SDIF Integration in i-Maestro Gesture tools

Thijs Koerselman[1], Oliver Larkin[1], Bee Ong[1], Nicolas Leroy[2], Jean-Philippe Lambert[2], Diemo Schwarz[2], Fabrice Guedy[2], Norbert Schnell[2], Frederic Bevilacqua[2], and Kia Ng[1]
[1]ICSRiM - University of Leeds, School of Computing & School of Music, Leeds LS2 9JT, UK
www.icsrim.org.uk
[2]IRCAM – CNRS STMS, 1 place Igor-Stravinsky, 75004 Paris, France
www.ircam.fr

info@i-maestro.org, *www.i-maestro.org*

Abstract

We discuss two prototype pedagogical applications that allow the study of string instrument bowing gesture using different motion capture technologies. Data exchange between the two applications is achieved using the Sound Description Interchange (SDIF) format which has, until now, been used primarily as a format for storing audio analysis data. We describe our method of storing motion- and analysis- data using the SDIF format.

1. Introduction

The storage and exchange of music-related gesture information is an emerging topic in the field of computer music [1][2][3]. Increased research into musical movement using sensors and motion capture systems has resulted in many researchers seeking a standardized file format for the description, storage and exchange of gesture data [1][2]. Currently there are a variety of formats that are used for storing motion data, none of which sufficiently meet the needs of musical gesture researchers since they are typically designed for more generic motion capture applications. Over the last ten years the Sound Description Interchange Format (SDIF) [4] has been adopted as the standard file format for storing audio analysis data, and allows many different audio applications to share data, thanks to its flexible, extensible structure. A similar extensible format is desirable for the exchange of gesture data where many different data types may be required for representing synchronous data from different devices including analysis data/descriptors [3].

The definition of taxonomy of gesture data types is the goal of projects such as the Gesture Description Interchange Format (GDIF) [3], Gesture and Motion Signals Format (GMS) [5], and Performance Markup Language (PML) [6].

A significant part of the i-Maestro Project involves the research and development of novel pedagogical tools and paradigms based on the analysis of musical gesture, with a focus on string practice training. The project, which is now in its third year, has produced two main software tools for this purpose: The Sound and Gesture Lab (S&G Lab) [7] and the 3D Augmented Mirror (AMIR) [8]. These tools offer a complementary approach to studying musical gesture, since they are each focused on different capture technologies and elicit different modes of interaction. Both tools are developed using Cycling 74's Max/MSP (www.cycling74.com) and make extensive use of external libraries. S&G Lab is based on IRCAM's FTM & Co. [9] and AMIR is based on the MAV Framework developed at ICSRiM [10].

1.1 Sound and Gesture Lab

S&G Lab allows for the manipulation and processing of sound and gesture data. At the lowest level, the application enables the simultaneous recording of audio and movement data. It offers the possibility to represent synchronized multi-modal data graphically as well as in terms of manipulatable internal representations and persistent files (using SDIF). On a higher level, the application permits the analysis of the data to extract musically relevant audio features from a live or a pre-recorded performance. This analytical view facilitates detailed study of a performance. Additionally, using real-time audio rendering opens the possibility for creating new correspondences between gesture and sound, therefore enabling the sonification of data and/or data analysis results for novel types of auditory feedback (sonification).

The application also provides recognition of a performance based on gesture data using the gesture-follower. This aspect supports music practice involving novel types of interaction paradigms between performer and computer.

The embedded score-follower can align the live input to a symbolic music representation, therefore creating a correspondence between multimodal signals and the symbolic representation of a performance.

These elements can be combined in different ways to support specific aspects of music lessons and scenarios for instrument practice, experimental improvisation and composition as well as allowing interactive approaches to music theory and ear training.

