Designing Time-Based Interactions With Multimedia

Eric Lee
Media Computing Group
RWTH Aachen University
52056 Aachen, Germany
eric@cs.rwth-aachen.de

ABSTRACT

The current model of time in multimedia frameworks poses particular problems when designing multimedia systems with time-based interaction. We propose to expand and extend an existing distinction between semantic time and real time from music and film theory to multimedia systems design. The semantic time concept also forms the foundation of a new software framework for multimedia systems that we are building, that, unlike most existing frameworks, includes mechanisms for time-based effects such as time-stretching.

Categories and Subject Descriptors: D.2.11: Software Architectures – *Patterns*; D.2.2: Design Tools and Techniques – *Software Libraries, User Interfaces*

General Terms: Design, Theory

Keywords: time design, time-based media, multimedia frameworks, multimedia systems

1. INTRODUCTION

Despite the abundance of research in multimedia processing algorithms and ever-increasing power of computers, building interactive multimedia systems where time plays an important role in the interaction remains rather difficult. For certain media such as audio and video, time plays an essential role in semantics (a single time instant of audio has no semantically meaningful interpretation, for example); thus, it is natural to extend the dimensions of interaction to include time. In fact, for certain types of content such as music, control over time is an essential part of expressivity, and musicians often take this control over time for granted.

Our previous work on interactive conducting systems [1, 5] allows users to control the speed, volume and instrumentation of a digital audio and video recording of an orchestra using basic conducting gestures. While evaluating these systems with users, we have confirmed that users most readily grasp the ability to control the tempo over the other parameters of volume and instrumentation.

Designing systems which support temporal interaction remain, unfortunately, quite difficult. We attribute these difficulties to insufficient theoretical constructs for designing/describing temporal interactions with time-based media, and an inability to incorporate time-based effects, such as time-stretching, into interactive multimedia systems using current, well-known multimedia frameworks such as Microsoft's DirectShow and Apple's QuickTime.

MM'05, November 6–11, 2005, Singapore.
ACM 1-59593-044-2/05/0011.

We describe below our research objectives and status of our current work in time design for interactive multimedia systems; this work can be divided into the conceptual, consisting of time design theory, and the practical, in the form of a new multimedia software framework.

2. TIME DESIGN CONCEPTS

Much of our work on time in interactive multimedia systems is inspired by the way time is handled in music; music theory, having evolved over hundreds of years, includes mechanisms for describing not only temporal duration in terms of bars, beats and notes, but also its time derivatives, such as tempo and acceleration.

Music researchers have, for many years, distinguished between a musical piece's *semantic time*, dictated by the musical score, and *real time*, dictated by the performance of the score. This distinction, for example, is the basis for Jaffe's work on tempo maps [3], a theoretical construct for studying musical performance and expressivity.

Coincidentally, this distinction between semantic time and real time is also made by cinematographers [2], who manipulate time by assembling cinematic shots filmed at different locations and times to produce an entirely new film sequence. This film sequence has, furthermore, a running (real) time, which is usually different from the story (semantic) time. Time in film theory, however, is treated more as an artistic mechanism, and thus lacks the formal equivalent of time maps for music.

Generalizing this concept of semantic time to other media, such as speech, is thus one of our objectives. Much like we use "steps" and "blocks" to more naturally describe distance in American cities, we believe there exist various units of semantic time depending on the media type and application.

3. SEMANTIC TIME FRAMEWORK

Current well-known multimedia frameworks, such as Microsoft's DirectShow and DirectSound, or Apple's QuickTime and Core Audio, require system designers and engineers to work in units of real time, such as video frames and PCM audio samples. Thus, it is difficult to incorporate time-based effects, such as time-stretching, in an interactive multimedia system, since users manipulate semantic time, such as timing between music beats. The non-trivial mapping between semantic time and real time complicate even a simple application that allows the user to adjust the speed of synchronized video and audio playback without any unwanted audio pitch-shifting artifacts.

In an attempt to alleviate these complications, we are designing a new software multimedia framework based on semantic time. Like many multimedia frameworks, our Semantic Time Framework allows developers to assemble a directed, acyclic graph of effects

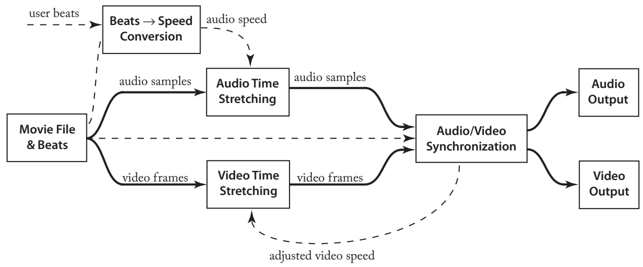

Figure 1: Block diagram of our previous interactive conducting system. Thick lines show the data flow, and dotted lines show control flow.

