A Summary of Formats for Streaming and Storing Music-Related Movement and Gesture data

Alexander Refsum Jensenius,[a] Nicolas Castagné,[b] Antonio Camurri,[c]
Esteban Maestre,[d] Joseph Malloch,[e] Douglas McGilvray [f]

[a]University of Oslo, Musical Gestures Group, a.r.jensenius@imv.uio.no
[b] ACROE, Grenoble, nicolas.castagne@imag.fr
[c] University of Genoa, Infomus lab DIST, antonio.camurri@unige.it
[d] Pompeu Fabra University, Music Technology Group, emaestre@iua.upf.edu
[e] McGill University, IDMIL, CIRMMT, joseph.malloch@mcgill.ca
[f] University of Glasgow, Centre for Music Technology, d.mcgilvray@elec.gla.ac.uk

Abstract

This paper summarises a panel discussion at the 2007 International Computer Music Conference on movement and gesture data formats, presents some of the formats currently in development in the computer music community, and outlines some of the challenges involved in future development.

1 Introduction

The rapid growth in research on enactive interfaces over the last years, and on movement and gesture in general, have shown the need for better methods, tools and techniques for handling what we will here refer to as *movement and gesture data*. One important challenge is the lack of generic formats for handling such data, something which often leads to compatibility problems when working with various hardware and software solutions. This issue has emerged as an important research topic in the *computer music* community over the last years. Considering that a computer music point of view may stimulate a larger discussion in the Enactive audience, this paper provides an overview of the solutions that are currently being worked on in this field.

While we have formats and standards for handling audio (AIFF, MPEG, etc.), audio analysis (SDIF), video (MPEG, QT, etc.), music notation (MusicXML), musical control data (OSC), etc., there are no widespread formats, nor structured approaches, for handling music-related movement and gesture data. In fact, most researchers store their data without using any specific format, or use the format of the specific device or application at hand [5]. This is a practical problem not only for the single researcher working with various types of equipment, but it also effectively hinders the sharing of data, tools and research methods between institutions.

Movement and gesture related studies have gained interest in the computer music community over the last years, and several research groups have started to work on solutions for standardising the way we store and stream movement and gesture data. Since several of these initiatives seemed to be unknown in the computer music community, we invited a number of researchers involved in the development of various movement and gesture data formats to a panel discussion at the 2007 International Computer Music Conference[1] in Copenhagen, Denmark [2]. This paper summarises the panel discussion, provides an overview of some of the formats in development by the authors, and points out some challenges for future development. Focusing mainly on the point of view of the computer music community, these formats may also be a starting point for a wider approach to encoding movement and gesture data.

2 Structuring Low level signals

A major challenge in the development of formats for handling movement and gesture data seems to be the lack of defining and structuring low-level movement and gesture *signals* or *streams*. We deal with low-level data representing performed movements and gestures, but there is no common agreement on how to describe, structure and encode such low-level data. While it is sometimes possible to work with device specific data, there is a growing need to record, store and exchange low-level data in a more generic way.

In the same way as the PCM audio format served as a foundation for the development of research on audio, we

[1] http://www.icmc2007.net

Proceedings of ENACTIVE/07
4th International Conference on Enactive Interfaces
Grenoble, France, November 19th-22th, 2007

find that the establishment of a generic, minimal format to structure and encode low level movement and gesture signals is crucial for further research on movement and gesture in fields such as computer music, enactive interfaces, computer graphics, virtual reality, etc.

2.1 Motion Capture Formats

The motion capture community has introduced a number of formats dedicated to storing and structuring motion capture data over the years. Some of these formats are used in the computer music community, but they are often far from sufficient for many of our needs, and therefore often create more problems than they solve.

One problem with several of these formats is that they are proprietary and designed to accompany specific hardware, something which does not give the openness and expandability that we look for. Another problem is that many of these motion capture formats focus on full-body motion-capture streams based on an articulated skeleton and a 3D-representation. This is often not general enough for many computer music applications where we are not only interested in describing human bodies, but also devices with different morphologies and dimensions, as well as information about tactility and haptics in the devices. In general, we therefore find existing motion capture formats too specific when it comes to dimensions, structure, number of degrees of freedom, and frequency characteristics (often also limited by storing in ASCII-files).