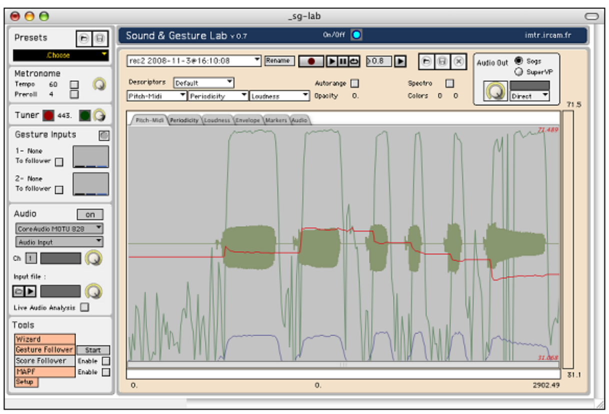

Figure 1. S&G Lab

1.2. The 3D Augmented Mirror

Like S&G Lab, AMIR is a tool for studying a musical performance. The software is designed as a real-time interface to a 3D Motion Capture system, which allows recording and analysis of motion data related to string performance, including bowing technique and the performers' posture. AMIR provides multimodal feedback in both real-time and non real-time (i.e. when playing back a recorded performance) modalities.

A typical setup would present the user with a similar interface to that shown in Figure 2. AMIR's interface is realized in OpenGL and displays a 3D rendering of the motion capture recording extended to visualize features such as bowing angles. Additionally the display contains synchronized video recording and visual feedback of analysis algorithms using running graphs and simplified diagrammatic representations. Most of the interface and the configuration of analysis modules are dynamic and can be changed by the user of the system. AMIR also features a sonification module which allows the analysis output to be mapped to sound.

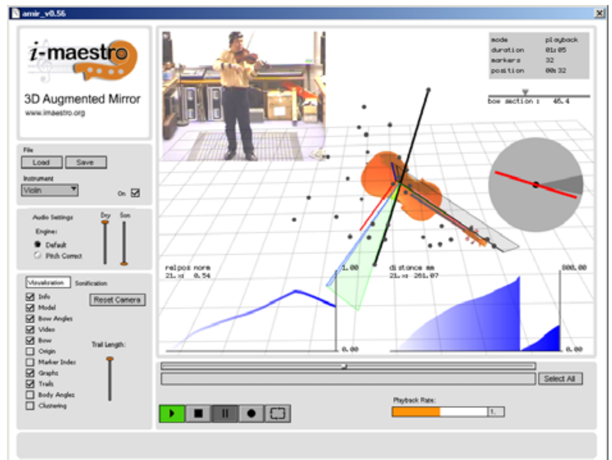

Figure 2. AMIR

2. Data Exchange

2.1. Motivation

By implementing the support for a common data format used by S&G Lab and AMIR, we allow the user to study the same performance using two complementary tools. This opens up new possibilities for visualizing and sonifying the data, providing a better insight and alternative ways for studying the players performance.

Data from the recording and analysis facilities in AMIR can be exported and used for further study in S&G Lab. Since S&G Lab has a broader function as a generic pedagogical tool for analyzing gesture and audio data, whereas AMIR has a more specific function, this data exchange currently only happens in one direction. At the time of writing there is no real need to exchange data in the opposite direction because AMIR is focusing on the analysis, visualization and sonification of 3D motion capture data. We are planning to integrate sensor input analysis into AMIR in the future, so we may reconsider bidirectional data exchange at a later stage.

2.2. The SDIF File Format

The SDIF standard arose out of the need to be able to store and exchange sound representation data between different analysis/synthesis programs, research teams, and institutions. The format enables anyone to interpret the information in the files correctly, needing as little external information as possible.

SDIF consists of a data format framework based on time-tagged frames containing matrices of binary numbers or text, and an extensible set of standard type declarations corresponding to different sound descriptions. Each frame holds a collection of one or more matrices containing the actual data. The file additionally specifies a list of all the non-standardized frame and matrix definitions used in the data. Since the standardized types are known by the SDIF framework, this information makes the file self-contained. The data descriptors in the file header provide the semantics that allow other applications to interpret the data correctly.

SDIF files typically include at least one NVT (Name Value Table) which is a special kind of matrix containing text data formatted into key-value pairs. These tables are often used to store global metadata such as time, date, version info, etc.

2.3. AMIR SDIF Integration

SDIF was chosen as the main file format for storing data in AMIR. The format allows us to combine the audio recording together with the motion capture data and the analysis in a single, standardized file type.

