 2001) and ITU-T (ITU-T, 2003) for wideband speech conversational applications in March 2001 and in July 2002 respectively. This is of far-reaching significance because, for the first time, the same codec has been adopted for wireless as well as wireline services. This eliminates the need for transcoding and facilitates the implementation of wideband voice applications and services across a wide range of communication systems and platforms. Therefore, AMR-WB has a broad

* Corresponding author.
E-mail address: shzhao@bit.edu.cn (S. Zhao).

0167-6393/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2012.04.003

range of potential applications in the migration of wireless or wired networks toward a single converged IP network. Consequently, AMR-WB has recently been selected by TISPAN as mandatory codec for terminals supporting wideband telephony services originating and terminating end to end IP media flows in NGN (Next Generation Network) (ETSI, 2009-12).

In IP environments, when routers are overloaded, they drop packets. Therefore, some voice packets inevitably disappear in the IP network. In addition, packets typically arrive at the terminating end with a delay jitter, which is a further result of the nature of the IP network. Hence, in real-time communications, a voice packet that arrives at the receiver endpoint too late is useless and equivalent to a lost packet. In order to mitigate the impact of such packet loss or frame loss, concealment techniques can be applied. However, the effect of frame losses can only be concealed to a certain degree and any loss will not only affect the reconstruction of the current frame but also impact the following frames due to the CELP (Code-Excited Linear Prediction) structure of AMR-WB with an adaptive codebook and the strong dependence between the parameters of the adjacent frames. Thus, the quality of the reconstructed speech will inevitably be degraded under packet loss conditions. The AMR-WB standard provides a non-normative specification of an error concealment module defining a minimum performance reference (3GPP, 2001). How to improve the capability of AMR-WB to combat packet loss over this reference remains a significant and challenging issue.

Sender-based loss recovery techniques which usually cause bandwidth consumption are promising for AMR-WB (Johansson et al., 2002), and the mode adaption of AMR-WB codec can be utilized for packet-switching networks to achieve both robustness and flexibility under various network conditions.

Forward error correction (FEC) and multiple descriptions coding (MDC) are two most promising sender-based loss recovery techniques. The performance of media-independent FEC (MI-FEC) and MDC is compared analytically in (Kim and Bastiaan Kleijn, 2006). That reference concludes that the side distortion optimized MDC (SD-MDC)generally performs better than MI-FEC. In (Podolsky et al., 1998), the performance of media-dependent FEC (MD-FEC) is evaluated by simulations and it is found that the MD-FEC has a consistently positive impact on the performance, provided the use of redundancy is carefully controlled. In (Altman et al., 2002), the performance of MD-FEC is evaluated based on queuing analysis and the analytical results show that MD-FEC may provide better performance given that the total packet loss probability does not increase too fast with the amount of FEC. The performance of MDC and SDC (Single Description Coding) is compared experimentally in (Apostolopoulos et al., 2002) and the results show that MDC with path diversity from content delivery networks can provide significant performance benefits over the conventional SDC. Recently, an analytical evaluation of the potential of MD-FEC and MDC for real-time audio has been investigated in (Gyorgy et al., 2006). The authors consider an ideal condition in which the set of code rates is continuous so that the optimal amount of redundancy can be set for the MD-FEC and that the side and central distortions of the MDC can be adjusted more smoothly such that the achievable gain can be increased. Under these assumptions, they conclude that MDC always performs better than MD-FEC. The above investigations use either analytical or experimental methods to evaluate the performance of FEC and MDC. The analytical evaluations provide the performance bounds for different loss recovery methods, which rather is of theoretical significance, whereas the experimental evaluations are based on specific codecs in specific communication systems, which is of more practical significance. To our best knowledge, however, either analytical or experimental evaluation of MD-FEC and MDC for AMR-WB still remains as an unexplored field.

The main contributions of this paper are threefold.

(1) Mathematical models of R-D bounds for different FEC and MDC methods are established which take into account some limitations of the AMR-WB codec, e.g., the discrete set of bit rates (modes) and the error propagation effect. Analytical comparisons for different FEC and MDC methods for the AMR-WB codec are made based on these mathematical models. Moreover, the mathematical models, initially established for the AMR-WB codec, can also be applied to other prediction-based speech codecs by adjusting parameters of those models in correspondence with them.
(2) Extensive experimental comparisons are made to obtain some advantageous configurations of FEC and MDC for the AMR-WB codec. Then, the differences between the analytical and experimental results are discussed. It is the first time that an experimental comparison is made between the performances of FEC and MDC for a specific speech codec.
(3) Based on the comparison results, an adaptive optimization system for AMR-WB coded speech transmission is proposed that selects the optimal packet loss recovery scheme in accordance with the actual network conditions.

The remainder of the paper is organized as follows. Section 2 describes the existent FEC and MDC techniques which are suitable for AMR-WB. The R-D bounds for different FEC and MDC techniques are compared analytically in Section 3. Section 4 compares the performance of different FEC and MDC techniques experimentally and an adaptive optimization system for the AMR-WB codec is proposed in Section 5. Section 6 concludes the paper.