Perceptual Model-Based Information Hiding in Audio Signals

Othman O. Khalifa, Omar Mahmoud El Mahdi A. Al Khazmi and Sheroz Khan

Electrical and Computer Engineering
International Islamic University Malaysia
khalifa@iium.edu.my

ABSTRACT

Audio data hiding is the process of embedding information into an audio signal so that the embedded information is inseparable from it and imperceptible to the listener. Information hiding is a multi- disciplinary area that combines signal processing with cryptography, communication theory, coding theory, information theory and the theory of human auditory and visual systems where information is hided within a host signal. A data hiding system should be robust, meaning that the embedded data could be decoded from the combined signal, even if it is distorted or attacked. This paper examines information hiding in speech signals. A perceptual model-based information hiding in speech signal is developed.

Keywords: Data hiding, Steganography, Watermarking, Security

1. Introduction

Digital media production and distribution has witnessed dramatic growth in recent years. Data hiding techniques have developed a strong basis for steganography area with a growing number of applications like digital rights management. The availability of the Internet, low-cost and reliable storage devices, and high speed networks has made it possible to replicate and distribute digital information easily. This has created a need for protection and enforcement of intellectual property rights for digital media to prevent its illegal copying/reproduction. The need has spurred research in data hiding in digital media due to its applications in watermarking, annotation, and steganography [2]. Most of the work in data hiding has been concentrated on hiding a small amount of information such as copyright data or a watermark in images and video segments [2-5]. However, general requirements, challenges and principles of hiding data in an audio are the same as those for embedding information in a video. Robustness of the hidden data, for example, is a key requirement for successful embedding and retrieval of the data. In other words, standard signal processing operations, such as noise removal and signal enhancement, must not result in loss or degradation of the embedded information. Additionally, for covert communication, the embedded information must withstand channel noise and intentional attacks or jamming on the signal. Also important in covert communication is the resilience of the hidden information to stay hidden to pirates during their intentional or unintentional attempts at detection. A measure of effectiveness of data embedding is the probability of detection of hidden data. Clearly the more robust the host medium – image, video, or audio – to attacks and common operations, the higher would be its effectiveness.

1. Perceptual Modeling

Auditory perception is based on the critical band analysis in the inner ear where a frequency-to-location transformation takes place along the basilar membrane. The auditory system is usually modeled as a bandpass filterbank, consisting of strongly overlapping bandpass filters with bandwidths around 100 Hz for bands with a central frequency below 500 Hz and up to 5000 Hz for bands placed at high frequencies. If the highest frequency is limited to 24000 Hz, 26 critical bands have to be taken into account. An important property of a watermarking system is to make sure that the presence of embedded information is transparent to the end user (certain applications may also require an audible/visible watermark). To perceptually hide information one would have to study the perceptual properties of the host signal. Audio watermarking poses greater challenges than image/video watermarking, mainly because of the wider dynamic range and higher sensitivity to AWGN of the HAS, compared to the Human Visual System (HVS). But in contrast to its large dynamic range, its differential range is limited. This means that louder sounds can mask out weaker ones [10]. The HAS also has a number of other properties which can be exploited to perceptually hide the distortions caused by embedding informationTwo properties of the HAS dominantly used in watermarking algorithms are frequency masking and temporal masking.

2.1 Frequency masking
It is a frequency domain phenomenon where a low level signal, e.g. a pure tone (the maskee), can be made inaudible (masked) by a simultaneously appearing stronger signal (the masker), e.g. a narrow band noise, if the masker and maskee are close enough to each other in frequency.

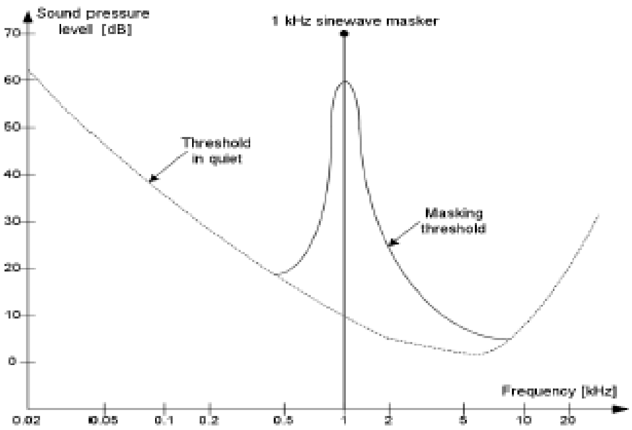

Fig. 1. Frequency masking in the human auditory system (HAS)

