Chapter 8

Enhancement of Speech Quality in Telephony Communications by Steganography

Naofumi Aoki
Hokkaido University, Japan

ABSTRACT

Steganography can transmit supplementary data without changing conventional data formats. The concept of high value-added communications is drawn from this advantage of steganography. As a specific application of the concept, this chapter describes two topics about the enhancement of the speech quality in telephony communications by steganography. A packet loss concealment technique and a band extension technique are explained. These techniques employ steganography for transmitting side information for improving the performance of signal processing. In addition, this chapter describes an efficient steganography technique devised for G.711, the most common codec for telephony speech standardized by ITU-T. The proposed technique, named selective LSB replacement technique, outperforms the conventional one in order to decrease the degradation caused by embedding side information into speech data by steganography.

INTRODUCTION

There is a somewhat negative image of steganography since it may be employed for illegal secret communications (Singh, 2000). However, any technology can be a poison or a medicine depending on its usage. The concept of high value-added communications potentially indicates a positive usage of steganography. Transmitting supplementary data by steganography, new functions may be realized without changing conventional data formats.

Of course, it might be easier to design completely new systems for such new functions. However, it costs enormously to replace conventional standards that are used widespread. The concept of high value-added communications based on steganography can be a solution for adding new functions to the conventional standards without losing their compatibility (Aoki, 2007b).

DOI: 10.4018/978-1-4666-2217-3.ch008
This concept gives a research framework of steganography that employs supplementary data as side information for improving the performance of signal processing. One of the practical targets is controlling the quality of multimedia information without increasing its apparent data size. It mainly focuses on the enhancement of communications quality at the receiver by using secret side information embedded at the sender. It has been indicated that the concept may potentially be applicable to some practical applications such as packet loss concealment for reliable speech communications (Aoki, 2007b; Ito, Konno, Ito, & Makino, 2010), band extension for high-fidelity speech communications (Aoki, 2007b; Ito, Handa, & Suzuki, 2009), and the improvement of a speech codec (Ito & Makino, 2009), and so on. Other recent trends are controlling the quality of entertainment applications such as manipulating vocal signal in music data at the receiver by using secret side information embedded at the sender (Sasaki, Hahm, & Ito, 2011).

In order to appeal the effectiveness of the concept, this article describes two topics about the enhancement of the speech quality in telephony communications by steganography. A packet loss concealment technique and a band extension technique are explained (Aoki, 2003a, 2003b). This article describes how the proposed technique outperforms the conventional one.

A packet loss concealment technique is employed for reliable telephony communications. Packet loss causes silent gaps that degrade the speech quality in telephony communications. It inevitably occurs in recent telephony communications systems based on VoIP (Voice over IP) that transmits speech data in best-effort networks such as the Internet. In order to enhance the speech quality, this article proposes a technique in which side information is embedded into speech data by steganography, and employed for restoring the gap frames caused by packet loss (Aoki, 2003a, 2003b). This article describes how the proposed technique outperforms the conventional one.

A band extension technique is employed for high-fidelity telephony communications. It enhances the clarity of the telephony speech. Since the sampling rate of the conventional telephony speech is 8 kHz, the frequency components higher than 4 kHz are not transmitted. Due to this limitation, the speech quality is somewhat muffled. In order to enhance the speech quality, this article proposes a technique in which side information is embedded into speech data by steganography, and employed for restoring the frequency components higher than 4 kHz (Aoki, 2006b, 2006c, 2007a). This article describes how the proposed technique outperforms the conventional one.

**STEGANOGRAPHY FOR TELEPHONY SPEECH**

As shown in Figure 1, steganography embeds secret message into the redundancy of cover data. Such redundancy may be a container for transmitting the secret message. The capacity for the secret message depends on the size of the redundancy.

In this article, speech data encoded with G.711 is chosen to be the cover data. G.711 is the most common codec for telephony speech standardized by ITU-T (International Telecommunication Union Telecommunication Standardization Sector) (ITU-T, 1988). It consists of μ-law and A-law schemes designated PCMU and PCMA, respectively. PCMU is mainly employed in North America and Japan. It encodes 14 bit speech data into 8 bit compression data at an 8 kHz sampling rate. PCMA is mainly employed in Europe. It encodes 13 bit speech data into 8 bit compression data at an 8 kHz sampling rate. Due to the simple logarithmic quantization algorithms, there remains a plenty of redundancy in the speech data encoded with G.711.

The LSB (Least Significant Bit) replacement technique is known as one of the simplest steganography technique (Cox, 2008). It just embeds the secret message into the LSB of the cover data. It is based on the fact that the LSB of the cover data obtained from multimedia information such