Chapter 12
Music Onset Detection

Ruohua Zhou
Queen Mary University of London, UK

Josh D. Reiss
Queen Mary University of London, UK

ABSTRACT

Music onset detection plays an essential role in music signal processing and has a wide range of applications. This chapter provides a step-by-step introduction to the design of music onset detection algorithms. The general scheme and commonly-used time-frequency analysis for onset detection are introduced. Many methods are reviewed, and some typical energy-based, phase-based, pitch-based and supervised learning methods are described in detail. The commonly used performance measures, onset annotation software, public database and evaluation methods are introduced. The performance difference between energy-based and pitch-based method is discussed. The future research directions for music onset detection are also described.

INTRODUCTION

The audio signal is often considered to be a succession of the discrete acoustic events. The term music onset detection refers to detection of the instant when a discrete event begins in a music signal. Music onset detection plays an essential role in music signal processing and has a wide range of applications such as automatic music transcription, beat-tracking, tempo identification and music information retrieval.

Different sound sources (instruments) have different types of onsets that are often classified as “soft” or “hard”. The human perception of the onset is usually related to the salient change in the sound’s pitch, energy or timbre. Hard onsets are characterized by sudden increases in energy, whereas soft onsets show more gradual change. Hard onsets can be well detected by energy-based approaches, but the detection of soft onsets remains a challenging problem. Let us suppose that a note consists of a transient, followed by a steady-state part, and the onset of the note is at the beginning of the transient. For hard onsets, energy changes

DOI: 10.4018/978-1-61520-919-4.ch012
are usually significantly larger in the transients than in the steady-state parts. Conversely, when considering the case of soft onsets, energy changes in the transients and the steady-state parts are comparable, and they do not constitute reliable cues for onset detection any more. Consequently, energy-based approaches fail to correctly detect soft onsets. Stable pitch cues enable to segment a note into a transient and a steady-state part, because the pitch of the steady-state part often remains stable. This fact can be used to develop appropriate pitch-based methods that yield better performances, for the detection of soft onsets, than energy-based methods. However, although many approaches use energy information, only a few pitch-based methods have been proposed in the literature.

We discuss general scheme for onset detection and how to develop an onset detection algorithm step by step. Many existing methods are described, and a few typical methods are to be described in detail. Performance evaluation and future research directions will also be discussed. The organization of this chapter is listed as follows. The ALGORITHMS section explains the general scheme for music onset detection and typical algorithms for energy-based, phase-based, pitch-based and supervised learning approaches. In the section on PERFORMANCE EVALUATION, the evaluation of onset detection approaches is discussed, and several established evaluation results are presented. Finally, the section on FURTHER RESEARCH DIRECTIONS discusses possible research directions, inspired by human perception, which could be applied to the field of music onset detection.

**ALGORITHMS**

**General Scheme**

Many different onset detection systems have been described in the literature. As shown in Figure 1, they typically consist of three stages; time-frequency processing, detection function generation, and peak-picking (Bello et al., 2005). At first, a music signal is transformed into different frequency bands by using a filter-bank or a spectrogram. For example, the Short Time Fourier Transform (STFT) and the Resonator Time Frequency Image (RTFI) are two useful time-frequency analysis tools for onset detection. Then, the output of the first stage is further processed to generate a detection function at a lower sampling rate. Finally, a peak-picking operation is used to find onset times within the detection function, which is often derived by inspecting the changes in energy, phase, or pitch.

**Time-Frequency Processing**

Music signals are time-varying, and most of the analysis tasks require a joint time-frequency analysis. One commonly-used time-frequency analysis tool is Short Time Fourier Transform (STFT). The Fourier Transform and its inverse can transform signals between the time and frequency domains. It can make it possible to view the signal characteristics either in time or frequency domain, but not to combine both domains. In order to obtain a joint time-frequency analysis for non-stationary signals, the STFT cuts the time signal into different frames and