In AMIR we specify matrix and frame types for all of our data structures, since no suitable types exist in the current SDIF standard for storing gesture and motion related data. This is likely to change in the future since the standardization of gesture description is a key topic in current research into new musical interfaces [1][2][3].

The following table contains a list of the custom matrix definitions supported by the MAV framework. These matrix types are subject to change, and additional types are likely to be specified in the future.

XLAB	Text matrix with data labels formatted as a comma separated string (label01, label02, ...)
XINF	Text matrix with MAV data source information formatted as name-value pairs.
XPOS	Raw motion capture / positional data. Columns: x,y,z
XBDI	Bowing direction (where 0/1/2 == off / up / down bow). Columns: Direction
XBAN	Bowing angles. Columns: BowTilt, StringAngle, BridgeAngle.
XBSP	Bow speed in m/s. Columns: Speed.
XBAC	Bow accelerations in m/s^2. Columns: Acceleration
XBCP	Bow contact point, relative to the instrument orientation. Columns: x, y, z
XBPO	Bow position normalized to hair of the bow. Columns: Position
XBHI	Bowing height, normalized between the bridge and the end of the fingerboard. Columns: Height
XBDT	Distance traveled, measured in mm accumulated per bow stroke. Columns: Distance
XIOR	Instrument local-to-world transformation matrix (4x4). Columns: m00, m01, m02, m03, m10, m11 ... m33.
XIPO	Instrument position in world space. Columns: x,y,z.
XIRO	Instrument XYZ rotation in world space. Columns: x,y,z

Table 1. MAV Matrix Types

The XLAB and XINF matrices are currently only relevant to the MAV framework, and they are used to tag additional information to each of the streams. Since both matrices contain static data they are typically only written once to the first data frame of the stream.

The data structure of every MAV object in an AMIR configuration is bundled in a separate SDIF stream. In order to facilitate certain automated tasks for saving and loading data, the software needs to know the origin of each stream. The XINF matrix provides this information using key-value pairs with the source objects' class and ID. Additional keys can be used to add other object specific information. XINF uses the same internal string formatting as the standard NVT type.

All of the data sources in MAV can be labeled with names. In a motion capture system labels are typically used for each of the markers to easily distinguish their placement on the body or instrument. These labels are stored using the XLAB text matrix containing the values formatted as a comma separated string.

In AMIR the user is free to select any combination of the available data structures for SDIF i/o simply by specifying a list of IDs. With a single user action the data is then automatically formatted or parsed using the specified file. The SDIF implementation can be completely hidden from the user. Figure 3 shows a simplified high level layout of an example SDIF file exported by AMIR. This example contains a mono audio recording together with the content of two MAV objects; one containing the 3D motion capture data, the other containing the output from the bowing analysis algorithms.

