

TOWARDS A GESTURE DESCRIPTION INTERCHANGE FORMAT

ALEXANDER REFSUM JENSENIUS^{1,2}, TELLEF KVIFTE¹, ROLF INGE GODØY¹

¹Musical Gestures Group, Department of Musicology, University of Oslo

²Input Devices and Music Interaction Lab, Schulich School of Music, McGill University

INTRODUCTION

We see the need for a Gesture Description Interchange Format (GDIF) to describe, store, retrieve and share information about music-related gestures.

Current general purpose gesture formats developed within the motion capture industry and biomechanical community (e.g. C3D) focus mainly on describing low-level motion of body joints. We are more interested in describing gesture qualities, performer-instrument relations, and gesture-sound relationships in a coherent and consistent way.

A common musical gesture format will simplify working with different software, platforms and devices, and allow for sharing data between institutions.

NEEDS

- Realtime gesture control
- Non-realtime storage for analysis and annotation

DATA

- Raw data from various devices
- Pre-processed data from the devices
- Semantic gesture descriptors
- Gesture analysis
- Annotations

FORMAT REQUIREMENTS

- Free and open for everyone to use
- Simple to use and get started with
- Human-readable
- Multiplatform
- Flexible and extendable

IMPLEMENTATION

For realtime purposes we are currently using Open Sound Control (OSC) to share information between computers. The current implementation is therefore focused around developing GDIF as an OSC namespace. For storing the information it is possible to write OSC strings to a file with time-tags, but it might also be relevant to consider using XML or IFF for storage.

RAW DATA FROM DEVICES

Commercial controllers typically have a well-defined format (e.g. MIDI), but since they all differ we find it important to create a general way of storing the information. Examples of interfaces/controllers that are popular in music technology research:

- Sensor interfaces (MIDI, USB, OSC)
- Human interface devices (joysticks, gamepads)
- MIDI instruments
- Wacom graphical tablets
- Polhemus electromagnetic tracker
- Vicon motion capture system

OSC namespaces for such devices can be implemented like:

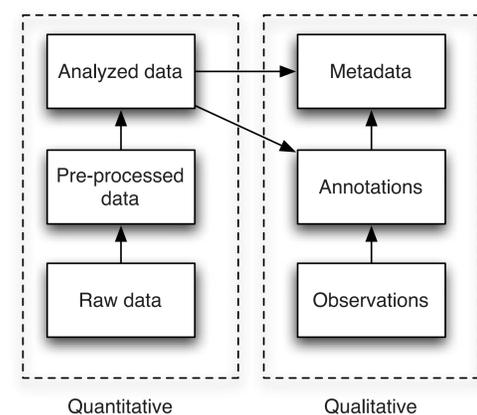
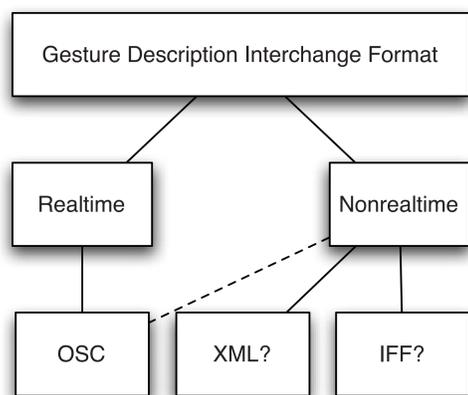
```
/midi/note/[pitch] [velocity]
/midi/ctl/[number] [value]
```

```
/hi/[number] [value]
```

```
/wacom/[x, y, pressure, xtilt, ...] [values]
```

```
/polhemus/[x, y, z, azimuth, elevation, roll] [values]
```

We think it is important to keep a namespace which is easily understandable for humans, even though it is less efficient in terms of processing.



PRE-PROCESSED DATA

Removing signal noise and normalizing the data is usually necessary before carrying out further analysis. We usually prefer to normalize data to a 0.-1. range, or possibly -1.-1. if movement is happening around a defined axis. But for example MIDI note information is better kept as is to avoid confusion. In current experimental setups we also tend to output several different layers of pre-processed data next to each other.

QUALITATIVE OBSERVATION DATA

For experimental setups it is necessary to store several layers of information:

- Session information describing equipment used etc
- Manual observation data
- Annotations
- General comments

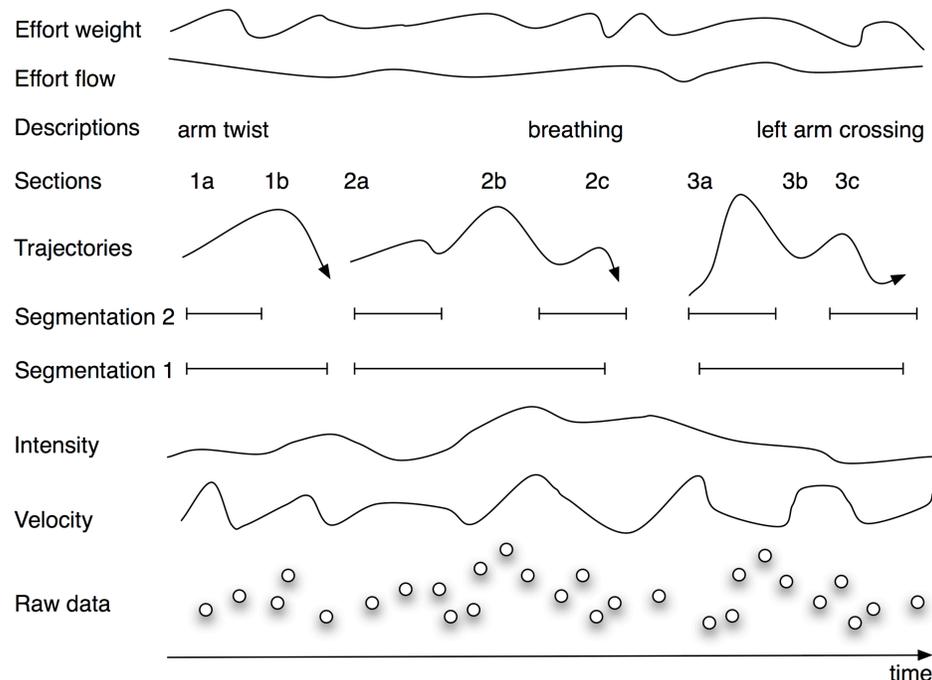
Such information is often crucial for understanding the quantitative data, and should be readily available in the format.

TOWARDS SEMANTICS

Working towards semantic descriptors of musical gestures is important for facilitating mapping. These could include Laban's Effort parameters (weight, space, time, flow), Juslin's emotion parameters (anger, grief, fear, joy), and other semantic gesture descriptors like directions (up-down, left-right) and intensity (high-low). Developing such descriptors are at the core of several research projects around the world, and an open GDIF format could allow storing such qualitative information in a structured way.

SEVERAL LAYERS

Human perception and cognition work at different attention levels and analytical perspectives all the time. Thus we find it important that gesture information is stored in such a way that the information is easily scalable both in precision and time resolution. Below is a draft displaying different types of information that we might be interested in storing in a GDIF file:



OPEN QUESTIONS

There are mostly open questions at this stage, but some important ones to consider are:

- Which type of time-base should be used?
- How should different files/layers be synchronized?
- What precision and resolution is necessary?
- Should data be stored sequentially or grouped based on segmentation?

ACKNOWLEDGMENTS

This research has been funded by the Norwegian Research Council. Thanks to Marcelo M. Wanderley, Mark Marshall and Nils Peters for comments and suggestions.