AN EFFICIENT TRANSCODING SCHEME FOR G.729 AND G.723.1 SPEECH CODECS: INTEROPERABILITY OVER THE INTERNET

RONG-SAN LIN¹, JIA-YU WANG¹ AND JENG-SHYANG PAN²

¹Department of Computer Science and Information Engineering
Southern Taiwan University
No. 1, Nan-Tai Street, Yungkang Dist., Tainan City 710, Taiwan
rslin@mail.stut.edu.tw

²Department of Electronic Engineering
National Kaohsiung University of Applied Sciences
Chien Kung Campus 415, Chien Kung Road, Kaohsiung 807, Taiwan

Received July 2010; revised October 2011

Abstract. This paper proposes an efficient conversion algorithm for G.729 and G.723.1 speech codecs to reduce computational complexity of the communications between the G.729 and G.723.1 speech codecs. The proposed transcoding method incorporates four processes: line spectral pair (LSP) interpolation, pitch conversion, fast adaptive-codebook search, and fast fixed-codebook search. To reduce search computations, we propose a fast adaptive codebook search algorithm that uses residual signals to predict the candidate gain-vectors of the adaptive codebook. For the fixed codebook, we propose a fast search algorithm that uses an energy function to predict the candidate pulse positions. Other codec parameters are directly converted in parametric levels without executing the decoding process. Simulation results show that the proposed methods can reduce total computational complexity by 65.8%, with a shorter coding delay compared with the commonly used decode-then-encode tandem approach. Objective and subjective evaluations were used to verify that the proposed transcoding scheme provides speech quality comparable to the tandem approach.

Keywords: Speech coding, G.723.1, G.729, Transcoding, Tandem, Fast codebook search, LSP, Pitch

1. Introduction. Speech transmission is the dominant service in telecommunications networks and in the multimedia domain, specifically the emerging Voice over IP (VoIP) protocol. VoIP is based on existing data network services [1,2] and speech processes were proposed in various domains such as the perceptual speech hashing algorithm [3,4]. For integrated multimedia services, multimedia compression standards can reduce the data rate dramatically. Various speech compression standards have been recommended by the International Telecommunication Union (ITU) for different applications. Currently, the ITU-T G.723.1 [5-9] and G.729 [10-12] speech codecs are considered the best standards for very low bit rate telephony services. The G.723.1 codec is recommended for H.323 Internet phone systems and the H.324 digital videophone service in public switching telephone network (PSTN) systems [13]. The G.723.1 speech coder has been used on the Internet extensively, for example, in the built-in software NetMeeting in Microsoft windows and other VoIP communication systems. The G.729 codec with lower coding delay and available selection of multiple data rates is the most popular speech codec used in H.323 systems, providing multimedia communication over Internet protocol to achieve guaranteed quality of service for real-time voice, data and video, or any combination of the three, including video telephony [14]. The gateway server processes packet data