Formats for Streaming and Storing Music-related Movement and Gesture Data

Alexander R. Jensenius	Oslo
Benjamin Knapp [Antonio Camurri]	SARC [Genova]
Nicolas Castagné	ACROE, Grenoble
Esteban Maestre	Pompeu Fabra
Joseph Malloch	CIRMMT, McGill
Stuart Pullinger [Douglas McGilvray]	Glasgow
Diemo Schwarz	IRCAM
Matthew Wright [Matthew Wright]	CNMAT / CCRMA

ICMC Copenhagen 2007

Alexander R. Jensenius	GDIF
Benjamin Knapp [Antonio Camurri]	EBF
Nicolas Castagné	GMS
Esteban Maestre	GDIF-XML
Joseph Malloch	GDIF-OSC
Stuart Pullinger [Douglas McGilvray]	PML
Diemo Schwarz	SDIF
Matthew Wright [Matthew Wright]	SDIF + OSC

Outline

Introduction to the panel
Introduction by each panellist
Open discussion
Closing remarks

Streaming Synchronisation Storage

APML	Affective Presentation Markup language
AML	Avatar Markup Language
AOA	Adaptive Optics Format
ASF / AMC	Acclaim motion capture formats
BRD	Flock of Birds motion capture format
BVA / BVH	Biovision motion capture formats
C3D	Vicon motion capture format
CSM	3D Studio Max format
EBF	EyesWeb Binary Format
GDIF	Gesture Description Interchange Format
GMS	Gesture Motion Signal
MCML	Motion Capture Markup Language
MPEG 4/7	Motion Picture Expert Group formats
MPML	Multimodal Presentation Markup Language
MURML	Multimodal Utterance Representation Markup Language
OSC	Open Sound Control
PML	Performance Markup Language
SDIF	Sound Description Interchange Format
SLML	Sign Language Markup Language
VHML	Virtual Human Markup Language (VHML)

ENACTIVE Survey

30 %	use raw data (no format)
50 %	use a proprietary, home-made format
40 %	use the format of the device at hand
80 %	don’t use a unique format, but one per application
< 10 %	use a known, officially released format

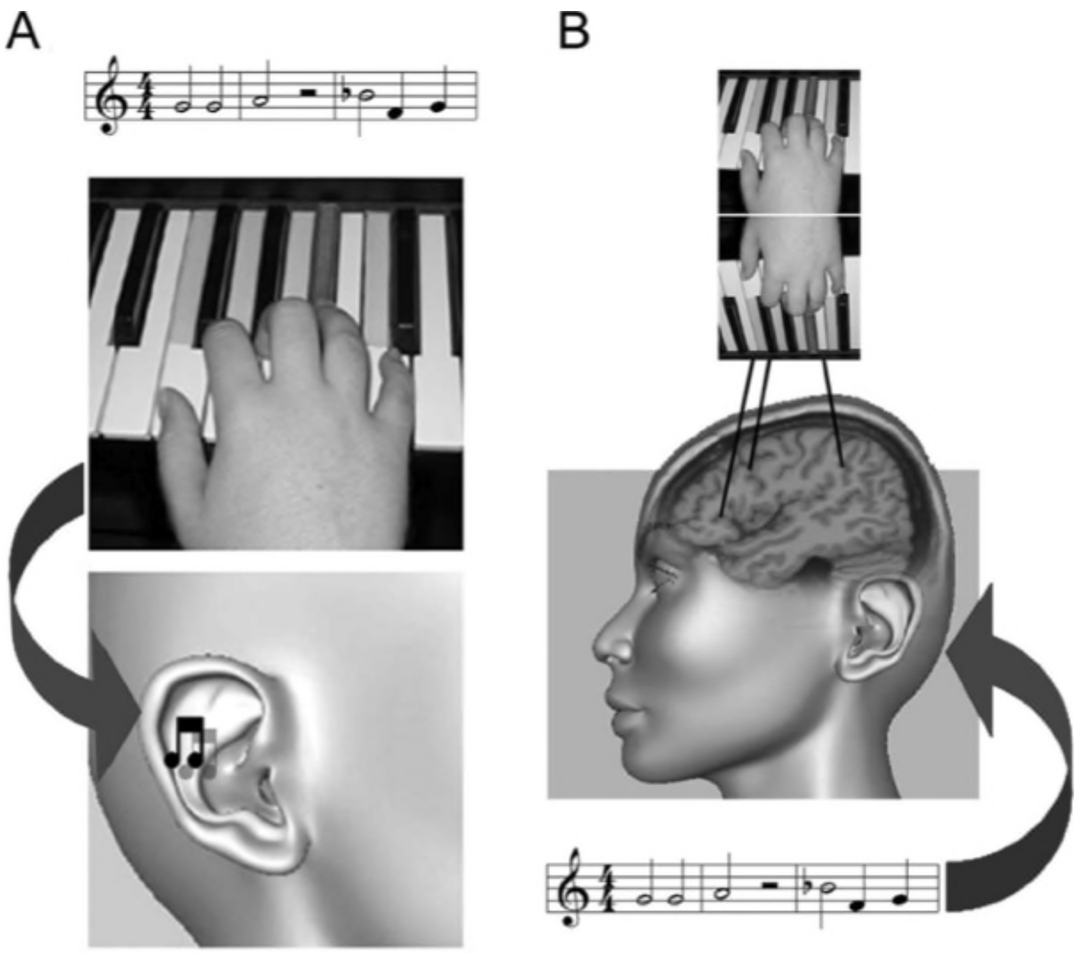
1. How do you currently work with music-related movement and gesture data?