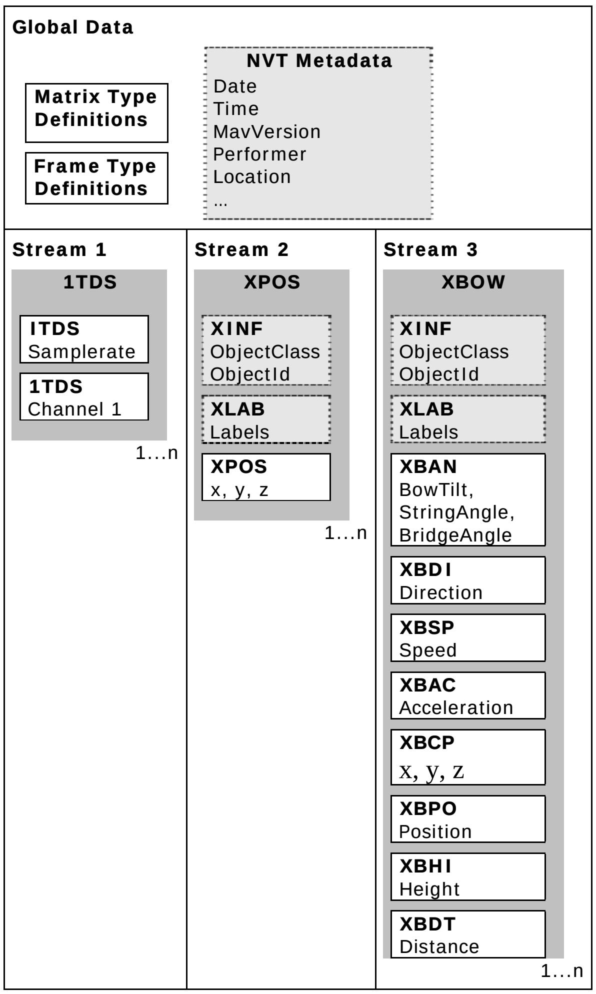

Figure 3. MAV SDIF File Layout

3. AMIR as an input to S&G Lab

In addition to the discussed file exchange, the two applications can also be used together in another modality, whereby AMIR streams data to S&G Lab in real-time via Open Sound Control (OSC) (www.opensoundcontrol.org). This is achieved easily since OSC is already used internally by S&G Lab to send data from sensors to the main application. In this way the gesture analysis parameters available from AMIR can be used as an input to S&G Lab.

4. Conclusion

We presented two applications for the study of musical gesture and discussed the integration and data exchange functionality between them. We have found SDIF to be a suitable data format for storing multi modal data representations. In this paper we described our current implementation for formatting motion data and related analysis outputs from AMIR. Future work in this direction would involve the adoption of any emerging format for the structured storage of gesture analysis data.

5. 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.

6. References

[1] Jensenius, A. Castagné, N. Camurri, A. Maestre, E. Malloch, J. McGilvray, D. “A Summary of Formats for Streaming and Storing Music-Related Movement and Gesture Data” in *Proc. of 4th International Conference on Enactive Interfaces,* Grenoble, France 2007

[2] Jensenius, A. and Camurri, A. and Castagné, N. and Maestre, E. and Malloch, J. and McGilvray, D. and Schwarz, D. and Wright, M., “The Need of Formats for Streaming and Storing Music-Related Movement and Gesture Data” in *Proc. of the International Computer Music Conference (ICMC2007),* Copenhagen, Denmark, 2007

[3] Jensenius, A. R., Kvifte, T. and Godøy R. I. "Towards a gesture description interchange format", in *Proc. of the International conference on New Interfaces for Musical Expression* (*NIME06)*, IRCAM, Paris, France

[4] Schwarz, D. and Wright, M. "Extensions and Applications of the SDIF Sound Description Interchange Format", *Proc. of the International Computer Music Conference (ICMC2000)*, Berlin, Germany, 2000

[5] Luciani, A., Evrard, M., Couroussé, D., Castagné, N., Cadoz, C., Florens, J-L. "A Basic Gesture and Motion Format for Virtual Reality Multisensory Applications" in *Proc. of the 1st international GRAPP Conference on Computer Graphics Theory and Applications*, Setubal, Portugal, 2006.

[6] Performance Markup Language (PML) http://www.n-ism.org/Projects/pml.php

[7] F. Guédy, N. Leroy, B. Bevilacqua, T. Grosshauser, N. Schnell, "Pedagogical experimentation using the Sound and Gesture Lab", in *Proc. of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS) - i-Maestro Workshop*, Barcelona, Spain, 2007.

[8] Ng, K., Larkin, O., Koerselman, T., Ong, B., Schwarz, D. Bevilaqua., F. "The 3D Augmented Mirror: Motion Analysis For String Practice Training", in *Proc. of the International Computer Music Conference (ICMC2007),* Copenhagen, Denmark, 2007.

[9] Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., Muller, R. "FTM - Complex Data Structures for Max", in *Proc. of the International Computer Music Conference (ICMC 05)*, Barcelona, Spain, 2005.

[10] Koerselman, T., Larkin, O., Ng, K. "The MAV Framework: Working with Motion data in Max MSP" in *Proc. of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS) - i-Maestro Workshop*, Barcelona, Spain, 2007.