PANEL: THE NEED OF FORMATS FOR STREAMING AND STORING MUSIC-RELATED MOVEMENT AND GESTURE DATA

Alexander Refsum Jensenius [a], *Antonio Camurri* [b], *Nicolas Castagné* [c], *Esteban Maestre* [d], *Joseph Malloch* [e], *Douglas McGilvray* [f], *Diemo Schwarz* [g], *Matthew Wright* [h]

[a] University of Oslo, Musical Gestures Group, a.r.jensenius@imv.uio.no
[b] University of Genoa, Infomus lab DIST, antonio.camurri@unige.it
[bc] ACROE, Grenoble, nicolas.castagne@imag.fr
[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
[g] IRCAM, Centre Pompidou, diemo.schwarz@ircam.fr
[h] CNMAT, UC Berkeley and CCRMA, Stanford University, matt@cnmat.berkeley.edu

ABSTRACT

The last decade has seen the development of standards for music notation (MusicXML), audio analysis (SDIF), and sound control (OSC), but there are no widespread standards, nor structured approaches, for handling music-related movement, action and gesture data. This panel will address the needs for such formats and standards in the computer music community, and discuss possible directions for future development.

1. INTRODUCTION

There has been a rapid growth in research on music-related movement, action and gesture over the last years. This development has particularly been driven by a number of large European collaborative projects (e.g. MEGA, CoGAS, S2S^2, Enactive Network, TaiChi) that have addressed various aspects of body movement control of musical sound. One of the main challenges that many research groups are faced with, is the compatibility problems between various hardware and software solutions used. This problem mainly arises due to the lack of formats and standards for music-related movement and gesture data. The situation also makes it difficult to share data among researchers and institutions, since there is no common way to structure data and related analyses.

There have been various initiatives in the computer music community to solve this problem over the last couple of years, including the Gesture Description Interchange Format (GDIF) [1] [13], Gesture Motion Signal (GMS) [2] [17] and Performance Markup Language (PML) [3]. However, these formats are still in development, and relatively unknown to the computer music community at large, and there may be other ongoing initiatives that we are not aware of.

This panel proposal is therefore intended for starting a discussion in the music technology community, and to see if we can agree on some future development lines. We see this as a natural follow-up of the more general discussion about formats and standards at ICMC 2004 [28]. Now is the time to focus on the need for standards for streaming and storing movement and gesture data.

2. VARIOUS FORMATS AND STANDARDS

2.1. Motion Capture Formats

A number of formats exist for storing motion capture data, many of which were designed for specific hardware, e.g. the AOA format used with optical tracker systems from Adaptive Optics, the BRD format used with the *Flock of Birds* electromagnetic trackers, and C3D [4] used for Vicon infrared motion capture systems. Several formats have also emerged for using motion capture data in animation tools, such as the BVA and BVH formats from Biovision, and the ASF and AMC formats from Acclaim [16], as well as formats used by animation software, e.g. the CSM format used by 3D Studio Max.

Some of these motion capture formats are used in our community, but often they create more problems than they solve. One problem is that they often focus on full-body motion descriptors, i.e. based on a full articulated skeleton, which does not always scale well for our types of applications. It also makes them less ideal as a starting point for creating a generic format for encoding movement and gesture data. Another problem is that most standards are mainly intended for storing low level descriptors only, and leave little room for storing mid- and high-level analytical results and annotations. Finally, the lack of ability to synchronise with various music-related data (audio, video, midi, OSC, notation, etc.), make them even less ideal.

[1] http://musicalgestures.uio.no
[2] http://acroe.imag.fr/gms/
[3] http://www.n-ism.org/Projects/pml.php
[4] http://www.c3d.org/

2.2. Movement-related Markup Languages

There have been many attempts to create XML based standards for motion capture and animation data, e.g. the Motion Capture Markup Language (MCML) [5], Avatar Markup Language (AML) [15], Sign Language Markup Language [8], Multimodal Presentation Markup Language (MPML) [24], Affective Presentation Markup language (APML) [7], Multimodal Utterance Representation Markup Language (MURML) [14], Virtual Human Markup Language (VHML) [1], etc. It is difficult to judge the success of these formats, and the relevance for music-related movement research, but none of these formats stand out as candidates to fulfill our needs.