2. What are your needs of formats and standards?

3. What are your suggestions for future development?

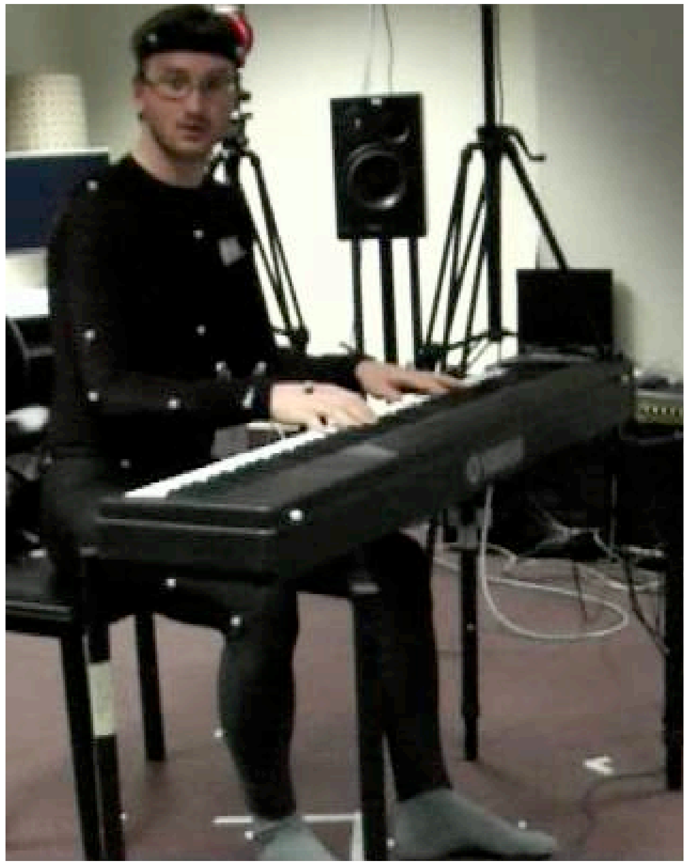
Alexander R. Jensenius

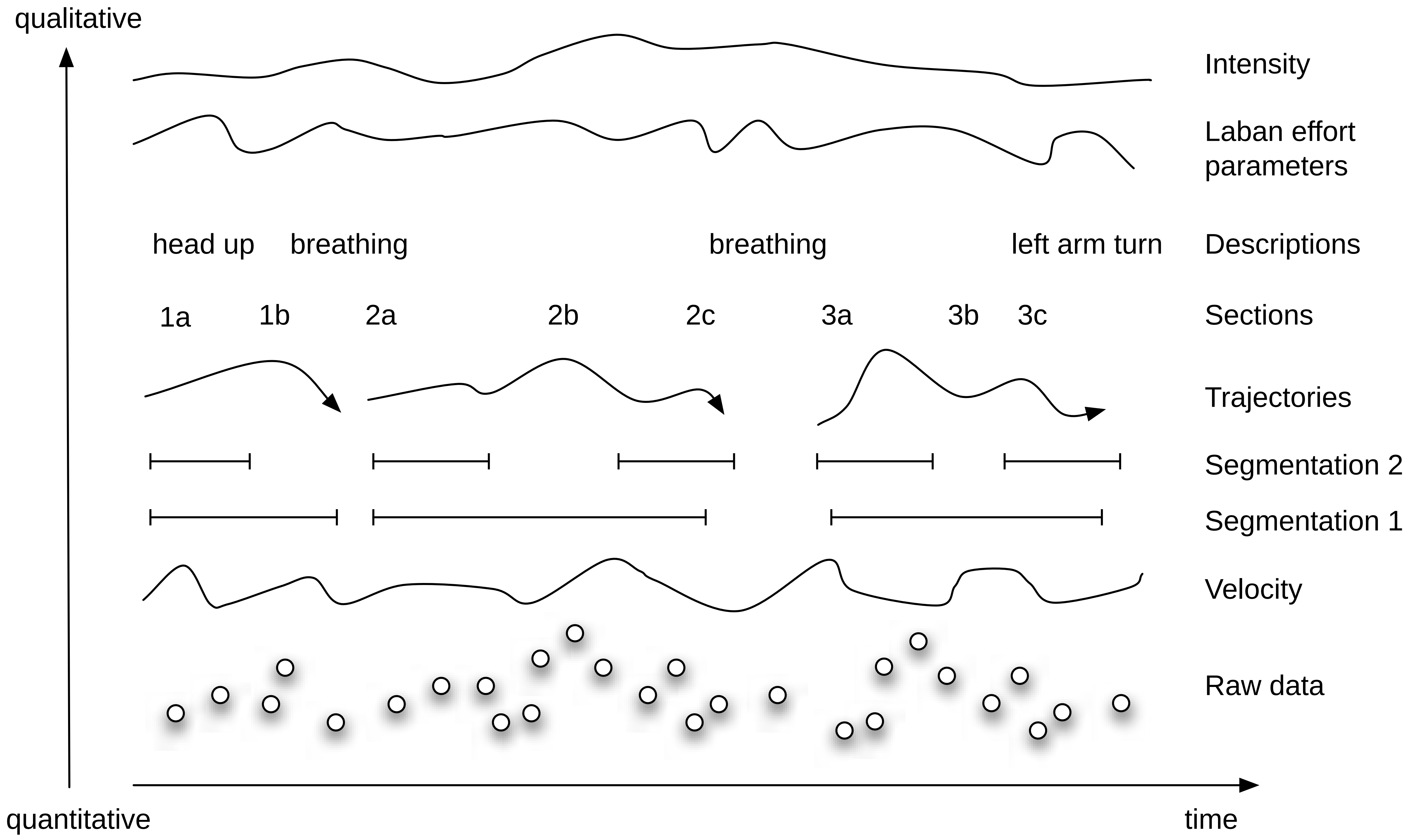
Musical Gestures Group
University of OSlo



