in Conference Abstracts and Applications of SIGGRAPH '03 (San Diego, July 27-31, 2003), ACM Press.

Cati Dance: self-edited, self-synchronized music video

Tristan Jehan
MIT Media Lab
tristan@media.mit.edu

Michael Lew
Media Lab Europe
lew@mle.ie

Cati Vaucelle
Media Lab Europe
cati@mle.ie

Abstract

This sketch presents a real-time system that aims to bridge a gap between machine listening technology, and self-editing, self-synchronizing video: a movie organizes itself from "listening" to music. In our current demonstration, a series of short video clips of "Cati" dancing, originally shot at different tempi, are arbitrarily sequenced, always in sync with the analyzed beat, i.e., if the music slows down, the dance slows down accordingly. When no beat is found, e.g., the music is mellow, then Cati stops dancing and waits, apparently bored.

1 Introduction

When Kenneth Anger presents a costumed character moving through a garden of fountains in "Eaux d'Artifice" in the 50's, he manipulates lights, fountains, and the water synchronized with Vivaldi's music. The drama of expectation is often driven by the rhythm between music and images. If today, existing systems allow us to synchronize them properly, it is typically a manual and time-consuming process. *Cati Dance*, however, engages the user to direct the manipulation of video by changing music on a portable MP3 player connected via audio to the computer. Not only he/she can express a musical choice, but also the video then reacts in a narrative manner, e.g., a lack of beat is causing the dancer Cati to stop dancing, and to wait, apparently bored. When the rhythm changes abruptly, Cati reacts in surprising ways, trying to keep up with the new beat. It takes generally a few seconds to adjust. The system extends the user's artistic expression by satisfying his/her expectation of the inter-relationship between music and video.

2 Description

The *Cati Dance* project is a real-time system that plays arbitrary video clips of a database, always in sync with incoming music. Our demonstration counts about 20 clips (from 5 to 15 seconds each) of a woman dancing (see Cati in Figure 1). We also use about 10 clips of Cati simply waiting (see Figure 2). The video database must be previously prepared, whereas the audio remains unknown in advance, and can arbitrarily be chosen. Preparation of a video clip consists of manually cropping the video sequence on particular visual *beats* (i.e., found visually in the movie, or with the help of the original audio track), and annotating the clip name with its total number of beats in a text file. The audio track can be discarded, and the video is re-sampled at 60 bpm, i.e., 1 second per beat. The system relies on two main components: an audio *beat tracker*, and a feedback-controlled *movie editor*.

Beat tracker: Our audio beat-tracking algorithm (also called foot-taping) was inspired by Eric Scheirer's [1], and performs equally well on most popular music, however using less processing power. It was required to save power on audio analysis, in order to edit, and display full-screen video on the same machine. Our model analyzes the audio signal with sliding G4-optimized-FFTs provided by Apple. The power spectrum is converted into an auditory-model filterbank, which mimics the human cochlea. A large series of comb filters return the resonant periodicity amplitude in each channel for 60 tempi ranging from 80 to 140 bpm, which then are summed across to give the global energy-per-tempo estimations. The maximum corresponds to the current music tempo. Analysis of the internal phase of the best-matching resonator, allows us to predict the location in time of the next *beat*. The internal gain gives us an estimation of *strength*.

Movie editor: The movie-editing algorithm takes only two inputs (the audio beat, and strength information), three arguments (two text files and a strength threshold), and displays the final movie on the screen. Each text file contains a list of names for the available video clips (a list for "dancing" clips, and a list for "waiting" clips), and their corresponding number of beats (for the "dancing" list only, as described above). The inferred beat of the video is constantly compared to the analyzed beat of the audio, and a feedback-controlled mechanism allows us to regulate the internal frame rate of the movie, changing its speed appropriately. The displayed frame rate is however kept constant at about 30 fps. A new clip is randomly selected from a list, loaded, and queued so that transitions between clips are seamless. If the strength parameter is below the given threshold, then the clips from the "waiting" list are chosen.

A video of the system in action can be found at:
http://www.media.mit.edu/~tristan/Projects/catidance.html

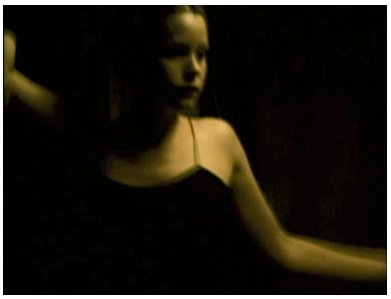

Figure 1. Screen shot of Cati dancing.

Figure 2. Screen shot of Cati waiting.

References

[1] SCHEIRER, E., 1998, *Tempo and Beat Analysis of Acoustic Music Signals,* Journal of the Acoustic Society of America, 103(1).