Chapter 7

Advanced Information Hiding for G.711 Telephone Speech

Akinori Ito
Tohoku University, Japan

Yōiti Suzuki
Tohoku University, Japan

ABSTRACT

G.711 is the most popular speech codec for Voice over IP (VoIP). This chapter proposes a method for embedding data into G.711-coded speech for conveying side information for enhancing speech quality such as bandwidth extension or packet loss concealment. The proposed method refers to a low-bit rate encoder to determine how many bits are embedded into each sample. First, a variable-bit rate data hiding method is proposed as a basic framework of the proposed method. Then, the proposed method is extended to achieve fixed bit rate data hiding. According to comparison experiments, the proposed method is proved to achieve higher speech quality compared with the conventional method. Moreover, the authors developed a low-complexity speech bandwidth extension method that uses the proposed data hiding method.

INTRODUCTION

Voice over Internet Protocol (VoIP) technology has been extensively used as a new infrastructure of the public phone network (Varshney, et al., 2002). Although several codecs are available for VoIP, G.711 (ITU, 1988), the simplest codec, is the most common one at present and is expected to remain so for the immediate future.

G.711 uses 64 kbit/s for conveying telephone-quality speech. Although its quality is enough to convey linguistic information, its quality is much lower than so-called wideband speech. Besides, as VoIP uses a connectionless communication channel such as Real-time Transport Protocol (RTP) (Schulzrinne, et al., 2003), the packet losses are not recovered by the transport protocol, therefore further degradation of speech is inevitable.

There have been several attempts to enhance G.711-coded speech, for both packet loss concealment (Komaki, et al., 2003) and bandwidth expansion (Aoki, 2006; Larsen & Aarts, 2004;
Packet loss concealment reduces the degradation of speech that is specific to IP-based communication (Perkins, et al., 1998), whereas bandwidth expansion provides users with a value-added speech communication experience. These enhancement methods require side information to the speech data coded by the G.711 codec. As the bit rate of the additional data is not very high, adding more side information to the original speech data is not a major problem in terms of bit rate. The problem with the additional side information is how to convey the side information with the speech itself. If we simply add the side information to the original speech data, we have to remodel the existing data format or communication protocol, but using one’s own data format or protocol prevents communication with most terminals that can handle only the standard protocols. Therefore, it is desirable that the speech data containing the additional data for enhancement is downward compatible with ordinary speech data coded by G.711.

An approach based on data hiding solves this problem. Data hiding is a technique to embed certain data into the original media data (or host signal) such as image, video, and speech, without significantly degrading the quality of the original media data (Petitcolas, et al., 1999).

Data hiding has been usually used for steganography and watermarking. Steganography is a method of secret communication, where the existence of a communication channel is kept secret. If data hiding is applied to steganography, the embedded information is regarded as more important than the host signal and so must be kept secret. Watermarking is a method to embed secret information into media data to keep track of the distribution of that data, and its most common purpose is copyright protection. If data hiding is applied to watermarking, the embedded information must be kept secret and degradation of the host signal should be imperceptible. Moreover, in these applications, the embedding methods must be robust against attacks.

Conversely, in the present study, data hiding is used to embed side information for enhancing the host signal and is hidden for the sake of compatibility. Thus, the embedded information need not be kept secret. In addition, we do not need to consider any attack on the hidden data in this case because the VoIP data are transmitted digitally. On the other hand, a data hiding method for compatibility should embed much data compared with a data hiding for steganography or watermarking, while keeping the quality of the host signal high.

Several data hiding methods have been developed with the aim of enhancing of multimedia data with backward compatibility. Marvel et al. (1999) proposed a spread spectrum image steganography, which is the very first work for using spread spectrum technique for data hiding. They pointed out that the steganography technique can be used to convey additional information such as in-band captioning. Yilmaz and Alatan (2003) proposed a method for using data hiding to carry side information for error concealment of video stream. Moreover, Smith et al. (2005) proposed a method to embed sensor data to RFID time series, which enables to create a new communication channel without changing a protocol. As our target is an audio signal, especially speech signal for VoIP, we are now studying other data hiding methods for VoIP. Several methods for hiding data in VoIP based on G.711 have been developed. For example, Aoki proposed a data-hiding-based speech enhancement approach for the G.711 codec (Aoki, 2007), based on LSB substitution for packet loss concealment and bandwidth expansion. Although this approach degrades the speech quality, it can be used for any protocol. Similar approach was also proposed by Vary and Geiser (2007), Tian et al. (2009), and Mazurczyk and Lubacz (2010).

In this chapter, we propose a novel data hiding method for G.711-coded speech. Our approach...