Actions → Sounds

Sounds → Actions





qualitative

Intensity

Laban effort parameters

head up breathing breathing left arm turn

Descriptions

1a 1b 2a 2b 2c 3a 3b 3c

Sections

Trajectories

Segmentation 2

Segmentation 1

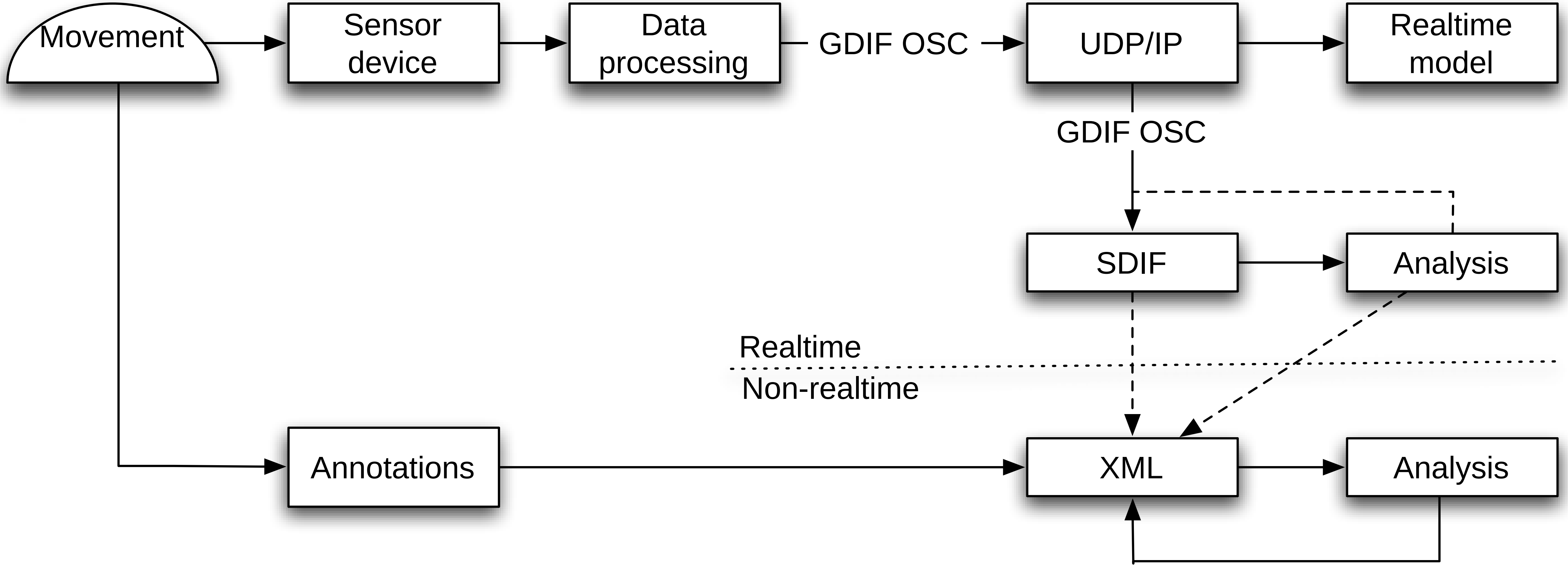
Velocity

Raw data

quantitative

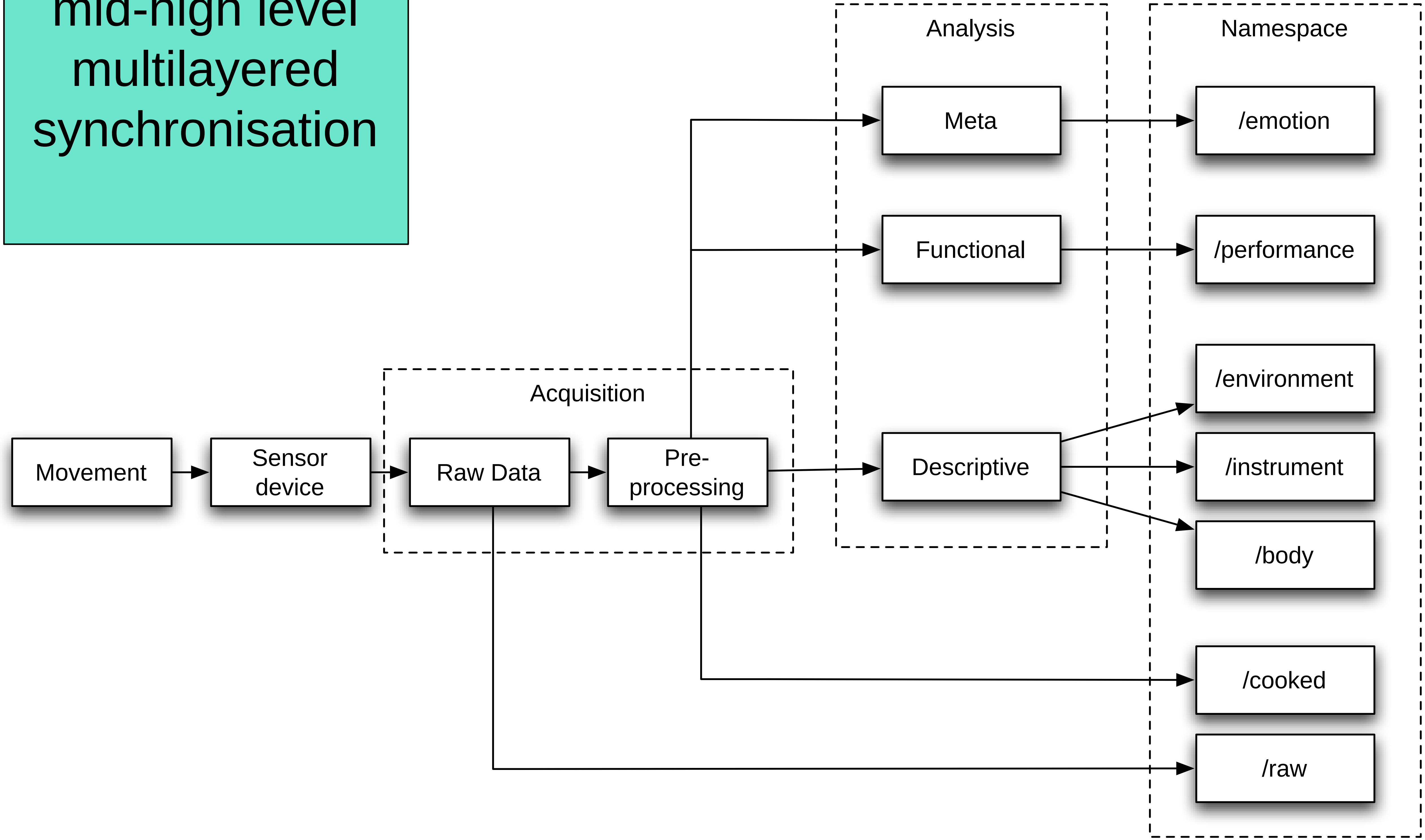
time

GDIF Gesture Description Interchange Format



GDIF Gesture Description Interchange Format

mid-high level multilayered synchronisation



Movement / Action / Gesture
Music Technology Group – Pompeu Fabra University

- Tangible musical instruments
- Gesture-based musical instrument synthesis
- Voice-driven interfaces
- Expressive performance analysis
- Embodied music interaction (ensemble performance)

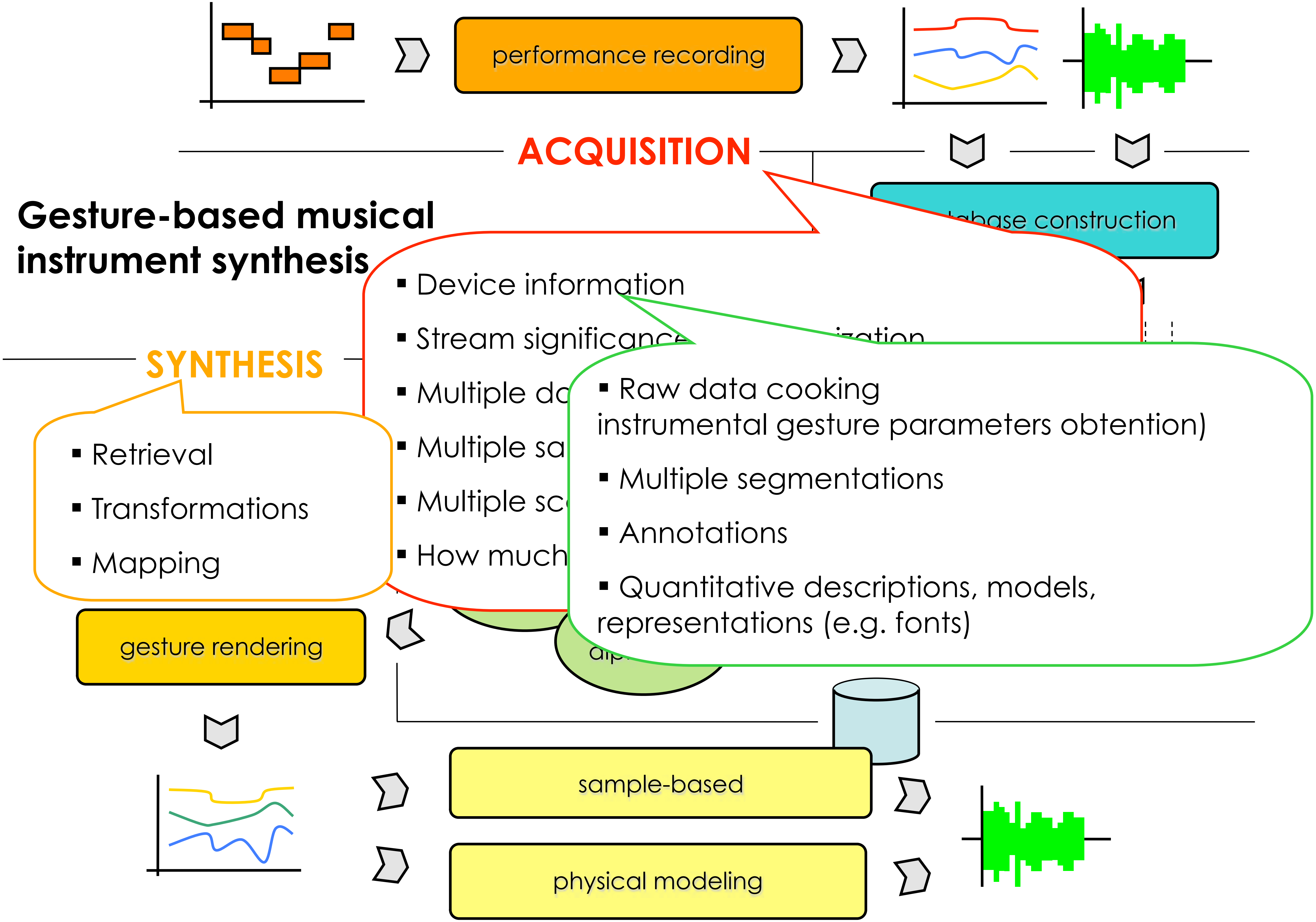
Application context
Gesture-based musical instrument synthesis

- Long term research line
- Focus on excitation-continuous instruments, instrumental gestures
- Explicitly introduce the performer into synthesis chain
- Study and model correlations between domains

Performer: score VS movement / action

Instrument: movement / action VS sound

Gesture-based musical instrument synthesis



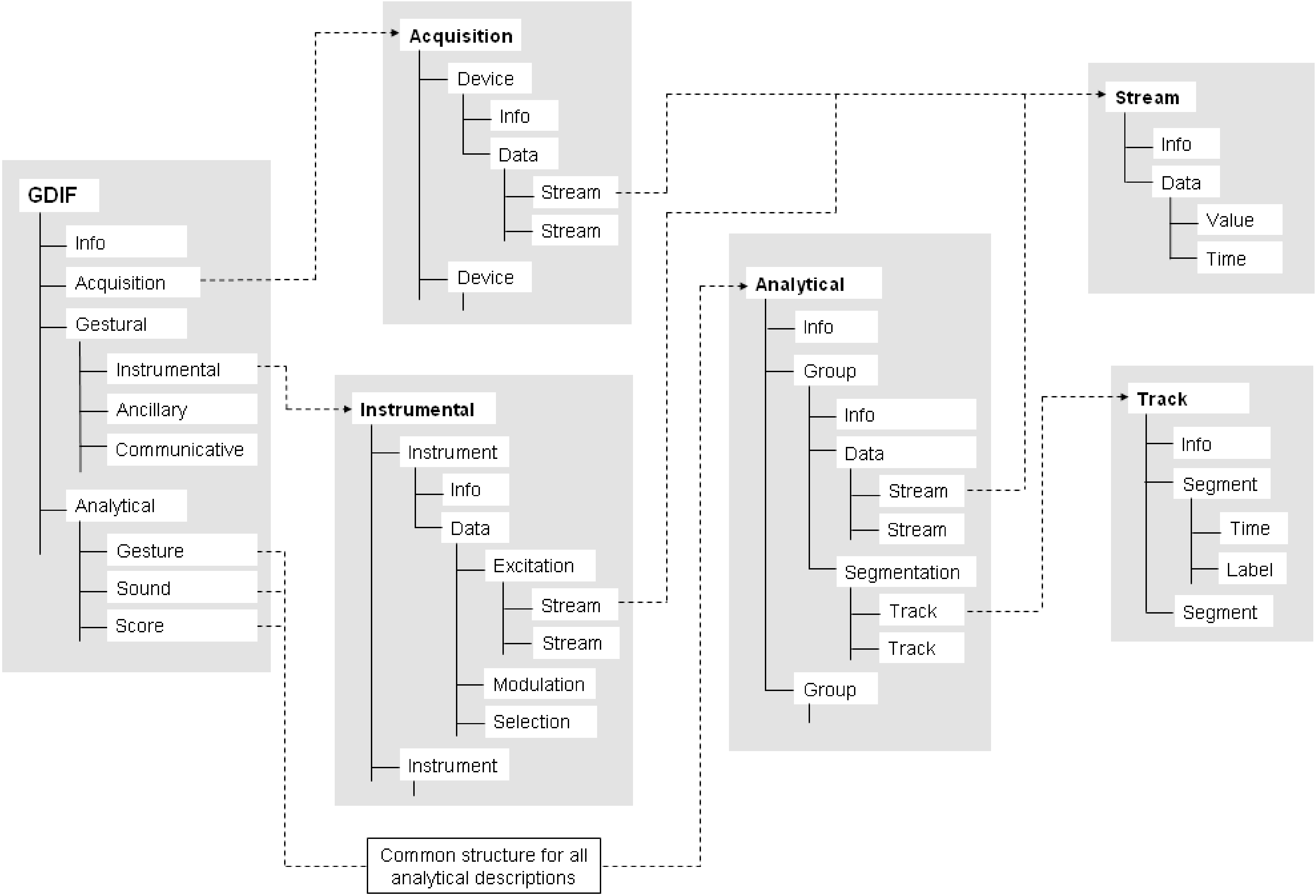
- Device information
- Stream significance ization
- Multiple da
- Multiple sa
- Multiple sc
- How much