2.2 Temporal masking
In addition to frequency masking, two phenomena of the HAS in the time domain also play an important role in human auditory perception. Those are pre-masking and postmasking in time [34]. The temporal masking effects appear before and after a masking signal has been switched on and off, respectively (Figure 2.3). The duration of the premasking is significantly less than one-tenth that of the post-masking, which is in the interval of 50 to 200 milliseconds. Both pre- and post-masking have been exploited in the MPEG audio compression algorithm and several audio watermarking methods.

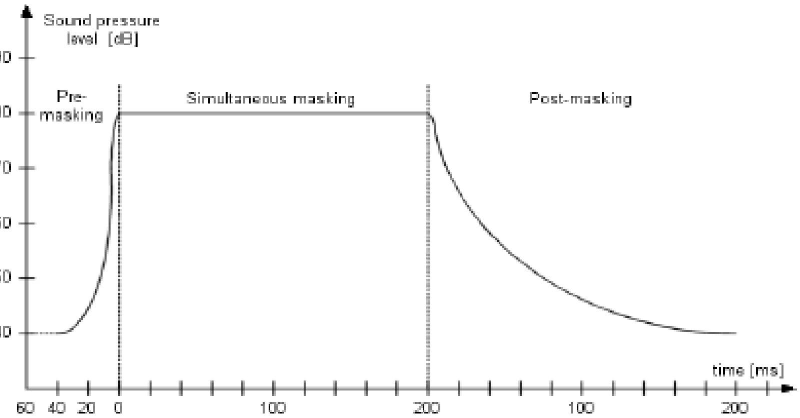

Fig. 2. Temporal masking in the human auditory system (HAS).

Psychoacoustical Masking Thresholds of Hearing
Auditory masking is a perceptual property of the human auditory system in which the presence of a strong tone renders the hearing of a weaker tone in its temporal or spectral neighborhood imperceptible. Also, a pure tone is masked by a wide-band noise if the tone occurs within a critical band. The ability of an audio signal to make other audio signals in its neighborhood (both in time and frequency) imperceptible to the human ear is called masking. A loud sound distorts the absolute threshold of hearing and thus renders

inaudible a weaker sound that may otherwise be audible [10]. We can observe in Figure 2.1, which depicts the masking effect, that the masker hides/masks any signal falling within its masking threshold (like the masked sound). The threshold in quiet is the minimum sound pressure level of a signal that can be heard by the HAS in the absence of other sounds. Frequency masking is based on the observation that the human ear cannot perceive frequencies at lower energies when these frequencies are present in the vicinity of tone- or noise-like frequencies at higher energies.

As with the design of coders, the masking phenomenon can be used to embed data in an audio with negligible perceptual difference between the original, unembedded audio and the data-embedded audio. Figure 3 General procedure for embedding data in perceptually masked locations. The first step in exploiting the masking property for coding or data embedding is to determine the masking threshold level. For an utterance of speech the masker frequencies – tonal and noise-like – and their power levels are computed from frame to frame. A global threshold of hearing based on the maskers is determined for each frame. Also, the sound pressure level for quiet, below which a signal is inaudible, is obtained.

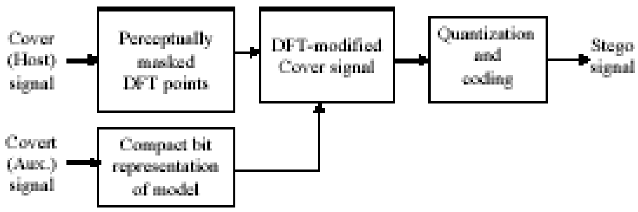

Fig. 3 General procedure for embedding data in perceptually masked locations

Figure 2 shows the normalized power spectral density (PSD), absolute quiet threshold, and threshold of hearing for a frame of utterance. The lowest spectral component around 2800 Hz in this figure, for instance, indicates that this component, being below the