The same seems to be the case for the movement-related parts of MPEG-4 [11] and MPEG-7 [22, 20], which both seem to be geared towards commercial multimedia applications.

2.3. GMS

The Gesture Motion Signal (GMS) format [9, 17] has been developed by the ACROE group in Grenoble, and is also used in the EU Enactive Network of Excellence [5] . It is a binary format based on the Interchange File Format (IFF) standard [23], and is mainly intended for structuring, storing and streaming low-level movement and gesture signals. It was designed as a proposal for a generic structure for raw movement and gesture signals, for which there is currently no format available.

2.4. GDIF

The Gesture Description Interchange File Format started as a collaborative project between the University of Oslo and McGill University [13, 21], and is currently also being developed by researchers from Pompeu Fabra university [19]. The main focus of GDIF is to create structures to handle different levels of movement data: from raw data to higher level descriptors, as well as secure synchronisation with other types of data and media. GDIF is currently being developed as a namespace for OSC, an extension to SDIF, and as an XML description. This allows for both streaming and storage, as well as compatibility with software and hardware in the computer music community.

2.5. PML

The Performance Markup Language (PML) is developed as an extension to the Music Encoding Initiative (MEI) [6] [25] at the University of Glasgow. The main idea is to create a structured approach to annotate performance data in relation to musical notation.

[5] http://www.enactivenetwork.org

[6] http://www.lib.virginia.edu/digital/resndev/mei/

3. NEEDS

We see a number of different needs for working with music-related movement and gesture data. First of all, there are many unsolved conceptual and practical problems when it comes to structuring raw data from various devices (e.g. MIDI instruments and NIMEs) in a generic way. This is further complicated by our needs to formalise descriptors for associated body movement data, and various mid- and high level features. This section will outline some of the needs we see for future research.

3.1. Different Types of Raw Data

A first step is to work towards a generic structure for raw movement and gesture signals. We are usually working with a large number of different hardware devices, all of which use different protocols, formats and, in a few cases, standards. For example:

- *Motion capture systems.* Such systems typically output data at high speeds (up to 4000 Hz) for a number (anything from 2 to 50) of multidimensional markers (often 3 or 6 degrees of freedom (DOF)). Motion capture systems usually use their own proprietary formats for storing the data.
- *MIDI devices*. Most commercial instruments only output MIDI, which is an event-based protocol for command signals, and thus hardly corresponds with movement and gesture signals.
- *Commercial controllers*. Game controllers, graphical tablets, mice, and other commercial devices usually comply to a well-known protocol and use more or less well-defined ranges and resolutions. As with MIDI, there are no standards for describing the functionality of the devices, or the movements and gestures associated with them.
- *Custom made instruments and devices*. We often work with special sensor systems and custom made interfaces, many of which exist in only one example. While many of the devices rely on some type of protocol for data transfer (MIDI, OSC, etc.), there is no structured way for handling the movements and gestures performed on such devices.

One of the biggest challenges seems to be the lack of good definitions of movement and gesture signals or streams and how they should be structured. This is very different from the audio world, where a sound signal can be identified by certain properties, e.g a sampled signal at 8-96KHz, made of tracks, often stereo, 2D, 5+1D, etc.

Even though we may build on proposals from the motion capture community, e.g. 3D skeleton models, these are not sufficient for our needs. We are interested not only in describing bodies, but also various devices which are highly versatile in their morphology and dimensions. This is further complicated by our interest in working with different types of data resolutions and sampling rates. Finally, we are also interested in defining information about tactility and haptics in the devices.

3.2. A multilayered approach

While many formats allow for storing one level of data, e.g. raw or analysed data, we see the needs for streaming and storing multiple levels: *raw*, *pre-processed* and *analysed* data. This is no so important for streaming solutions, but for offline analysis we find it important to be able to store multiple streams of analytical results for the same raw data. This will make it possible to carry out collaborative studies between research institutions and comparative studies on the same material.