- Raw data cooking instrumental gesture parameters obtention)
- Multiple segmentations
- Annotations
- Quantitative descriptions, models, representations (e.g. fonts)

- Retrieval
- Transformations
- Mapping

Application context

Violin Performance DB Creation



- GDIF
  - Info
  - Acquisition
  - Gestural
    - Instrumental
    - Ancillary
    - Communicative
  - Analytical
    - Gesture
    - Sound
    - Score

- Acquisition
  - Device
    - Info
    - Data
      - Stream
      - Stream
  - Device

- Instrumental
  - Instrument
    - Info
    - Data
      - Excitation
        - Stream
        - Stream
      - Modulation
      - Selection
  - Instrument

- Analytical
  - Info
  - Group
    - Info
    - Data
      - Stream
      - Stream
    - Segmentation
      - Track
      - Track
  - Group

- Stream
  - Info
  - Data
    - Value
    - Time

- Track
  - Info
  - Segment
    - Time
    - Label
  - Segment

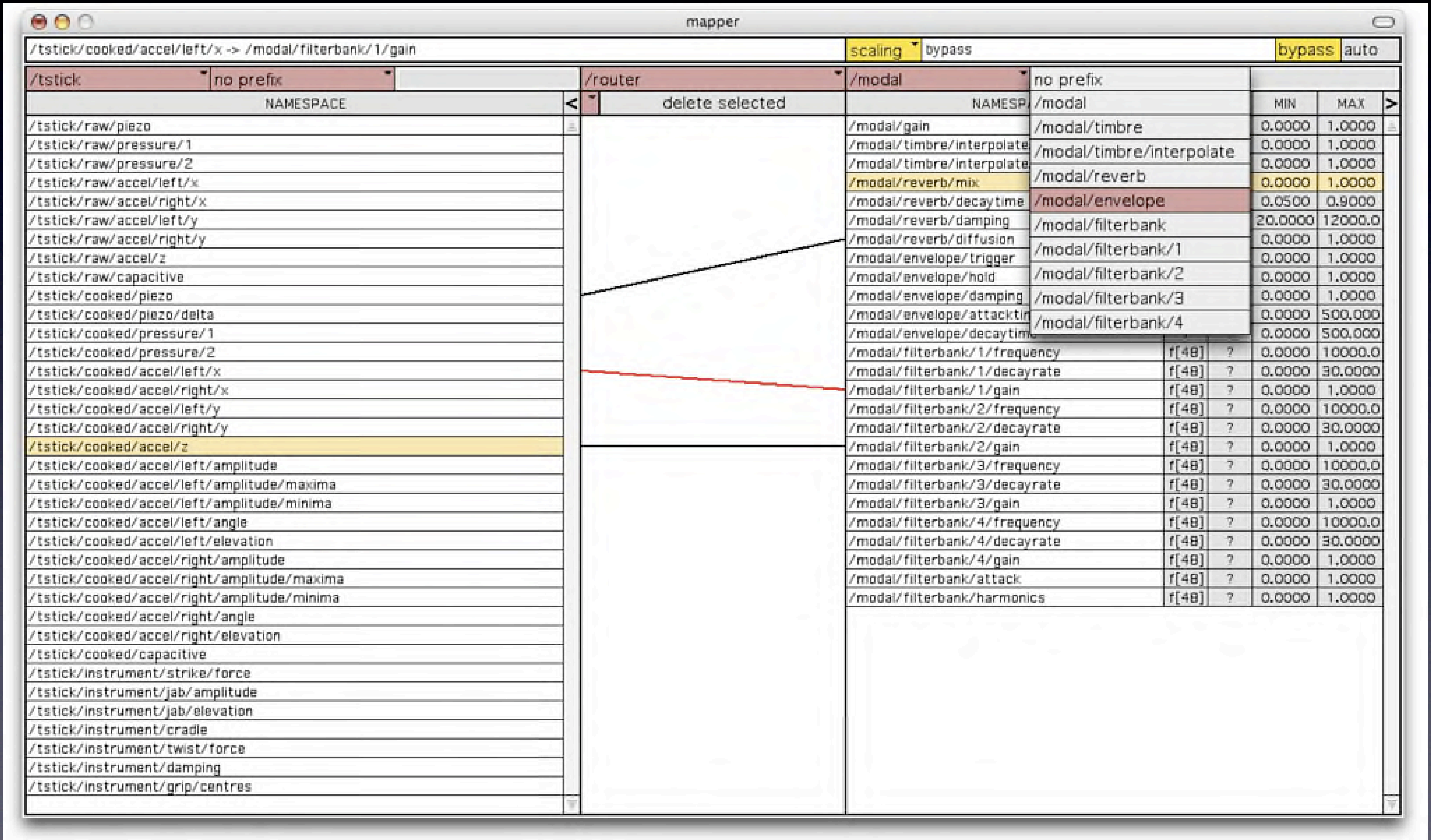
Common structure for all analytical descriptions

Joe Malloch

- Digital musical instruments
  - conception
  - design
  - construction
  - mapping
  - evaluation
  - performance
- Other gesture controlled systems

Tools for collaborative DMI mapping

Based on parts of the GDIF proposal



/tstick/cooked/accel/left/x -> /modal/filterbank/1/gain

NAMESPACE

/tstick/raw/piezo
/tstick/raw/pressure/1
/tstick/raw/pressure/2
/tstick/raw/accel/left/x
/tstick/raw/accel/right/x
/tstick/raw/accel/left/y
/tstick/raw/accel/right/y
/tstick/raw/accel/z
/tstick/raw/capacitive
/tstick/cooked/piezo
/tstick/cooked/piezo/delta
/tstick/cooked/pressure/1
/tstick/cooked/pressure/2
/tstick/cooked/accel/left/x
/tstick/cooked/accel/right/x
/tstick/cooked/accel/left/y
/tstick/cooked/accel/right/y
/tstick/cooked/accel/z
/tstick/cooked/accel/left/amplitude
/tstick/cooked/accel/left/amplitude/maxima
/tstick/cooked/accel/left/amplitude/minima
/tstick/cooked/accel/left/angle
/tstick/cooked/accel/left/elevation
/tstick/cooked/accel/right/amplitude
/tstick/cooked/accel/right/amplitude/maxima
/tstick/cooked/accel/right/amplitude/minima
/tstick/cooked/accel/right/angle
/tstick/cooked/accel/right/elevation
/tstick/cooked/capacitive
/tstick/instrument/strike/force
/tstick/instrument/jab/amplitude
/tstick/instrument/jab/elevation
/tstick/instrument/cradle
/tstick/instrument/twist/force
/tstick/instrument/damping
/tstick/instrument/grip/centres

NAMESP			MIN	MAX
/modal/gain			0.0000	1.0000
/modal/timbre/interpolate			0.0000	1.0000
/modal/timbre/interpolate			0.0000	1.0000
/modal/reverb/mix			0.0000	1.0000
/modal/reverb/decaytime			0.0500	0.9000
/modal/reverb/damping			20.0000	12000.0
/modal/reverb/diffusion			0.0000	1.0000
/modal/envelope/trigger			0.0000	1.0000
/modal/envelope/hold			0.0000	1.0000
/modal/envelope/damping			0.0000	1.0000
/modal/envelope/attacktin			0.0000	500.000
/modal/envelope/decaytim			0.0000	500.000
/modal/filterbank/1/frequency	f[48]	?	0.0000	10000.0
/modal/filterbank/1/decayrate	f[48]	?	0.0000	30.0000
/modal/filterbank/1/gain	f[48]	?	0.0000	1.0000
/modal/filterbank/2/frequency	f[48]	?	0.0000	10000.0
/modal/filterbank/2/decayrate	f[48]	?	0.0000	30.0000
/modal/filterbank/2/gain	f[48]	?	0.0000	1.0000
/modal/filterbank/3/frequency	f[48]	?	0.0000	10000.0
/modal/filterbank/3/decayrate	f[48]	?	0.0000	30.0000
/modal/filterbank/3/gain	f[48]	?	0.0000	1.0000
/modal/filterbank/4/frequency	f[48]	?	0.0000	10000.0
/modal/filterbank/4/decayrate	f[48]	?	0.0000	30.0000
/modal/filterbank/4/gain	f[48]	?	0.0000	1.0000
/modal/filterbank/attack	f[48]	?	0.0000	1.0000
/modal/filterbank/harmonics	f[48]	?	0.0000	1.0000

