

STATISTICAL CORRECTION OF TRANSCRIBED MELODY NOTES BASED ON PROBABILISTIC INTEGRATION OF A MUSIC LANGUAGE MODEL AND A TRANSCRIPTION ERROR MODEL

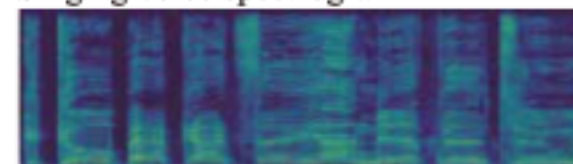
Yuki Hiramatsu Go Shibata Ryo Nishikimi Eita Nakamura Kazuyoshi Yoshii

Graduate School of Informatics, Kyoto University, Japan

ABSTRACT

This paper describes a statistical post-processing method for automatic singing transcription that corrects pitch and rhythm errors included in a transcribed note sequence. Although the performance of frame-level pitch estimation has been improved drastically by deep learning techniques, note-level transcription of singing voice is still an open problem. Inspired by the standard framework of statistical machine translation, we formulate a hierarchical generative model of a transcribed note sequence that consists of a music language model describing the pitch and onset transitions of a true note sequence and a transcription error model describing the addition of false notes.

Singing voice spectrogram



Automatic singing transcription
(AST)

Transcribed note sequence



Key estimation

C

Error correction
(Markov language model + Transcription error model)

Candidate note sequences



Candidate selection
(LSTM language model)

-10

-5