Yet another problem with many of the motion capture formats is the lack of possibilities for synchronising low-level data with mid- and high-level analytical results, as well as other types of data (e.g. music notation) and media (e.g. audio and video). This calls for more generic formats that can synchronise data with various resolutions and sampling rates (see section 3).

2.2 GMS

The Gesture and Motion Signal (GMS) format[2] has been developed in the EU Enactive Network of Excellence[3] by a subgroup of partners headed by the ACROE group [4]. It is a binary format intended for structuring, storing and streaming low-level movement and gesture signals as generically as possible, not only for computer music applications.

In GMS, a *gesture scene* can be encoded at any frequency rate (e.g. 100 Hz to a few tens KHz) and it is based on a two-level structure made of *gesture channels* and *gesture units* (Figure 1). A gesture channel allows for structuring the dimensions of the performed gesture; it can correspond either to an intensive variable (e.g. position) or extensive variable (e.g. force), that can be either 1D, 2D or 3D. A gesture unit is made of a group of channels, and allows for structuring various recorded points/forces in a meaningful manner.

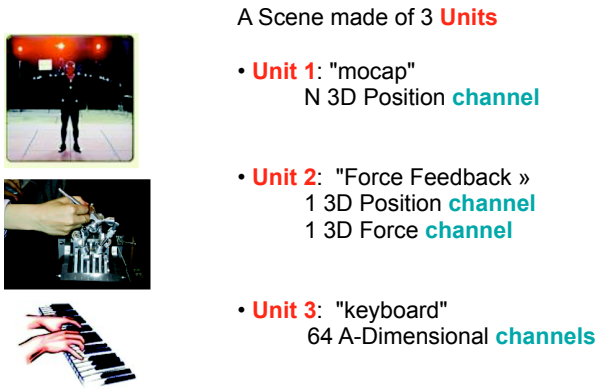

Figure 1: An example of a gesture scene structured and encoded with GMS

3 Mid- and high-level data

Much of the analysis and usage of music-related movement and gesture data is happening at what may be called mid- and high levels, e.g. focusing on phrases, expressivity, emotional response, etc. For this reason we need to find solutions to handle such data in a structured manner and to synchronise such data with the low-level data they are often derived from. There are several research groups involved in finding solutions for handling the structuring of such data, and three formats are currently being developed: GDIF, PML, XMI.

3.1 GDIF

The development of the Gesture Description Interchange Format (GDIF)[4] is a collaborative effort between researchers at the University of Oslo, McGill University and Pompeu Fabra university [3]. The focus is on creating structures for handling different levels of movement and gesture data: from raw data to higher level descriptors, as well as secure synchronisation with other types of data and media.

GDIF development is mainly focused on *what* to store and not *how* to store it, and is therefore based on existing formats and protocols, e.g. OSC, SDIF and XML (Figure 2). This allows for both streaming and storage, as well as compatibility with various computer music software and hardware. For realtime control, GDIF has been tested to control spatialisation [8] and creating a more structured and flexible approach to setting up mappings between various sensor devices and sound engines [7]. For the analysis of musical gestures,

[2] http://acroe.imag.fr/gms/
[3] http://www.enactivenetwork.org
[4] http://musicalgestures.uio.no

an XML-based implementation of GDIF is being developed for creating performance databases, exemplified through violin performance in [6].

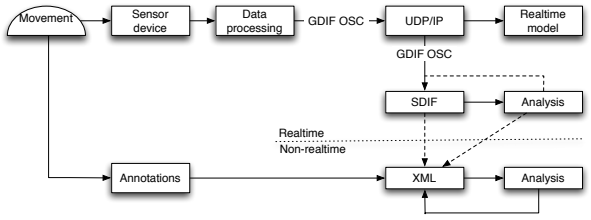

Figure 2: A GDIF setup for handling both streaming and storage of data using OSC, SDIF and XML

3.2 PML

Performance Markup Language (PML)[5] is an XML-based representation intended to facilitate the investigation of issues relating to musical performance. To investigate these issues, it is necessary to analyse performance artefacts in the context of the score. Therefore, the basic content of a PML file comprises the score, a basic markup of performance events, and a description of the correlation between individual performance and score objects.