IDMIL • CIRMMT

- Motion-capture
  - Vicon System 460, Vicon MX, BTS Smart, NDI Optotrak, NDI Certus, Phoenix VisualEyez, Polhemus Liberty
  - Performance database
  - Mocap workshops
- Haptics
- Mapping
- Sonification
- Sensor development
- Many collaborations with other institutions, groups and projects

Example

Needs:

Streaming

Synchronization

Storage

Sharing

Need ability to record/store/analyse/share:

- multichannel audio
- multi-angle video
- multichannel sensor data
- commercial controller data
- motion-capture data
- force-feedback
- vibrotactile feedback

multiple sample rates & data types

multiple analyses

segmentation data

annotation

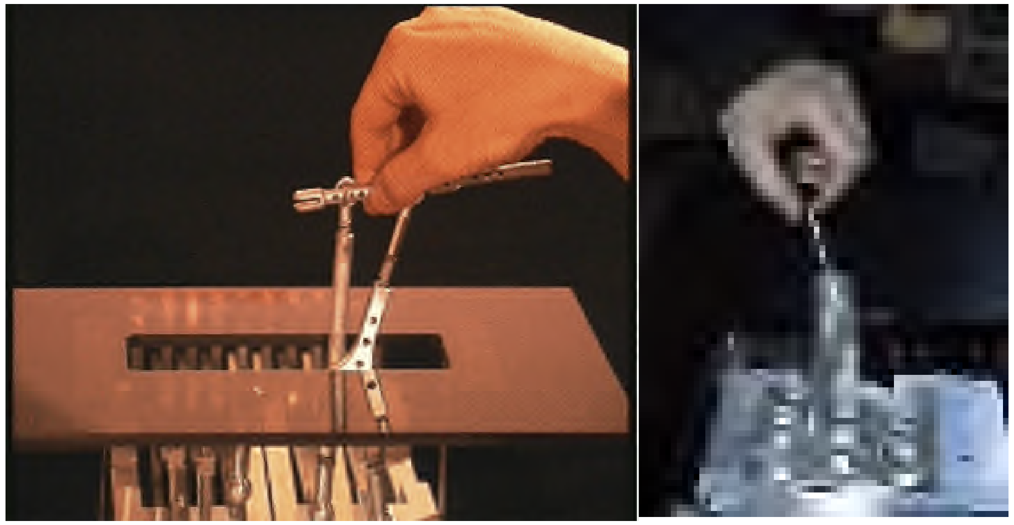
metadata

scores

Suggestions for development

- Allow streaming/storage of low, mid, and high-level information
  - include raw data
  - multiple perspectives
- Work together with other institutions
  - Share data and tools
- GDIF

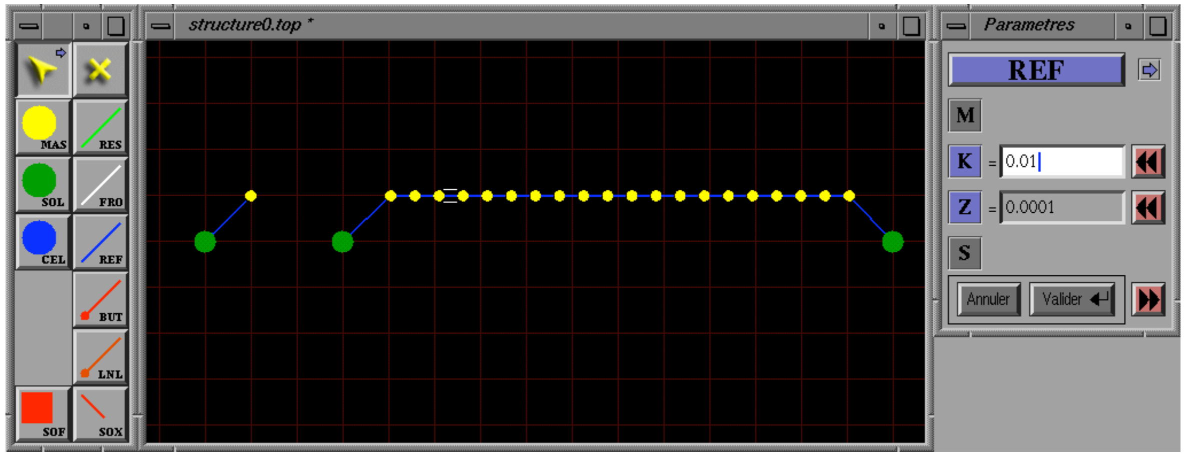
Nicolas CASTAGNE, **ACROE**, and ICA laboratory, INPG, France



Low Level Gesture

ERGOS force-feedback technologies

Low Level Gesture



K = 0.01

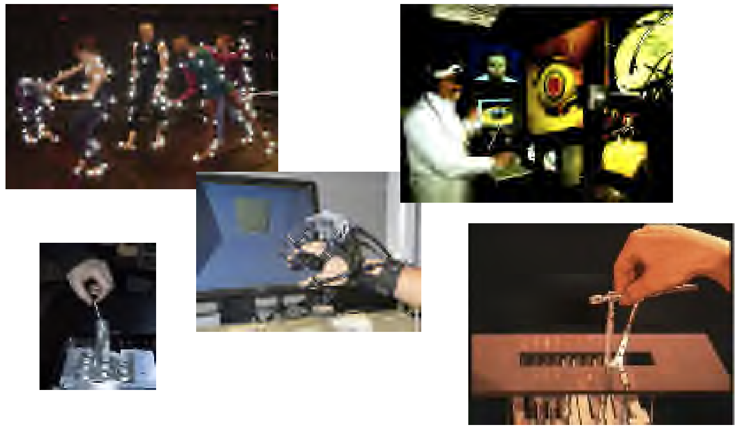
Z = 0.0001

Low Level Gesture

GENESIS

Music creation by physical modeling
Synthesis and use of movement/gesture data

OBSERVATIONS



1/ **Low-level gesture data & signals**,

that encode precisely a perfomed gesture,

are becoming a central mean.

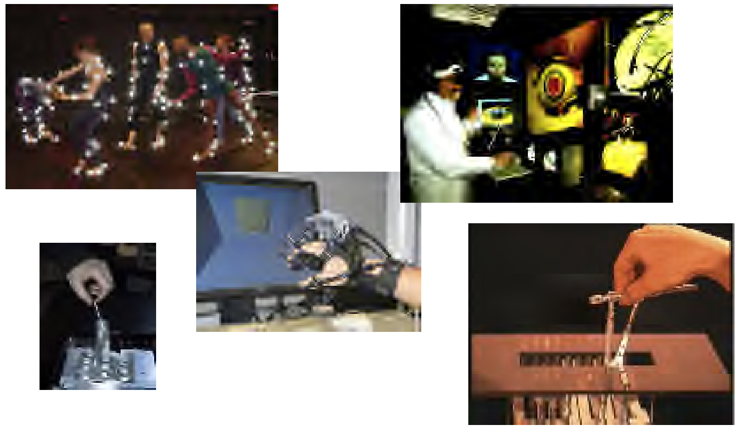
2/ **This category of data are mostly encoded without any format**, or with proprietary, device specific formats.

3/ **There is no appropriate format** to structure and encode low-level gesture data.

Existing formats are either:

Not Generic enough
Not Minimal enough
Not Efficient enough
Not Low Level enough

VIEW POINTS



The design of a **generic structuring and encoding of low-level gesture data is crucial today**
in order to allow structuring, storing, exchanging, analyzing, etc. gesture data.
(as well as PCM audio formats rooted the development of digital audio)

This question should be approached in a multidisciplinary context,
including device designers, Haptics, VR, HCI, Computer Graphics, Computer Music,
unless we will miss important things.

Ideally, it **should better be studied before proposing higher level formats** for the
encoding of higher level gesture, more symbolic, features,
unless we will "miss a step"

The question is very open. It is a **difficult research question**,
given the high versatily of gesture and gesture devices

=> WHAT are low level gesture data ?

