論文の内容の要旨

論文題

Harmonic acoustical models and polyphonic probabilistic musicological models applied to multiple pitch transcription of musical signals
（調波的音響モデルと音楽論的統計多重音モデルによる
音楽音響信号の多重ピッチ分析）

氏名　　ラチンスキ　スタニスワフ アンジェイ

（本文）Multiple pitch transcription, or automatic music transcription, is a process, in which a machine analyzes an input audio signal containing a recording of a musical performance and automatically infers the musical score that was being played. Effectively, this means estimating onset times, durations and pitches of all the notes present in the recording. This problem is one of the earliest, most difficult and most important tasks of the field of Music Information Retrieval (MIR), which is a dynamically developing and interdisciplinary field of science that aims at automatically retrieving information from musical data, both acoustic and symbolic. Multiple pitch transcription is a relatively new problem: due to the inherently complex nature of musical audio signals, it was not given much attention before the advent of computational methods in the 1990s. Since then it has been rapidly growing. Nevertheless, despite all the efforts, it remains largely unsolved.

The goal of this work is to design, develop and implement a top-down automatic transcription system that would approach the problem in a similar way the human musicians do: by applying prior knowledge in musicology to the process. This is in contract to the current transcription methods, which can be characterized as bottom-up, and which rarely make use of musicological knowledge. Such top-down, probabilistic approach has been already very successful in the field of speech recognition.

First, a high-performance acoustic model called Harmonic Nonnegative Matrix

Approximation (HNNMA) is proposed. This model is used to pre-process the input signal and extract note candidates from the input signals, which are then utilized in the higher levels of the system. Then, one of the first musicological models in the field appliable to the task of automatic music transcription is considered, and a usable implementation of such a model is proposed. This model is based on Dynamic Bayesian Networks (DBNs), a generalization of the commonly used Hidden Markoc Model (HMM). Finally, the proposed models are integrated in a probabilistic manner through a salience model into a working multiple pitch transcription system. Inference of the latent multiple pitches is done by means of a modified frontier algorithm, which is a generalization of the Viterbi decoding of HMMs.

Experimental evaluation of the acoustic model alone was done and it showed improvement over other state-of-the-art methods. Theoretical analysis of the of the predictive power of the musicological model by means of conditional and marginal cross-entropies was performed. Obtained results suggest that this model is capable of capturing basic characteristics of musical note sequences by encoding rules of music theory in a probabilistic framework. Finally, the integrated transcription system is tested on audio recordings. The usage of the musicological model shows significant improvement over a system consisting only of the acoustic and salience models.