These multidimensional data sets should also be synchronised with various other types of data and media files (audio, video, midi, osc, notation, etc.). It is also important to be able to store qualitative data, e.g. observations, various types of metadata (e.g. expressive and emotional features [4, 3, 2]) and annotations synchronised with the quantitative data.

3.3. Streaming

Both for running experiments and for creating performance systems we need solutions for streaming data. The large variability of our data in terms of resolution and speed makes it a challenge to create a format which is both efficient and flexible enough. It is also a challenge to find solutions for streaming multiple streams based on different segmentation modes, or time lines.

When dealing with streaming in the context of computer music, Open Sound Control (OSC) has emerged as a standard in the research community. While the openness of OSC has certainly been liberating, it has also made it difficult for OSC-enabled systems to communicate efficiently. Attempts have been made to move towards uniform OSC namespaces (such as [29] and recent discussions in the OSC community), but there does not seem to be any consensus on how to actually describe such information.

Creating a structured approach to handling movement and gesture data within OSC is the current main priority of GDIF development. This implies formalising the structure of how to encode raw data and associated movement and gesture data using OSC namespaces [12].

3.4. Storage

There are several different needs when it comes to storing movement and gesture data. A typical scenario is the need for storing data for local analysis and retrieval from specific experimental setups. In such cases it is important to store enough descriptors and metadata to make the data sets clear and understandable for others.

A different type of use is the creation of shared databases. The need for sharing data between researchers and institutions is growing rapidly, and it would be of great interest to be able to compare data and analytical results. This would require a much more rigid way of storing and annotating data so that the utility can be useful for other researchers.

3.5. Synchronisation

An important point here is synchronisation with different other types of data. Synchronisation is, obviously, crucial when working with music-related data. When it comes to audio, and results of audio analysis, the Sound Description Interchange Format (SDIF) [27] offers the necessary framework, and is currently available in a number of software and programming environments [26]. The SDIF specification and implementation has already tackled a number of challenges relating to synchronisation of multiple streams of data, including high-speed data streams, and might also be extended to store movement-related data streams.

More conceptual problems arise when we want to synchronise with data which is based on relative (or no specific) time coding, e.g. symbolic music notation. Performance recordings based on musical notation usually vary considerably, and creating solutions for "time warping" data sets or creating musical "keyframes" have to be explored further. Here it is probably possible to integrate formats like MusicXML [7] [10, 6], Performance Markup Language (PML) [18], and the Music Encoding Initiative (MEI) [25].

4. PANEL DISCUSSION

The objective of the panel is to start a discussion in the computer music community about the need for formats and standards relating to movement and gesture data. This will be addressed by the following three questions to each of the panellists:

- How do you currently work with music-related movement and gesture data?
- What are your needs of formats and standards?
- How do we proceed from here?

All panellists are working with performance and/or analysis of music-related movements and gestures, and most are also involved in development of many of the formats, standards and frameworks presented in the paper. Hopefully, this discussion will increase the interest for these topics, and lead to continued development and more collaborative projects in the future.

5. REFERENCES

[1] S. Beard and D. Reid. MetaFace and VHML: A First Implementation of the Virtual Human Markup Language. *AAMAS workshop on Embodied Conversational Agents-let's specify and evaluate them*, 2002.

[2] 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*, Forthcoming 2007.

[7] http://www.recordare.com/xml.html

[3] A. Camurri, G. De Poli, M. Leman, and G. Volpe. Toward communicating expressiveness and affect in multimodal interactive systems for performing arts and cultural applications. *IEEE Multimedia*, 12(1):43–53, 2005.

[4] A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gesture: The eyesweb expressive gesture processing library. In *Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, pages 460–467. Springer-Verlag, Berlin Heidelberg, 2004.

[5] H. Chung and Y. Lee. MCML: motion capture markup language for integration of heterogeneous motion capture data. *Computer Standards & Interfaces*, 26(2):113–130, 2004.