=> Common work is needed

GMS format - Gesture and Motion Signal format

Annie Luciani, Mathieu Evrard, Damien Courousse, Nicolas Castagne , Claude Cadoz, Jean Loup Florens, 2006

GMS is first proposal for a:

- as generic as possible so far
- low-level,
- binary
- minimal

format for

organizing and storing

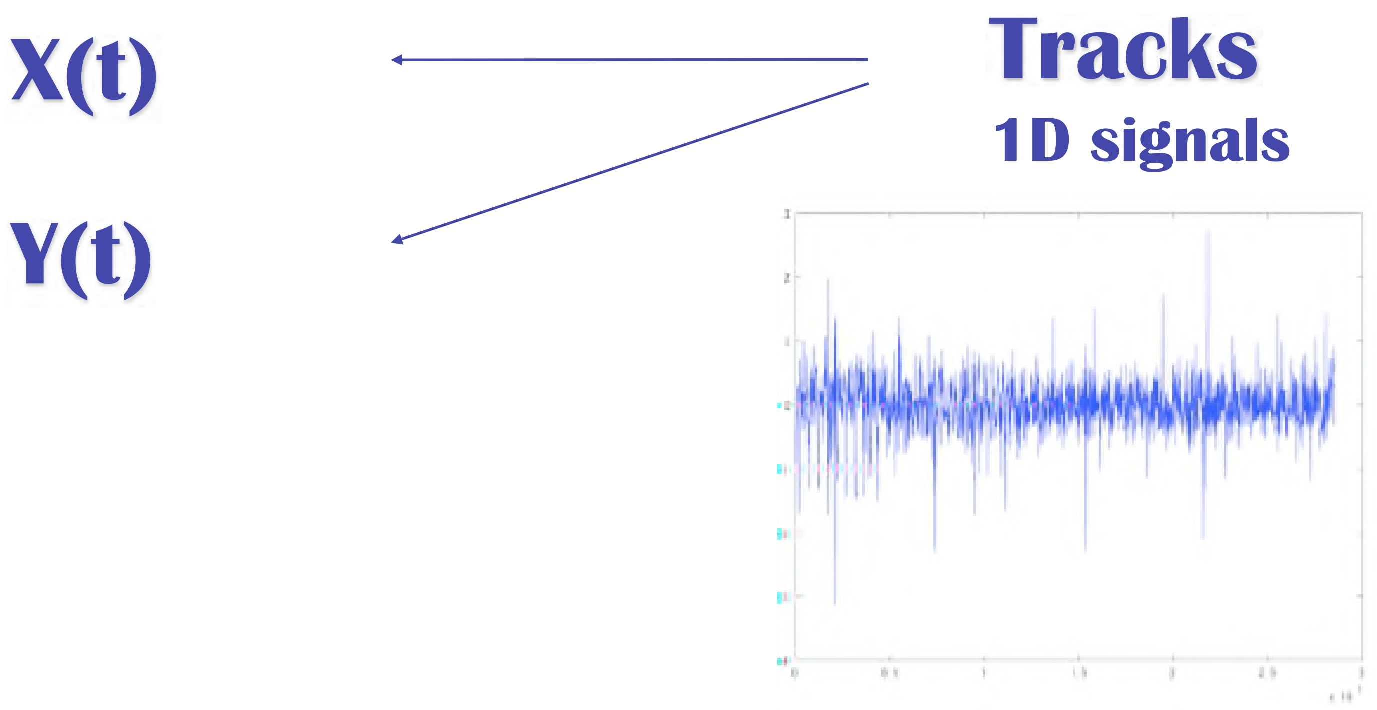
low level Gesture Signals and Gesture streams

GMS format - Gesture and Motion Signal format

GMS organizes the Morphological Versatility of gesture signals

GMS format - Gesture and Motion Signal format

GMS organizes the Morphological Versatility of gesture signals



X(t)

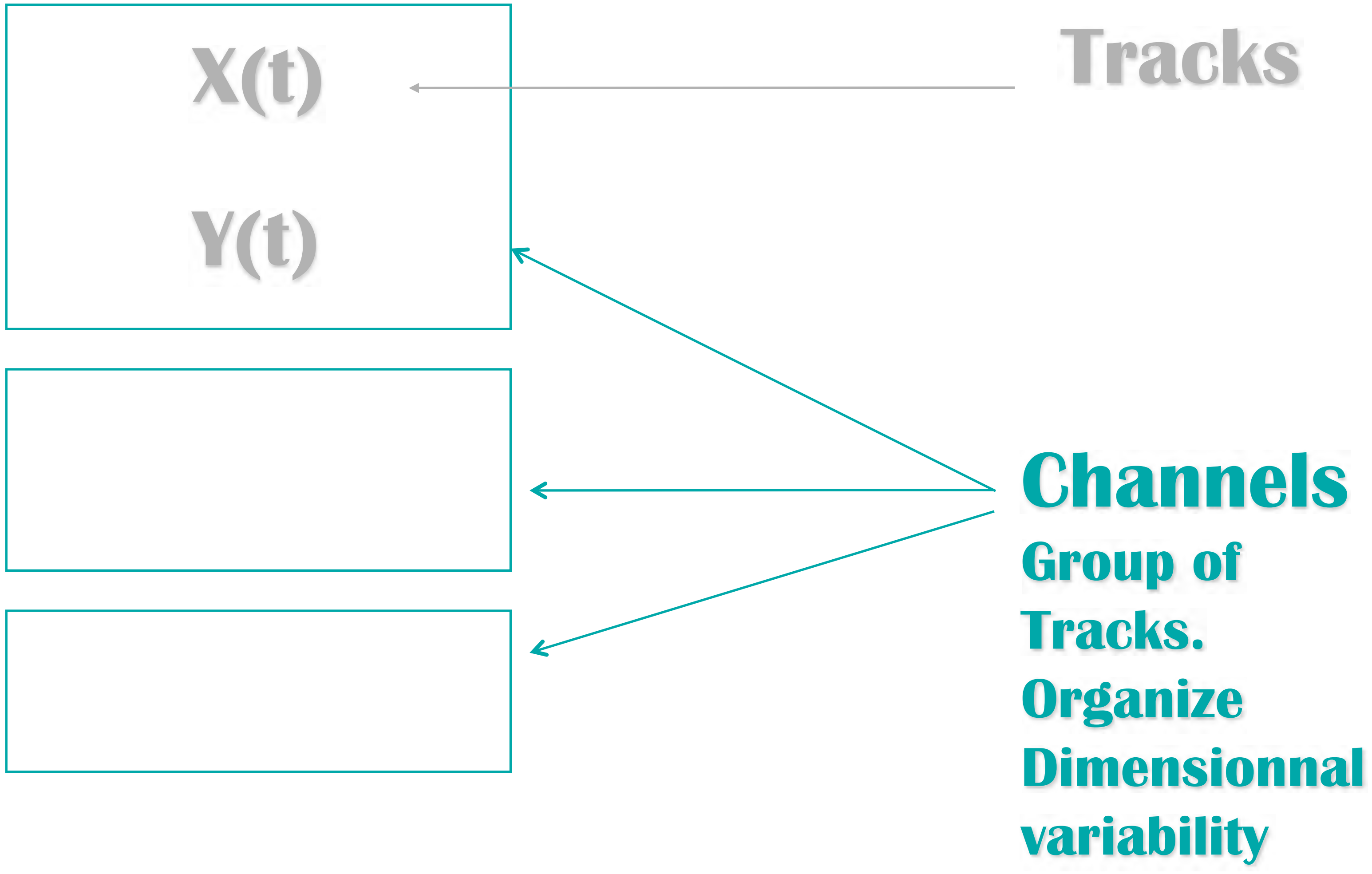
Y(t)