The strict hierarchy of XML is not naturally suited to the multiple overlapping structures which are required to adequately describe musical information. Therefore, PML encourages information to be stored within separate hierarchies. These hierarchies can cross reference information in other informational hierarchies using internal relational links and pointers to locations within external files in formats such as PCM audio or GMS low level signals. Therefore PML allows existing formats to be combined into one representational system, allowing existing tools for manipulation of score, audio, video and gesture to be used.

3.3 EyesWeb XMI

The new EyesWeb XMI[6] (eXtended Multimodal Interaction) proposes a multi-layered gesture processing framework containing three layers: *MoCap*, *trajectories*, *cues and gestures*. This allows for working with: i) real-time, multimodal processing and interactive systems, and ii) data analysis and synchronised processing of pre-recorded data (see for example [1]).

EyesWeb XMI is supporting various geometric data types and compound data types such as collections to face the multifaceted and multi-layered problems that arise in gesture analysis research. It is possible to represent and process point-light display data, multicamera and multisensor data, as well as collections of different data and expressive gesture cues. A full set of automatic converters between the different layers and data types issupported.

[5] http://www.n-ism.org/Projects/pml.php

[6] http://www.eyesweb.org

4 Other Formats and Protocols

A critical aspect in computer music is the need to synchronise data with other types of data and media. Besides formats for handling audio, video, images and notation, there are two formats that have been established as ”standards” over the last decade: SDIF and OSC.

4.1 SDIF

The Sound Description Interchange Format (SDIF) was originally developed for handling audio and audio analysis data [10], and has been implemented in a number of software and programming environments [9]. The SDIF specification and implementation has already tackled a number of challenges relating to synchronisation of multiple streams of exogenous data, including high-speed data streams.

Even though SDIF was originally developed for storing audio data, it is a ”container” format that could easily be extended to carry necessary low-, mid- or high-level movement and gesture data (section 2 and 3). This, however, still requires development of taxonomies and structures for such data, as currently being developed in GMS, GDIF and PML.

4.2 OSC

Open Sound Control[7] (OSC) is an open, transport-independent, message-based protocol for communication between music hardware and software systems [12]. OSC has received increased interest over the years and is currently the *de facto* communication standard in the computer music research community, and is also slowly being introduced in various commercial systems (as an alternative to MIDI).

OSC does not solve the encoding and structuring of movement and gesture data, only the transport of the data. That is why it is necessary to develop solutions for a structured approach to creating OSC namespaces for streaming movement and gesture data, such as GDIF.

5 Summary

Standards-making seems to be an ongoing, iterative activity in the computer music community [11], and one can argue that the most successful formats in use are the ones that started by solving a specific problem for later

[7] http://www.opensoundcontrol.org

to be developed into a more generic standard. Such a bottom-up approach is, indeed, the approach taken by several of the authors in their various developmental efforts, including GDIF, GMS, PML and EyesWeb XMI.

The panel discussion at ICMC, and this follow-up paper has presented some of the current challenges and research efforts when it comes to movement and gesture data formats in the field of computer music. Some key elements for future development are to:

- create solutions for both performance (streaming) and analysis (storage).
- define, structure and encode low-level continuous movement and gesture signal data with different frequencies, resolutions, dimensions, etc., including various feedback loops.
- define, structure and encode mid- and high-level analytical data and descriptors, and synchronise these with related low-level data.
- handle synchronisation with musical notation, other types of data (e.g. annotations) and media (audio and video)
- support already existing formats and protocols used in the community, e.g. SDIF, OSC, MusicXML.

The various formats developed by the authors approach different aspects of the above-mentioned problems, and by uniting research efforts it may be possible to ensure interoperability between the different formats. An important point here is that of cross-disciplinary collaboration. Similar problems relating to structuring and encoding movement and gesture data are currently being tackled by researchers in various fields, and much research still needs to be carried out both conceptually and technologically. By joining efforts, we may be able to more efficiently reach our goals of generic solutions for handling movement and gesture data. Hopefully this paper may stimulate such further collaboration.