[6] S. Cunningham. Suitability of MusicXML as a Format for Computer Music Notation and Interchange. *Proceedings of IADIS Applied Computing 2004 International Conference, Lisbon, Portugal*, 2004.

[7] B. De Carolis, C. Pelachaud, I. Poggi, and M. Steedman. APML, a mark-up language for believable behavior generation. *Life-like Characters. Tools, Affective Functions and Applications*, pages 65–85, 2004.

[8] R. Elliott, J. R. W. Glauert, J. R. Kennaway, and I. Marshall. The development of language processing support for the visicast project. In *Assets '00: Proceedings of the fourth international ACM conference on Assistive technologies*, pages 101–108, New York, NY, USA, 2000. ACM Press.

[9] M. Evrard, D. Couroussé, N. Castagné, C. Cadoz, J.-L. Florens, and A. Luciani. The gms file format: Specifications of the version 0.1 of the format. Technical report, INPG, ACROE/ICA, Grenoble, France, September 2006.

[10] M. Good. MusicXML: An Internet-Friendly Format for Sheet Music. *XML Conference and Expo*, 2001.

[11] B. Hartmann, M. Mancini, and C. Pelachaud. Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. *Proceedings of Computer Animation, 2002*, pages 111–119, 2002.

[12] A. R. Jensenius. GDIF Development at McGill. Short Term Scientific Mission (STSM) Report, COST 287 Action ConGAS, February 2007.

[13] A. R. Jensenius, T. Kvifte, and R. I. Godøy. Towards a gesture description interchange format. In *Proceedings of New Interfaces for Musical Expression, NIME 06, IRCAM - Centre Pompidou, Paris, France, June 4-8*, pages 176–179, 2006.

[14] A. Kranstedt, S. Kopp, and I. Wachsmuth. MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. *Proceedings of the AAMAS Workshop on 'Embodied conversational agents–Let's specify and evaluate them'*, 2002.

[15] S. Kshirsagar, N. Magnenat-Thalmann, A. Guye-Vuillème, D. Thalmann, K. Kamyab, and E. Mamdani. Avatar markup language. In *EGVE '02: Proceedings of the workshop on Virtual environments 2002*, pages 169–177, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[16] J. Lander. Working with motion capture file formats. *Game Developer*, January, 1998.

[17] 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.

[18] J. MacRitchie, N. J. Bailey, and G. Hair. Multi-modal aquisition of performance parameters for analysis of chopin's b flat minor piano sonata finale op.35. In *DMRN+1: Digital Music Research Network One-day Workshop 2006, Queen Mary, University of London, 20 December 2006*, 2006.

[19] E. Maestre, J. Janer, A. R. Jensenius, and J. Malloch. Extending gdif for instrumental gestures: the case of violin performance. In *Proceedings of the International Computer Music Conference*, Submitted 2007.

[20] B. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. John Wiley and Sons, 2002.

[21] 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, 6-11 November, New Orleans*, pages 360–366, San Francisco, 2006. International Computer Music Association.

[22] J. Martinez, R. Koenen, and F. Pereira. MPEG-7: the generic multimedia content description standard, part. *Multimedia, IEEE*, 9(2):78–87, 2002.

[23] J. Morrison. Ea iff 85: Standard for interchange format files. Technical report, Electronic Arts, January 14, 1985.

[24] H. Prendinger, S. Descamps, and M. Ishizuka. MPML: A markup language for controlling the behavior of life-like characters. *Journal of Visual Languages and Computing*, 15(2):183–203, 2004.

[25] P. Roland. The Music Encoding Initiative (MEI). *Proceedings of the First International Conference on Musical Applications Using XML*, pages 55–59, 2002.

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

[27] 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 International Computer Music Conference, Ann Arbor, Michigan*, pages 276–279, 1998.

[28] 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.

[29] M. Wright, A. Freed, A. Lee, T. Madden, and A. Momeni. Managing complexity with explicit mapping of gestures to sound control with OSC. In *Proceedings of the 2001 International Computer Music Conference, Habana, Cuba*, pages 314–317, 2001.