Conducting A Realistic Electronic Orchestra

Jan O. Borchers
Computer Science Department
Stanford University
Stanford, CA 94305-9020, USA
borchers@stanford.edu

Wolfgang Samminger
Telecooperation Research Group
University of Linz
4040 Linz, Austria
sam@tk.uni-linz.ac.at

Max Mühlhäuser
Telecooperation Group
Darmstadt University of Technology
64283 Darmstadt, Germany
max@informatik.tu-darmstadt.de

ABSTRACT

Personal Orchestra is the first system to let users conduct an actual audio and video recording of an orchestra, using an infrared baton to control tempo, volume, and instrument sections. A gesture recognition algorithm interprets user input, and a novel high-fidelity playback algorithm renders audio and video data at variable speed without time-stretching artifacts. The system is installed as a public exhibit in the HOUSE OF MUSIC VIENNA.

KEYWORDS: Conducting, orchestra, exhibit, gestures, time-stretching

INTRODUCTION AND RELATED WORK

The HOUSE OF MUSIC VIENNA is an exhibition center for Vienna's rich musical culture. *Personal Orchestra* is an interactive exhibit for this environment to let visitors conduct a recording of the Vienna Philharmonic Orchestra. Most conducting systems use gesture recognition for use by professional conductors, but play back only synthesized audio (MIDI) and no video [6, 7, 1, 4, 10, 5]. Others render the orchestra, but as artificial 3-D characters [3], or use real audio and video material, but do not let the user control tempo [9].

DESIGN

The UI design was based on our *HCI design patterns* for interactive exhibits [2]. We created a dedicated room reminiscent of a concert hall. A large projection shows the orchestra rehearsing until a user enters the podium and picks up an infrared baton. After language and piece selection and a brief introduction, the user can start conducting.

With museum visitors as users, we chose a simple up/down conducting style. Vertical gesture size controls overall volume. Conducting in the direction of instrument sections raises their volume. If the user keeps "teasing" the orchestra by conducting extreme tempi, the players react quite naturally—they stop, one of them gets up and complains. Otherwise, the piece ends with applause for the conductor.

SYSTEM ARCHITECTURE

Baton movements are sensed by a tracker and sent to a first computer, *POServer*. It determines tempo, volume, and orchestra section emphasis by gesture recognition and prediction, and sends this via our TCP-based, textual *Personal Orchestra Control Protocol (POCP)* to a second machine, *POClient*. That machine plays back the audio and video, continuously adjusting playback parameters to follow the conducting, and reporting back current movie positions. During language and piece selection, and upon finishing or breaking off, *POServer* sends POCP commands to *POClient* to display corresponding screens and movies.

Client and server run on Apple Power Macs G4/500 under Mac OS 9, Java & QuickTime. Audio and video are streamed from disk. *POClient* sends video to a 110in rear projection, and audio into a 2x2 high-end system with front and rear speakers for locatable sound sources and ambient audio.

GESTURE RECOGNITION AND PREDICTION

A Buchla *Lightning II* [8] translates input from an infrared-emitting, battery-operated baton, received by a tracker mounted under the screen, into standard 7-bit MIDI controller signals representing baton coordinates $x(t)/y(t)$ and baton button state.

$-$ to $+$ sign changes of $y'(t)$ (trajectory bottom) are "downbeats", and correspond to movie positions we marked as the audible beats in the music. User tests showed conductors expect to hear each beat shortly before this turning point; we shifted the downbeat movie positions accordingly. Playback speed is then adjusted so that the orchestra always follows the conductor. However, even at the same speed, the orchestra may be out of phase with the conductor. Also, when the conductor speeds up, a remainder of the current beat has not yet been played when the next downbeat arrives, so the orchestra needs to "catch up" (similarly for slowing down).

To tackle this, we adapted an algorithm used for synchronizing networked clocks: since jumping ahead or back in time would create discontinuities, synchronization has to be done by changing clock speeds. To catch up, playback speed is increased above the new conducting rate until movie and conductor are in sync, then leveled back off to the actual new conducting speed. The larger this time window for catching up, the longer it takes the orchestra to get back in sync

with the conductor, but the more tolerant it becomes against tempo jitter from inexperienced conductors. This "tightness" parameter can be changed at system startup to match the conductor's abilities.

To avoid constantly shifting emphasis between instruments when users conduct with a large x amplitude, the system changes emphasis only after the average conducting direction has remained in a section for a few beats. This slower reaction also makes the system behave more naturally.