Tracks

1D signals

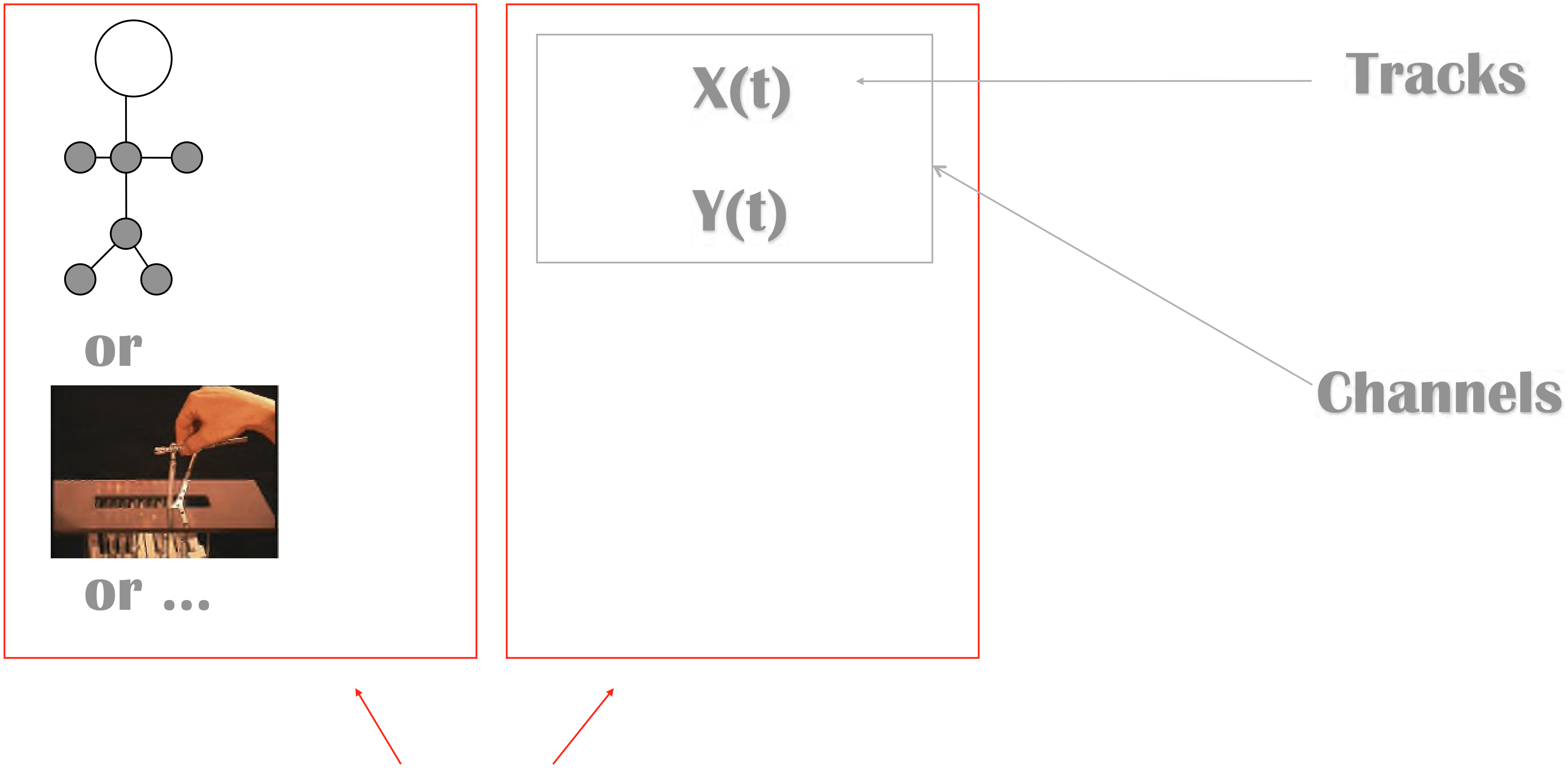
GMS format - Gesture and Motion Signal format

GMS organizes the Morphological Versatility of gesture signals



GMS format - Gesture and Motion Signal format

GMS organizes the Morphological Versatility of gesture signals



or

or ...

X(t)

Y(t)

Tracks

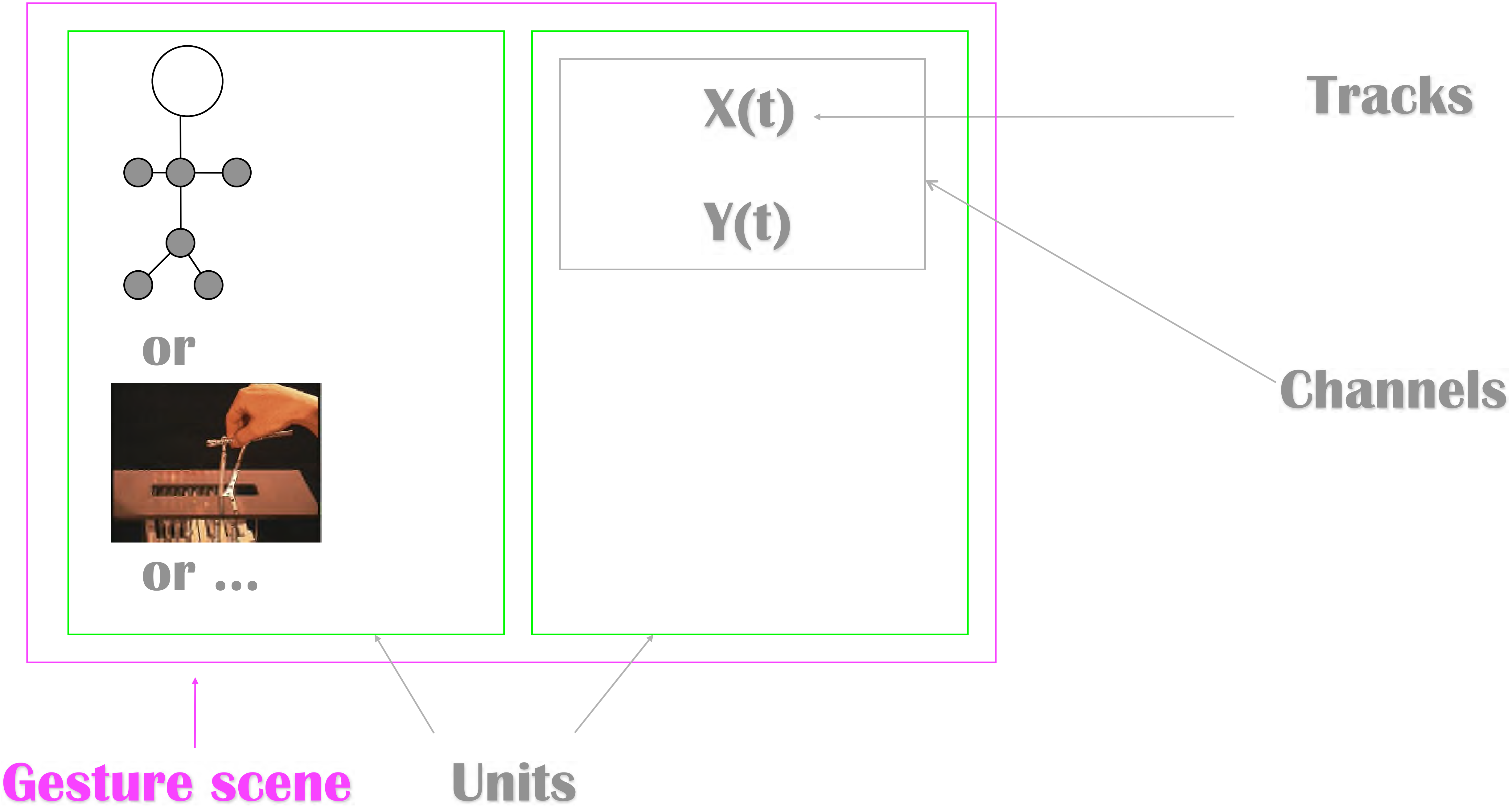
Channels

Units

group of channels. Organize Structural Versatility

GMS format - Gesture and Motion Signal format

GMS organizes the Morphological Versatility of gesture signals



X(t)

Y(t)

or

or ...

Tracks

Channels

Gesture scene

Units

GMS format - Gesture and Motion Signal format

- Gesture **Track**: a basic digital signal
  Meaningless in itself as for the performed gesture

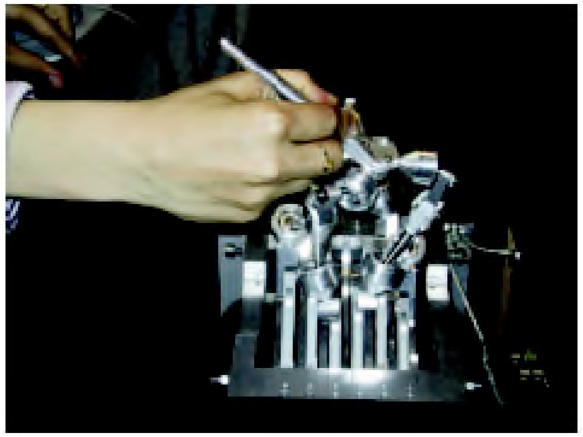
- Gesture **Channel**: made of various tracks
  A channel is A-Dimensionnal, 1D, 2D, or 3D
  A channel is of a certain type: position, force…
  Ex: a channel for a moving point, a unique force…

- Gesture **Unit**: a group of channel
  Units support the functionnal organisation of the perfomed gesture
  Ex: a single character in motion capture, a piano keyboard…

- Gesture **Scene**: various units
  The scene defines the framerate & the duration of the signal

