



TITLE:

# Shouted speech detection using hidden markov model with rahmonic and mel-frequency cepstrum coefficients

AUTHOR(S):

Fukumori, Takahiro; Nakayama, Masato; Nishiura, Takanoobu; Nanjo, Hiroaki

---

CITATION:

Fukumori, Takahiro ...[et al]. Shouted speech detection using hidden markov model with rahmonic and mel-frequency cepstrum coefficients. The Journal of the Acoustical Society of America 2016, 140(4): 3057-3057: 2aSPb7.

ISSUE DATE:

2016-10

URL:

<http://hdl.handle.net/2433/229399>

RIGHT:

Copyright 2016 Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America. The following article appeared in 'The Journal of the Acoustical Society of America 140, 3057 (2016)' and may be found at <https://doi.org/10.1121/1.4969503>.; There are hidden parts depending on the permission condition of the publisher in this pdf.

GMM has difficulty in representing temporal changes of the speech features. In this study, we further expand the previous method using hidden Markov model (HMM) which has state transition to represent the temporal changes. Through objective experiments, the proposed method using HMM could achieve higher detection performance of the shouted speech than the conventional method using GMM.

2a TUE. AM

**2aSPb7. Shouted speech detection using hidden markov model with rahmonic and mel-frequency cepstrum coefficients.** Takahiro Fukumori, Masato Nakayama, Takanobu Nishiura (Ritsumeikan Univ., 1-1-1, Nojihigashi, Kusatsu 525-8577, Japan, [fukumori@fc.ritsumeikai.ac.jp](mailto:fukumori@fc.ritsumeikai.ac.jp)), and Hiroaki Nanjo (Kyoto Univ., Kyoto, Japan)

In recent years, crime prevention systems have been developed to detect various hazardous situations. In general, the systems utilize the image information recorded by a camera to monitor the situations. It is however difficult to detect them in the blind area. To address the problem, it is required to utilize not only image information but also acoustic information occurred in such situations. Our previous study showed that two acoustic features including rahmonic and mel-frequency cepstrum coefficients (MFCCs) are effective for detecting the shouted speech. Rahmonic shows a subharmonic of fundamental frequency in the cepstrum domain, and MFCCs represent coefficients that collectively make up mel-frequency cepstrum. In this method, a shouted speech model is constructed from these features by using a gaussian mixture model (GMM). However, the previous method with