In this paper we present a system being developed in the context of the i-Maestro project \[1\][2], for the analysis of gesture and posture in string practice training using 3D motion capture technology. The system is designed to support the teaching and learning of bowing technique and posture, providing multimodal feedback based on real-time analysis of motion capture data. We describe the pedagogical motivation, technical implementation and features of our system.

1. INTRODUCTION

String players often use mirrors to observe themselves practicing. Recently we have seen the increased use of video recording in music education as a practice and teaching tool \[3][4]. For the purpose of analysing a performance these tools are not ideal since they offer a limited 2D, fixed perspective view. Using 3D Motion Capture technology it is possible to overcome this limitation by visualising the instrument and the performer in a 3D environment \[2][3]. Furthermore, the motion capture data can be analysed to provide relevant feedback about the performance. Feedback can then be delivered to the users in a variety of ways in both real time (online) and delay-time (offline) situations. This can help teachers to illustrate different techniques and can also to identify a student’s problems. It can help students to develop awareness of their playing and to practice effectively.

This paper presents a system for the analysis and visualisation of 3D motion data with synchronised video and audio. It discusses the interface between Max/MSP and a motion capture system, a library of Jitter objects developed for analysis and visualisation, and preliminary work on a sonification module for real-time auditory feedback based on analyses of motion capture data.

2. RELATED WORK

Although a great deal of research in the area of new musical interfaces looks at the violin family, the majority of this work lies in the context of contemporary composition and performance or lab-based gesture analysis. There are few examples of these interfaces being developed specifically for music pedagogy applications.

3D motion capture has been used in \[4\] to assist in piano pedagogy by visualising the posture of a professional performer and comparing it to that of the student. Motion capture has also been featured in a double bass instruction DVD \[5\] to help illustrate elements of bowing technique. The use of auditory feedback to teach musical instrument skills has been discussed in \[6\]. For related research on the sonification of motion data, see \[8\] and \[9\]. For the mapping of motion data to music see \[7\].

3. ANALYSES AND FEEDBACK

Through our own experience, discussion with teachers, and by studying existing string pedagogy literature we have identified areas of interest for analysis. These include:

- Relationships between the bow and instrument body (e.g. the angle between the bow and the bridge, the part of the bow that is used and the point at which the bow meets the strings).
- 3D trajectories/shapes traced by the bow and bowing arm (see Figure 1)
- Body Posture (e.g. the angles of the neck, bowing arm and wrist)

Figure 1. A screenshot of the 3D augmented mirror interface showing synchronised video and motion capture data with 3D bowing trajectories.
4. 3D MOTION CAPTURE

The details of the motion capture process have been covered sufficiently elsewhere [7][8][9] so here we briefly explain our setup.

4.1. System

We use a VICON 8i Optical Motion capture system\(^1\) with twelve infra red cameras (200fps) and a passive marker system. The high frame rate allows us to capture rapid bowing motions in detail and slow them down significantly maintaining an adequate resolution.

4.2. Markers and Models

We attach markers to the bow (see Figure 3), the instrument and the performer’s body. Markers are placed strategically in order to:

1. maximise visibility to the cameras
2. minimise the risk of displacement
3. minimise interference with the performance
4. minimise damage to the instrument
5. provide the necessary information for the required data analyses.

During the performance it is desirable to use as few markers as possible, to reduce the amount of marker data and to allow the performer as much freedom as possible. This may result in the loss of some information that may be necessary for data analyses. The solution is to create a static model of the instrument with a large number of markers. Once we have this model we can remove some of the markers and use the simplified model for the performance (see Figure 2). The system reconstructs the locations of the missing markers in real time by applying offsets based on the static model.

5. 3D AUGMENTED MIRROR

Our system, which we call the “3D Augmented Mirror”, is implemented in Max/MSP and Jitter using the MAV (Motion Analysis and Visualisation) framework. Rather than providing a fixed configuration, our system allows different analysis and visualisation modules to be scripted dynamically. This functionality is essential since we do not want to restrict users to certain types of analysis nor do we want to fix the user interface to any particular configuration.

The 3D Augmented Mirror supports synchronised recording and variable rate playback of motion capture data, audio and video.

5.1. Interface to the Motion Capture System

We developed an application (called VICONbridge) which streams the data from the VICON system to Max using the TCP/IP protocol. The application parses incoming motion data and sends the filtered and reformatted data to the \texttt{jit.net.recv} object.

5.2. MAV Framework

The MAV framework is designed as a collection of algorithms and Jitter objects based on a shared library. The objects perform various analysis and visualisation tasks using low level data processing and dynamic binding to realise a highly effective and flexible framework. Although we do not explicitly deal with video processing, the Jitter API offers a great amount of freedom over the standard Max API. Since Jitter objects are not “physically” bound to a Max patcher, it is possible to dynamically script the MAV objects at runtime\(^2\).

We chose Lua\(^3\) as the scripting language to implement the dynamic environment within Max, using the \texttt{jit.gl.lua} external developed by Wesley Smith\(^4\).

