Available online at www.sciencedirect.com
ScienceDirect
Computer Speech & Language 54 (2018) 61–70
www.elsevier.com/locate/csl

A scalable wideband speech codec using the wavelet packet transform based on the internet low bitrate codec☆

Koji Seto, Tokunbo Ogunfunmi*

Department of Electrical Engineering, Santa Clara University, Santa Clara, CA 95053, USA

Received 4 May 2017; received in revised form 8 March 2018; accepted 4 September 2018
Available online 8 September 2018

Abstract

Most recent speech codecs employ code excited linear prediction (CELP) and transmit side information to improve speech quality under packet loss. Another approach to achieve high robustness to packet loss is to use the frame independent coding scheme based on the internet low bitrate codec (iLBC). The scalable wideband speech codec based on the iLBC was previously presented and outperformed G.729.1 at most bit rates according to the objective quality. This paper presents improvements to the previous work. Specifically, we employ the wavelet packet transform (WPT) instead of the modified discrete cosine transform (MDCT) to enhance the quality, and evaluate the proposed codec based on both the objective and subjective quality measures. The objective quality evaluation results show that clear improvement is achieved and that the proposed codec outperforms G.729.1 at the bit rate of 18 kbps or higher under clean channel conditions and has higher robustness to packet loss than G.729.1. The informal subjective test results also show similar trends.

Keywords: Wavelet packet transform (WPT); Internet low bitrate codec (iLBC); Packet loss; Scalable coding; Speech coding; Voice over Internet protocol (VoIP)

1. Introduction

Importance of scalability and wideband coding capability as feature enhancement of speech coding has been recognized in modern communication systems and applications (Ogunfunmi et al., 2015; Martin et al., 2008). Wideband codecs have been standardized (ITU-T G.729.1, 2006; ITU-T G.711.1, 2008; ITU-T G.718, 2008; Valin et al., 2012; 3GPP TS 26.445, 2015) and the availability of wideband telephony has been increasing rapidly. In the course of the transition from narrowband to wideband telephony, it is important to provide backwards compatibility with existing networks and terminals. One of the promising approaches is the use of bandwidth extension techniques, among which the bandwidth scalable speech coding provides significantly improved quality by adding enhancement bitstream layers to the core layer of a bitstream with narrowband speech data. Furthermore, the scalable bitstream structure facilitates the rate adaptation based on network conditions. In particular, the bitstream can be truncated at any

☆ This paper has been recommended for publication by Prof. R. K. Moore.

* Corresponding author.
E-mail address: togunfunmi@scu.edu (T. Ogunfunmi).

https://doi.org/10.1016/j.csl.2018.09.001
0885-2308/ 2018 Elsevier Ltd. All rights reserved.

point of communication systems to reduce the bit rate without the need of a feedback on network conditions. Therefore, a scalable bitstream provides higher flexibility and easier adaptation to sudden change of network conditions, which can be used to reduce packet loss rates. The bitstream scalability is certainly one of the essential features of speech codecs for voice over Internet protocol (VoIP) (Ogunfunmi and Narasimha, 2012) networks.

Most of the recent speech coding standards adopt code excited linear prediction (CELP) (Schroeder and Atal, 1984) as a core layer coding method. Whereas CELP relies on the pitch prediction across frame boundaries to achieve high coding efficiency, it introduces error propagation when frame loss occurs. Thus, side information is usually transmitted to mitigate the problem. Another approach to achieve high quality under packet loss is to use frame independent coding. The typical choice of the codec is the internet low bitrate codec (iLBC) (Andersen et al., 2002; Ogunfunmi and Narasimha, 2010). The iLBC does not require side information to achieve high robustness to packet loss; however, its benefit comes at the expense of relatively high operating bit rates. Whereas the CELP codec with side information has been actively researched and standardized, the iLBC-based codec has not been studied enough. This paper explores the latter approach.

The scalable wideband speech codec designed specifically for IP networks to provide high robustness to packet loss was introduced in Seto and Ogunfunmi (2012) and we presented its performance enhanced version in Seto and Ogunfunmi (2014a), which adopted the subband coding framework used in G.729.1 (Ragot et al., 2007). In this paper, we propose a novel scalable wideband speech codec which provides further improvements in quality compared to the previous work (Seto and Ogunfunmi, 2014a). In particular, the wavelet packet transform (WPT) (Coifman et al., 1991; Coifman et al., 1992; Vetterli and Kovačević, 1995; Strang and Nguyen, 1996) is employed instead of the modified discrete cosine transform (MDCT) in the enhancement layers to improve speech quality. In contrast to the MDCT, the wavelet transform (Vetterli and Kovačević, 1995; Strang and Nguyen, 1996; Rioul and Vetterli, 1991) can be used to better capture localized waveforms in time domain; therefore it is useful for encoding highly non-stationary signals such as transients (Seto and Ogunfunmi, 2014b). The proposed codec uses the WPT to encode the coding errors from lower- and higher-band codecs. Since these error signals are considered to be highly non-stationary, the WPT should be better suited than the MDCT.

The quality evaluation was performed using both objective tests and informal subjective tests. The objective evaluation results indicate that the quality of the proposed codec is improved in comparison with the previous work and is better than that of G.729.1 at bit rates of 18 kbps or higher in clean channel conditions and that it has better robustness to packet loss than G.729.1. The informal subjective evaluation results reflect similar trends.

The paper is organized as follows: Section 1 presents the structure of the proposed codec. The details of the WPT are described in Section 2. Section 3 provides the quality evaluation based on the objective tests and the subjective tests, and Section 4 concludes the paper.

2. Proposed codec

The proposed codec is a scalable wideband extension of the multi-rate iLBC (Seto and Ogunfunmi, 2013) and has basically the same structure as the codec introduced in Seto and Ogunfunmi (2014a). The multi-rate iLBC works as a narrow-band core layer codec and is characterized by frame-independent coding and multi-rate operation.

The block diagram of the encoder is illustrated in Fig. 1. The wideband input signal sampled at 16 kHz is processed using 20 ms frames. The input is decomposed into two sub-bands using a Quadrature Mirror Filter (QMF) analysis filter bank. The lower-band signal is first high-pass filtered and encoded by the multi-rate iLBC, which produces a bitstream of the core layer (Layer 1). The coding error resulting from multi-rate iLBC is perceptually weighted and transformed into the wavelet transform domain using the WPT. The higher-band signal is pre-processed by a low-pass filter after spectrally folding. The resulting signal is encoded by the time-domain bandwidth extension (TDBWE) (Ragot et al., 2007), which generates a Layer 2 bitstream. The coding error resulting from the TDBWE is decomposed into wavelet coefficients by the WPT. The resulting two sets of wavelet coefficients cover whole frequency range of wideband input signal. Those wavelet coefficients are divided into two sub-bands at either 1 kHz or 2 kHz and each sub-band is separately quantized using the scalable algebraic vector quantization (AVQ) specified in G.718 (Vaillancourt, et al., 2008; Makinen, et al., 2005), which generates Layer 3 and Layer 4 bitstreams. To further enhance quality, the quantization errors from Layers 3 and 4 are jointly quantized again by the scalable AVQ and a Layer 5 bitstream is produced.