Real Time Tracking and Visualisation of Musical Expression

Simon Dixon, Werner Goebl, and Gerhard Widmer

Austrian Research Institute for Artificial Intelligence, Schottengasse 3, A-1010 Vienna, Austria.
{simon,wernerg,gerhard}@ai.univie.ac.at

Abstract. Skilled musicians are able to shape a given piece of music (by continuously modulating aspects like tempo, loudness, etc.) to communicate high level information such as musical structure and emotion. This activity is commonly referred to as expressive music performance. The present paper presents another step towards the automatic high-level analysis of this elusive phenomenon with AI methods. A system is presented that is able to measure tempo and dynamics of a musical performance and to track their development over time. The system accepts raw audio input, tracks tempo and dynamics changes in real time, and displays the development of these expressive parameters in an intuitive and aesthetically appealing graphical format which provides insight into the expressive patterns applied by skilled artists. The paper describes the tempo tracking algorithm (based on a new clustering method) in detail, and then presents an application of the system to the analysis of performances by different pianists.

1 Introduction

An expert musical performer is able to shape a given piece of music to communicate high level information such as musical structure and emotion. That is, the artist goes beyond what is prescribed in the written score and modifies, gradually or abruptly, the tempo or loudness or other parameters at certain places in the piece in order to achieve certain musical and emotional effects. This activity is commonly referred to as *expressive music performance*. Expressive performance is an element of central importance in art music and especially in classical music, where the performing artists have (or take) a lot of freedom in expressing their interpretation of the music and their individuality. At the same time, expressive performance is still a poorly understood phenomenon, both from a musical and a cognitive perspective. No formal models exist that would explain, or at least quantify and characterise, aspects of commonalities and differences in performance style.

This paper presents a step towards the automatic high-level analysis of this elusive phenomenon with Artificial Intelligence methods. We restrict our attention to two of the most important expressive dimensions: fluctuations in *tempo* and *loudness* (*dynamics*). A system is presented that is able to measure tempo and dynamics of a musical performance and to track their development over time. The system accepts raw audio input (e.g., from a microphone), tracks tempo and dynamics changes in real time, and displays the development of these expressive parameters in an intuitive and aesthetically appealing

C. Anagnostopoulou et al. (Eds.): ICMAI 2002, LNAI 2445, pp. 58–68, 2002.

graphical format which provides insight into the expressive patterns applied by skilled artists.

Measuring and tracking *dynamics* is rather straightforward. The (perceived) loudness of the music can be derived from the audio signal by applying well-known signal processing techniques and psychoacoustic principles. The difficult part is inferring the basic *tempo*, and tracking changes in tempo in real time. The main problems are detecting the onsets of notes in the raw audio signal (event detection), inferring the basic rate of beats or tempo and the most plausible metrical level (tempo induction), and the real time adaptation of the tempo hypotheses in response to newly incoming information (tempo tracking).

The main technical contribution of this paper is a real time algorithm that finds the tempo of a musical performance, keeping track of multiple hypotheses, rating and updating each of the hypotheses dynamically, and allowing the user to interactively switch between hypotheses (e.g., when the system has obviously chosen a wrong metrical level). At the heart of the tempo induction and tracking system is a fast on-line clustering algorithm for time intervals.

In the following, we describe the tempo tracking and clustering algorithm in detail, and then present an application of the algorithms in a real time visualisation system for expressive music performance. An example visualisation of two pianists playing the same piece demonstrates what kind of direct insight into expressive performance can be facilitated by this kind of system.

2 Real Time Tempo Tracking

Most music has as its rhythmic basis a series of pulses, spaced approximately equally in time, from which the timing of all musical events can be measured. This phenomenon is called the *beat*, and the individual pulses are also called beats. The rate at which beats occur defines the *tempo*, a value which varies over time. Sometimes a multiple or divisor of the tempo is perceived as an alternative tempo; these different rates are called *metrical levels*.

The task of a tempo induction and tracking system at any moment during its operation is to infer, from the observed inter-note time intervals, possible beat rates and select the one that most likely represents the perceived tempo of the piece. This is performed by a clustering algorithm which groups similar time intervals between note onsets, forming clusters which correspond to musical time units, such as half notes, quarter notes and dotted quarter notes. In a mechanical performance, these time intervals would be precisely integer or simple integer fraction multiples of the time between two consecutive beats. But in expressive performance, the categories are blurred and change over time, so the clustering algorithm must be robust to noise and able to adapt dynamically to drift in the cluster centres.

The architecture of the tempo tracker is shown in figure 1. The input signal is preprocessed in several stages to detect the onsets of musical notes (events), and this information is used by the multiple tempo tracking subsystem to create a set of tempo hypotheses which are updated dynamically as further input data arrives. The highest-ranking tempo hypothesis is given as output, but the ranking can be overridden by the user selecting