

PERCEPTUAL SPEECH HASHING AND PERFORMANCE EVALUATION

YUHUA JIAO¹, LIPING JI² AND XIAMU NIU¹

¹Information Countermeasure Technique Research Institute
Harbin Institute of Technology

2 Yikuang Street, Nan Gang District, Harbin, 150080, P. R. China
{ yuhua.jiao; xiamu.niu }@ict.hit.edu.cn

²Information Security Research Center
Harbin Institute of Technology Graduate School at Shenzhen
University Town of Shenzhen, 518055, P. R. China

Received March 2009; revised July 2009

ABSTRACT. *A speech perceptual hash algorithm in compression domain is proposed in this paper. Speech coded at very low bitrate requires hash algorithm with high compactness and robustness. LSF, which is the intermediate result of partial decoding, could model the changing shape of the speaker's vocal tract. They are used to generate hash value. Our proposed method satisfies the robustness and discrimination power requirement of perceptual hash with very low hash bitrate. It is also a computational efficient algorithm which could be applied to scenarios with power restriction and real-time communication requirement. For performance evaluation, the perceptual hashing is modeled as a Markov information source, and the entropy rate of the Markov source is proposed as a joint quantitative measure of discrimination power and compactness. The proposed model is universal with respect to theory and practice. We conclude that randomness of hashing is a measure of discrimination power and compactness rather than security. The security relies on the size of the secret key space.*

Keywords: Perceptual hashing, Markov source, Entropy rate, MELP speech, Randomness

1. Introduction. Perceptual hashing, also refers as robust hashing and fingerprinting, maps digital multimedia data into a compact digital digest which is unique to multimedia content. Unlike cryptographic hashing, perceptual hashing is sensitive to perceptual changes of content and tolerant to content preserving operations. Multimedia of different contents would be mapped to distinct and statistical random hash values.

Perceptual hashing for audio has been widely studied recently [1]. However, audio hashing algorithms are not suitable for speech. Firstly, audio hashing is not robust to low bitrate speech coding, because parametric speech codec is totally different from audio compression scheme. Secondly, the bitrate of audio hashing is too high to speech. The hash bitrate is about 2.7Kbps in [2], which is larger than the bitrate of MELP speech coder (2.4Kbps). There are few speech hashing algorithms in literature. In [8], a speech content authentication scheme integrated with CELP speech coders was proposed. However, there are many speech coding standards for different speech communication systems [9]. However, the proposed method was limited to CELP speech coders. MELP coder provides good speech quality and error resilience at very low bitrate. It is widely applied in IP phone, wireless communication, especially in secure communication [4].

In this paper, a perceptual speech hashing algorithm for MELP coded speech (PSHM) is proposed. The perceptual hash value is calculated from partially decoded linear spectral