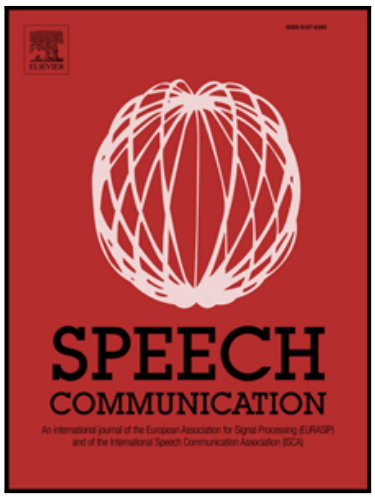

Dimensional paralinguistic information control based on multiple-regression HSMM for spontaneous dialogue speech synthesis with robust parameter estimation

Tomohiro Nagata, Hiroki Mori, Takashi Nose

PII: S0167-6393(16)30239-4
DOI: 10.1016/j.specom.2017.01.002
Reference: SPECOM 2428

To appear in: *Speech Communication*

Received date: 2 September 2016
Revised date: 27 December 2016
Accepted date: 4 January 2017

Please cite this article as: Tomohiro Nagata, Hiroki Mori, Takashi Nose, Dimensional paralinguistic information control based on multiple-regression HSMM for spontaneous dialogue speech synthesis with robust parameter estimation, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.01.002

Dimensional paralinguistic information control based on multiple-regression HSMM for spontaneous dialogue speech synthesis with robust parameter estimation

Tomohiro Nagata[a], Hiroki Mori[a], Takashi Nose[b]

[a] *Graduate School of Engineering, Utsunomiya University, 7-1-2, Yoto, Utsunomiya-shi, 321-8585 Japan*
[b] *Graduate School of Engineering, Tohoku University, 6-6-05, Aramaki Aza Aoba, Aoba-ku, Sendai-shi, 980-8579 Japan*

Abstract

This paper describes spontaneous dialogue speech synthesis based on the multiple regression hidden semi-Markov model (MRHSMM), which enables users to specify paralinguistic information of synthesized speech with a dimensional representation. Paralinguistic aspects of synthesized speech are controlled by multiple regression models whose explanatory variables are abstract dimensions such as pleasant-unpleasant and aroused-sleepy. However, in the training phase of the MRHSMM, estimated regression coefficients may have unreasonably large values, which cause fragility in the parameter generation with respect to paralinguistic information given to the synthesizer. For robust estimation of the regression matrices of the MRHSMM with unbalanced spontaneous dialogue speech samples, the re-estimation formulae were derived in the framework of the maximum a posteriori (MAP) estimation. By examining the synthesized speech, it was confirmed that the acoustic fea-

Email address: ken1@speech-lab.org (Tomohiro Nagata)