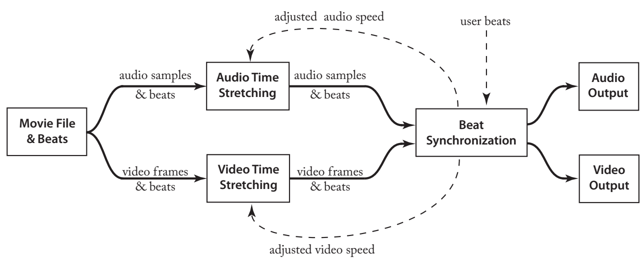

Figure 2: Block diagram for our latest conducting system, using the Semantic Time Framework. Compared to Figure 1, there are fewer components and links between them.

which operate on streams of audio and video. Unlike current, well-known multimedia frameworks, however, a stream in our framework contains semantic time metadata which is preserved throughout the processing chain[1].

As an initial step, we have designed and built a prototype framework based on these principles [4]. Using this framework, we have been able to create an interactive conducting system with a simpler and more elegant system design (see Figures 1, 2).

4. EVALUATION PLAN

In evaluating this work, we can consider the validity of our design theories, and the usefulness of the Semantic Time Framework. We plan to adopt an iterative design-implement-analyze approach, incrementally expanding and refining our theories at each design cycle. In the first iteration, which has already begun, we will incorporate this work in our interactive conducting systems. In the second iteration, we will expand our application to other computer music systems, such as a project we have begun for studying the relationship between dance movements and musical rhythm. In further iterations, we would then move beyond computer music systems, and expand our applications to include other forms of time-based media, such as speech or cinema.

We are still in the process of determining specific criteria by which our work can be evaluated. We will start by analyzing how our theories and framework are adopted in various interactive multimedia systems. We have already shown, for example, the validity of our concepts when applied to the specific case of an interactive conducting system for digital audio and video recordings. Adopting a similar approach and show how existing interactive multimedia systems with time-based interaction could benefit from this framework.

We strongly believe a complete evaluation of our work also requires feedback from other members of the research community. Thus, we are in the process of open sourcing the design and implementation of the Semantic Time Framework to enable others to adopt the framework in their systems.

[1]Apple's latest version of the Core Audio framework for digital PCM audio includes an "AUTimePitch" effect that allows one to adjust speed independently of pitch. However, the overall architecture of Core Audio has not changed, and implementing custom time-based effects using this framework remains non-trivial.

5. FUTURE DIRECTIONS

We have identified a number of possible directions for future work:

Time design patterns. As we continue to design and build interactive systems for time-based media, it is likely that we will continue to expand our repertoire of time design concepts. Once these concepts have matured, they may eventually become time design patterns, recipes which capture our experiences with designing such systems which can be reapplied to other, similar systems.

Dimensions of semantic time. We have established a clear distinction between the two dimensions of semantic time and real time thus far. Are there other dimensions of time? For example, a film sequence not only contains story time and real time, but also the source times of the individual clips which were reedited to form the scene.

Units of semantic time. We have established that music beats are useful units of semantic time for music. What units make sense for other media types, such as speech or cinema?

Time-based effects. Time-stretching is a rather semantically simple operation on the temporal axis, similar to scaling the brightness of an image. One could imagine more sophisticated time-based effects, such as one to remove "dead" time where the audio is silent, or an effect which adjusts the micro-timing within a beat to create a "swing" effect in music.

6. ACKNOWLEDGEMENTS

The author would like to think Jan Borchers and Rafael Ballagas for their helpful comments. *Personal Orchestra* was coordinated by Max Mühlhäuser, now at Darmstadt University, and in cooperation with the Vienna Philharmonic. *You're the Conductor* was developed in cooperation with Teresa Marrin Nakra at Immersion Music, and the Boston Symphony Orchestra.

7. REFERENCES

[1] J. Borchers, E. Lee, W. Samminger, and M. Mühlhäuser. Personal orchestra: A real-time audio/video system for interactive conducting. *ACM Multimedia Systems Journal Special Issue on Multimedia Software Engineering*, 9(5):458–465, March 2004. Errata published in next issue.

[2] D. Bordwell and K. Thompson. *Film Art: An Introduction*. McGraw-Hill, New York, 2003.

[3] D. Jaffe. Ensemble timing in computer music. *Computer Music Journal*, 9(4):38–48, 1985.

[4] E. Lee, T. Karrer, and J. Borchers. Towards a framework for interactive systems to conduct digital audio and video streams. To appear in: *Computer Music Journal*, 30(1), 2006.

[5] E. Lee, T. Marrin Nakra, and J. Borchers. You're the conductor: A realistic interactive conducting system for children. In *Proceedings of the NIME 2004 Conference on New Interfaces for Musical Expression*, pages 68–73, Hamamatsu, Japan, June 2004.