HIGH-QUALITY INTERACTIVE MULTIMEDIA TIME-STRETCHING

We recorded the orchestra, playing various pieces without a conductor, in broadcast quality from the conductor's perspective. Time-stretching the video interactively by repeating or dropping frames is straightforward, as long as frame rates do not drop below animation rates (around 12fps). With our rather static scene, extreme movements creating jerkiness at high speed, or strange-looking movements due to the "wrong" physics in slow-motion, did not pose a problem.

Changing audio speed, on the other hand, would normally also change audio pitch. Fourier-transforming the signal and retransforming it with a changed duration preserves pitch, but introduces reverb when slowing down since every part of the signal is prolonged equally. Simple granular synthesis (repeating or leaving out ca. 50ms packets of the audio signal to adjust its duration) creates echo effects when packets with attack sounds are repeated. By preexamining the audio, more complex algorithms get closer to the near-CD quality we required, but take 10-50x the clip duration for processing on current hardware.

For that reason, we intended to pre-time-stretch our audio material off-line in high quality at various speeds, and then blend over between these pre-stretched versions to change speed during playback. However, this would have introduced different time coordinate systems for each audio track. Instead, we pitch-shifted the audio in half-tone steps of a factor of $\sqrt[12]{2}$, up to +/–1 octave. This changes pitch without changing tempo (using the same techniques as time-stretching). Since pitch-shifted audio tracks keep the same duration, a tempo change now simply meant blending over to the appropriate audio track and simultaneously changing playback speed of the entire movie to bring that track back to its original pitch. To implement emphasizing of instrument sections, we pitch-shifted the three section tracks separately, and mix them in real time according to emphasis during playback.

EVALUATION

Hundreds of visitors use our system as *Virtual Conductor* in the HOUSE OF MUSIC VIENNA each day. In questionnaires we gave to 30 users there, 81% judged audio quality as "good", 19% as "sufficient". Video quality was judged to be "good" by 75%, "sufficient" by 21%. 93% considered the exhibit among the Top 3 in the HOUSE OF MUSIC VIENNA.

SUMMARY AND FUTURE WORK

Personal Orchestra lets users conduct not a synthetic MIDI/VRML orchestra, but a real audio/video recording of the Vienna Philharmonic, using a novel way to time-stretch multimedia in real time and high quality. Natural conducting gestures with an infrared baton control volume, section emphasis, and even tempo of the performance without noticeable artifacts—although the orchestra will not tolerate notoriously bad conducting. HCI design patterns for interactive exhibits guided our design. We are currently working on further improving the conducting recognition of the system. Please see http://www.actibits.com/ for technical information about the system, and http://www.house-of-music-vienna.at/ for information about the center.

REFERENCES

1. Jan O. Borchers: WorldBeat: Designing a baton-based interface for an interactive music exhibit. *Proc. CHI 97*, ACM, 1997, 131–138.

2. Jan O. Borchers: *A pattern approach to interaction design.* 264 pages, John Wiley & Sons, New York, 2001 (http://www.hcipatterns.org/).

3. T. Ilmonen: The Virtual Orchestra performance. *Proc. CHI 2000*, ACM, 2000.

4. M. Lee, G. Garnett, and D. Wessel: An Adaptive Conductor Follower. *Proc. International Computer Music Conference*, ICMA, San Francisco, 1992.

5. Teresa Marrin Nakra: Inside the Conductors Jacket: analysis, interpretation and musical synthesis of expressive gesture. *PhD thesis, Massachusetts Inst. of Technology,* February 2000.

6. Max V. Mathews: The Conductor Program and Mechanical Baton, in Max V. Mathews and J. R. Pierce, eds.: *Current Directions in Computer Music Research,* MIT Press, Cambridge, 1991.

7. David Pogue: The dangers of the Digital Orchestra. *New York Times Direct,* Apr 5, 2001.

8. R. Rich: Buchla Lightning II. *Electronic Musician,* **12**(8), Cardinal Business Media, Emeryville, CA, August 1996, 118–124.

9. Music Box (now closed), Sony Center Berlin, 2000, http://www.sony-center.de/sonycenter_eng/-entertainment/music_box/c_index.html.

10. S. Usa and Y. Mochida: A conducting recognition system on the model of musicians' process. *Journal of the Acoustical Society of Japan*, **19**(4), 1998.