Tempo extraction and beat tracking with tempex and beatex

Steffen Pauws
Philips Research
steffen.pauws@philips.com

ABSTRACT

Tempex and beatex refer to a single method to perform both the extraction of perceived tempo and the localization of beat positions of musical audio. The method uses a set of driving signals computed from the musical audio to drive resonator filter banks. The output is taken as the sum across frequency bands of like-frequency resonators.

1. INTRODUCTION

In the way musical tempo is perceived, it is commonly defined as the most prominent, isochronous pulse in music, to which a listener can tap or dance. It is expressed in number of beats per minute. Listeners can disagree on what can be identified as the most prominent pulse. However, distinct perceptions are mostly related to each other by an integer multiple. In other words, a person can tap along twice or thrice as many times than another person does. These subjective differences in ordinary human interpretation are a problem in a uniform and objective evaluation of automatic tempo detection methods for music.

The perception of tempo in music is based on hearing mutually time-correlated rhythmic accents. In general, accents in music coincide with musical onsets caused by, for instance, a percussive sound, the striking of a single note or chord, or a vocal tone. Before a tempo can be detected from music, it is necessary to have knowledge on the time positions and the relative strengths of these onsets in music.

A musical onset is perceived by an audible increase in loudness, an audible change in pitch, or an audible change in timbre. In the physical domain, these perceptual changes correlate with an increase of the power in the sound signal or by a change in the frequency spectrum of the sound over time. To this end, an automatic method for finding the musically relevant onsets is equipped to detect these physical phenomena. Many different methods are already realized and published in the literature. Knowledge on how the human auditory system distinguishes neighboring frequencies and similar loudness levels is indispensable in these methods.

A known robust method consists of measuring the loudness variations in a running sound signal by detecting large positive differences of a current loudness level with a recent minimum loudness level. As a result, an array of likely onsets with varying strengths is computed. This array is then used as a driving signal for a rhythm analysis method to find a musical tempo, the meter or even an interpretation of the rhythmic interpretation of a performer (for instance, to study the 'swing' in jazz). The rhythm analysis methods are based on correlation, a bank of comb filter or computer models of oscillators. The output of these analyses can consist of the period of the dominant tempo, the phase of the tempo, the beat strength, similar information on a less dominant tempo, meter or time signature and position of the down beat in music.

2. DRIVING SIGNAL

3. RESONATOR FILTER BANK