Acknowledgements

Thanks to Diemo Schwarz, Matt Wright, Stuart Pullinger and Ben Knapp for participating in the ICMC panel, and for valuable comments and suggestions.

References

[1] A. Camurri, G. Castellano, R. Cowie, D. Glowinski, B. Knapp, C. L. Krumhansl, O. Villon, and G. Volpe. The premio paganini project: a multimodal gesture-based approach for explaining emotional processes in music performance. In *Gesture Workshop, Lisbon*, 2006.

[2] A. R. Jensenius, A. Camurri, N. Castagne, E. Maestre, J. Malloch, D. McGilvray, D. Schwarz, and M. Wright. Panel: the need of formats for streaming and storing music-related movement and gesture data. In *Proceedings of the 2007 International Computer Music Conference*, Copenhagen, Denmark, forthcoming 2007. San Francisco: ICMA.

[3] A. R. Jensenius, T. Kvifte, and R. I. Godøy. Towards a gesture description interchange format. In *NIME '06: Proceedings of the 2006 International Conference on New Interfaces for Musical Expression*, pages 176–179, Paris, France, 2006. Paris: IRCAM – Centre Pompidou.

[4] A. Luciani, M. Evrard, N. Castagné, D. Couroussé, J.-L. Florens, and C. Cadoz. A basic gesture and motion format for virtual reality multisensory applications. In *Proceedings of the 1st international Conference on Computer Graphics Theory and Applications, Setubal, Portugal, March 2006*, Setubal, Portugal, 2006.

[5] A. Luciani, M. Evrard, D. Courousse, N. Castagne, I. Summers, A. Brady, P. Villella, F. Salsedo, O. Portillo, C. A. Avizzano, M. Raspolli, M. Bergamasco, G. Volpe, B. Mazzarino, M. M. Wanderley, A. R. Jensenius, R. I. Godøy, B. Bardy, T. Stoffregen, G. De Poli, A. Degotzen, F. Avanzini, A. Roda, L. Mion, D'Inca, C. Trestino, and D. Pirro. Report on Gesture Format. State of the Art. Partners' propositions. Deliverable 1 D.RD3.3.1, IST-2004-002114-ENACTIVE Network of Excellence, 2006.

[6] E. Maestre, J. Janer, M. Blaauw, A. Pérez, and E. Guaus. Acquisition of violin instrumental gestures using a commercial EMF tracking device. In *Proceedings of the 2007 International Computer Music Conference*, Copenhagen, Denmark, 2007. San Francisco: ICMA.

[7] J. Malloch, S. Sinclair, and M. M. Wanderley. From controller to sound: Tools for collaborative development of digital musical instruments. In *Proceedings of the 2007 International Computer Music Conference*, Copenhagen, Denmark, 2007. San Francisco: ICMA.

[8] M. T. Marshall, N. Peters, A. R. Jensenius, J. Boissinot, M. M. Wanderley, and J. Braasch. On the development of a system for gesture control of spatialization. In *Proceedings of the International Computer Music Conference*, pages 360–366, New Orleans, LA, 2006. San Francisco: ICMA.

[9] D. Schwarz and M. Wright. Extensions and applications of the SDIF sound description interchange format. In *Proceedings of the 2000 International Computer Music Conference*, pages 481–484, Berlin, Germany, 2000. San Francisco: ICMA.

[10] M. Wright, A. Chaudhary, A. Freed, D. Wessel, X. Rodet, D. Virolle, R. Woehrmann, and X. Serra. New applications of the sound description interchange format. In *Proceedings of the 1998 International Computer Music Conference*, pages 276–279, Ann Arbor, MI, 1998. San Francisco: ICMA.

[11] M. Wright, R. Dannenberg, S. Pope, X. Rodet, X. Serra, and D. Wessel. Panel: Standards from the computer music community. In *Proceedings of the 2004 International Computer Music Conference, Miami, FL*, pages 711–714, 2004.

[12] M. Wright, A. Freed, and A. Momeni. Opensound control: State of the art 2003. In *NIME '03: Proceedings of the 2003 International Conference on New Interfaces for Musical Expression*, 2003.