3. Data Hiding Procedures

Perception-based data hiding schemes for audio are influenced by properties of the human auditory system (HAS). Many researches were conducted to determine the audibility of embedded data at masked frequencies. In the past, different perception-based algorithms were proposed for data hiding/watermarking in audio data [4 -10]. These algorithms can be broadly classified according to the underlying technique of data embedding: perceptual masking [3, 4, 5, 9], direct sequence spread spectrum (DSSS) [3,4], and phase coding [3,7,8]. Algorithms based on phase coding [3, 7, 8] work well as far as imperceptibility of the embedded data is concerned, but suffer from some limitations e.g. some of them do not perform well against standard data manipulations and well against standard data manipulations and most of them carry small payloads i.e. the amount of information embedded. For example, the phase coding technique proposed in [3] can embed only 16-32 bits of data in one-second duration audio samples. The algorithm based on echo-based coding [6] can embed about 40-50 bits of data in one-second duration of an audio signal. However, data hiding in the LSBs of audio samples in time domain is one of the simplest watermarking algorithms with very high data rates of hidden information. However, the adjusting of the LSBs of audio samples introduces noise that becomes audible as the number of the LSBs used for data hiding increases.

CONCLUSION

Audio information hiding is the process of embedding information into an audio signal so that the embedded information is inseparable from it and imperceptible to the listener. A brief survey of information hiding in audio signals based on the Perceptual properties of Human Visual System was given. Hiding of data by modifying the power spectral density and the phase of speech frames at perceptually masked frequency points has been reported. Auditory masking were presented and Data Hiding Procedures were explained

References

[1]. Poulami Dutta, Debnath Bhattacharyya and Tai-hoon Kim, Data Hiding in Audio Signal: A Review, International Journal of Database Theory and Application Vol. 2, No. 2, June 2009.
[2]. Debnath Bhattacharyya, Poulami Dutta, Maricel O. Balitana, Tai-hoon Kim and Purnendu Das, Hiding Data in Audio Signal, ACN 2010, CCIS 77, pp. 23–29, 2010, Springer-Verlag Berlin Heidelberg 2010
[3]. Ted Painter and Spanias Andreas. Perceptual coding of digital audio. Proceedings of the IEEE, 88(4):451–513, Abril 2000.
[4]. Kriti Saroha, Pradeep Kumar Singh A Variant of LSB Steganography for Hiding Images in Audio, International Journal of Computer Applications (0975 – 8887) Volume 11– No.6, December 2010
[5]. Rashid Ansari, Hafiz Malik, Ashfaq Khokhar, Data-Hiding In Audio Using Frequency-Selective Phase Alteration, ICASSP 2004, pp. 389-392.
[6]. Ingemar Cox. Digital Watermarking and Steganography. Morgan Kaufmann Publishers, San Francisco, 2008.
[7]. Gopalan, K.G.; Benincasa, D.S.; Wenndt, S.J.;, Data Embedding In Audio Signals, Aerospace Conference, 2001, IEEE Proceedings, pp. 2713 - 2720
[8]. M. F. Mansour, and A. H. Tewfik, “Time-scale invariant audio data embedding,” Proc. IEEE International Conference on Multimedia and Expo, ICME, Japan, August 2001.
[9]. W. Bender, D. Gruhl, N. Morimoto and A.Lu, “Techniques for data hiding,” IBM Systems Journal, Vol. 35, Nos. 3 & 4, pp. 313-336, 1996.
[10]. M.D. Swanson, M. Kobayashi, and A.H. Tewfik, “Multimedia data-embedding and watermarking technologies,” Proc. IEEE, Vol. 86, pp. 1064-1087, June 1998.
[11]. J. Mielikainen, “LSB Matching Revisited,” IEEE Signal Processing Letters, Vol.13, No.5, pp.285-287, May 2006.
[12]. W. C. Kuo, L. C. Wuu, C. N. Shyi, and S. H. Kuo, “A Data Hiding Scheme with High Embedding Capacity Based on General Improving Exploiting Modification Direction method” HIS2009, Aug. 2009.
[13]. K. Gopalan, et al, “Covert Speech Communication Via Cover Speech By Tone Insertion,” U.S. Patent applied for, Oct. 2003.
[14]. R.J. Anderson and F.A.P. Petitcolas, “On the limits of steganography,” IEEE J. Selected Areas in Communications, Vol. 16, No. 4, pp.474-481, May 1998.