\(^1\) http://www.vicon.com
\(^2\) See Jitter API reference: http://www.cycling74.com/
\(^3\) http://www.lua.org
\(^4\) http://www.mat.ucsb.edu/~whsmith/
advantage of using Lua is its ability to link custom C modules. By linking Lua to the same dynamic library as the MAV externals we are able to access and share any data type or function between the library and Lua scripts, keeping CPU intensive tasks in C. We use this technique to separate visualisation algorithms from data analysis and processing.

5.3. Data Representation
Our system stores motion capture data using the Motion Analysis’ TRC format, with audio and video as separate files. We are currently adding support for SDIF [11] which will allow us to store synchronised audio, motion and analysis data in one file.

We use several SDIF data streams in order to combine multiple representations of data. The first stream contains standard time domain sampling (1TDS) matrices for the audio data. A second stream contains all motion capture data structured for every frame and stored in a custom matrix type with columns for ID, X, Y and Z data and a row for each marker. The third stream will contain multiple custom defined matrices for storing analysis data. SDIF data is represented by time-tagged frames, which allows streams at arbitrary data rates to be synchronised.

This approach is compatible with that described in [13] and would facilitate the correlation of data obtained with other types of gesture capture system e.g. IRCAM’s augmented violin [12].

6. ANALYSIS METHOD
Using the MAV library, the 3D Augmented Mirror system can perform a number of different data analyses by processing the 3D motion data. These analyses can be visualised using graphs and other graphical representations based on OpenGL (see Figure 1 and Figure 5).

6.1. Dynamic Coordinates System
The data from the motion capture system gives us the absolute location of each marker. To perform several of the desired analyses, it is necessary to derive a local co-ordinates system based on specific markers rather than using arbitrary world co-ordinates. For example, to monitor bowing, the local coordinates system will typically be positioned on the instrument body in order to focus on the interactions between the bow and the instrument.

6.2. Bow Stroke Segmentation
Using a local coordinates system we can extract the direction of the bowing motion by calculating the difference in distance (over consecutive frames) between the tip of the bow and the point at which the bow intersects the plane that is perpendicular to the fingerboard plane (see figure 5). A change of direction indicates a segmentation point (i.e. change of bow). This segmentation can reveal the connection between bow strokes and other characteristics of the performance.

6.3. Orientation
By choosing two markers, the user may define a vector and monitor its orientation in relation to the local coordinates system. By setting the two markers on the bow and a local origin on the bridge, we can use this process to analyse the angle of the bow in relation to the bridge.

6.4. Bowing Gestures
Our system provides the option to visualise the trajectories drawn by bowing movements (see Figure 1). Taking this further, we are segmenting and studying individual shapes. There are a number of interesting features that may be extracted from these shapes, for example the direction and orientation of the shape, its bounding volume and its “smoothness”. We are particularly interested in studying the relationship of these shapes to the music being played and to different bowing techniques. For example, Figure 6 shows the passage being played in Figure 1, which creates a clear figure of eight shape.

7. SONIFICATION OF DATA ANALYSES
We are using sonification as an alternative means of representing the information gathered from analyses of 3D motion data. This has potential for a number of reasons:

---

http://www.motionanalysis.com
Sound is the primary feedback mechanism in instrumental performance, therefore it is logical to assume that to a musician auditory feedback may be more relevant and useful than visual feedback.

When a student is practicing, they may not be able to use visual representations [6]. Their eyes may be occupied reading the score.

Sonification allows better temporal resolution which is appropriate for monitoring rapid movements such as those involved with bowing technique.

Sonification may allow multiple streams of information to be communicated to the user simultaneously.

7.1. Implementation
We have implemented a sonification tool in Max/MSP, initially using the angle of the bow in relation to the instrument as input data in order to monitor “parallel bowing”. We provide different options including controlling synthesised sounds and also processing the instrument sound in order to create a sonification that is related to the performance. We believe this may offer a more appropriate and usable feedback than synthesised sounds.

Sonification may be discreet or continuous depending on the user’s preference. Discreet mode allows two separate settings to monitor (user defined) “correct” and “incorrect” angles of the bow. In continuous mode, the angle of the bow is sonified so the performer will be made aware of the angle as they play. The parameter mapping may be adjusted for both modes so that deviations in the input can provide a different range of sound modification.

Since it may be undesirable to constantly hear the sonification, we provide a threshold control. This allows the sonification to interject only when necessary.

8. CONCLUSION AND FUTURE WORK
We presented a system for the analysis of 3D motion capture data within the Max/MSP environment and multimodal delivery of analysis data. Although this system is firstly aimed at studying string performance, it is by no means limited to this functionality. Development is ongoing, and we plan to add more analysis, visualisation and sonification possibilities based on further discussion with teachers and students. This process will involve trials and validation in real pedagogical situations.

We plan to integrate our system with other technologies related to the i-Maestro project such as MPEG SMR [10] and score/gesture following [12]. Primarily we are interested in ways of annotating gesture analysis data onto the score and indexing motion capture recordings to musical phrases.

9. ACKNOWLEDGEMENTS
The i-Maestro project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i-maestro.org). Thanks to all i-Maestro project partners and participants, for their interest, contributions and collaboration.

10. REFERENCES