GMS format - Gesture and Motion Signal format

Example



A Scene made of 3 **Units**

- **Unit 1**: "mocap"
  N 3D Position **channel**

- **Unit 2**: "Force Feedback »
  1 3D Position **channel**
  1 3D Force **channel**

- **Unit 3**: "keyboard"
  64 A-Dimensional **channels**

Matt Wright (CNMAT and CCRMA, soon UVic)

matt@cnmat.berkeley.edu

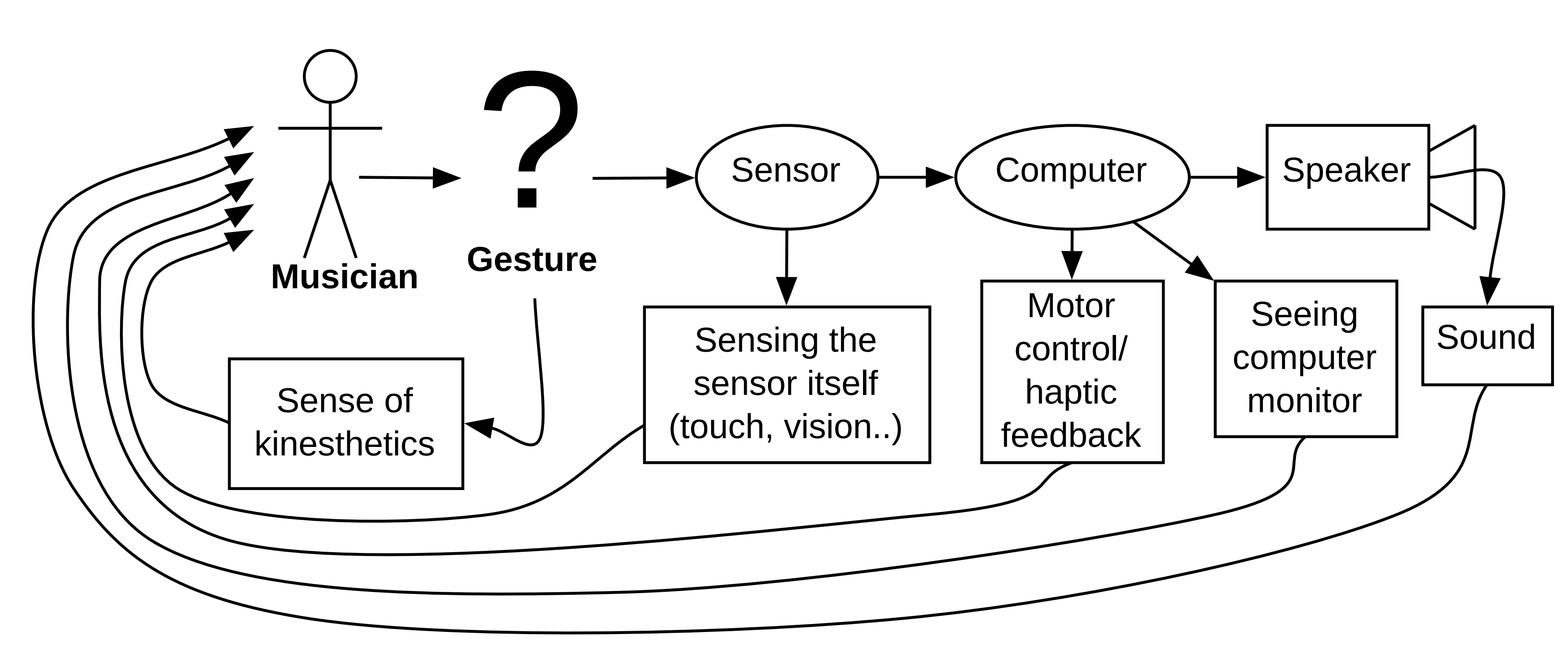
I mainly use movement-related data for realtime mapping of my gestures to sound control.

I also did one motion capture project:

[redacted]

The Role of Feedback in Producing Gesture

Musicians' physical motion is practically meaningless without knowing the context. Musicians constantly adapt their gestures based on auditory, haptic, and visual feedback:



This example is from the NIME context (w/ computer-controlled haptic feedback). *Any* situation where a musician makes gesture will provide lots of feedback, which will in turn influence the gestures produced.

Some Advice on Promoting New Standards in the Computer Music Community

- Start by implementing something that solves your own problems, then generalize.
- Supply free code in the form of full-fledged working examples designed to be copied and modified.

    Don’t just provide a library. (E.g., the OSC-Kit)

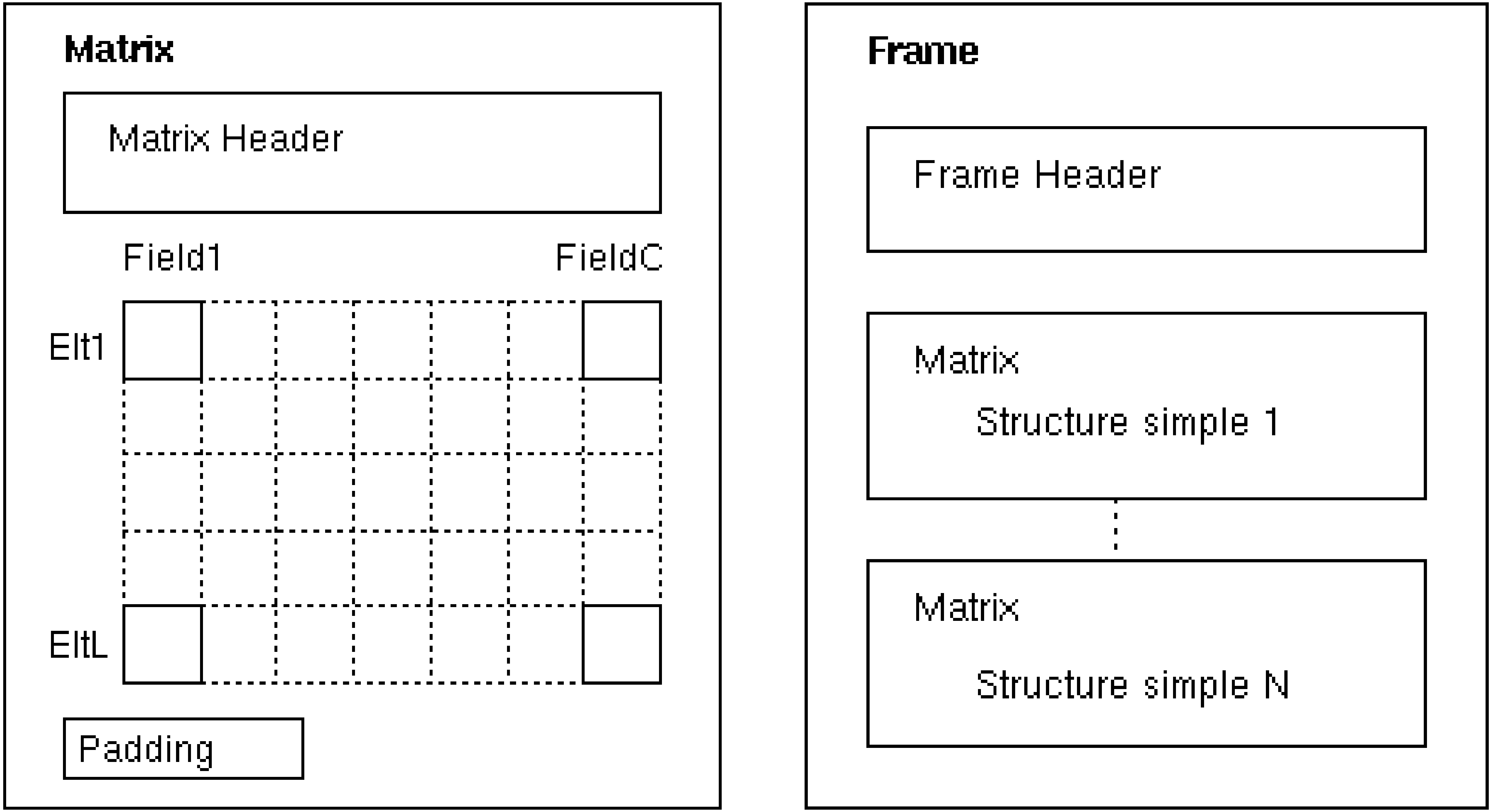
    Developers are easily put off by seeming complexity; they often prefer to write a limited and possibly incorrect implementation from scratch instead of using open-source resources.

- In our community, standards-making and support of interchange seems to be an ongoing iterative activity.

SDIF Sound Description Interchange Format **in 2 minutes**

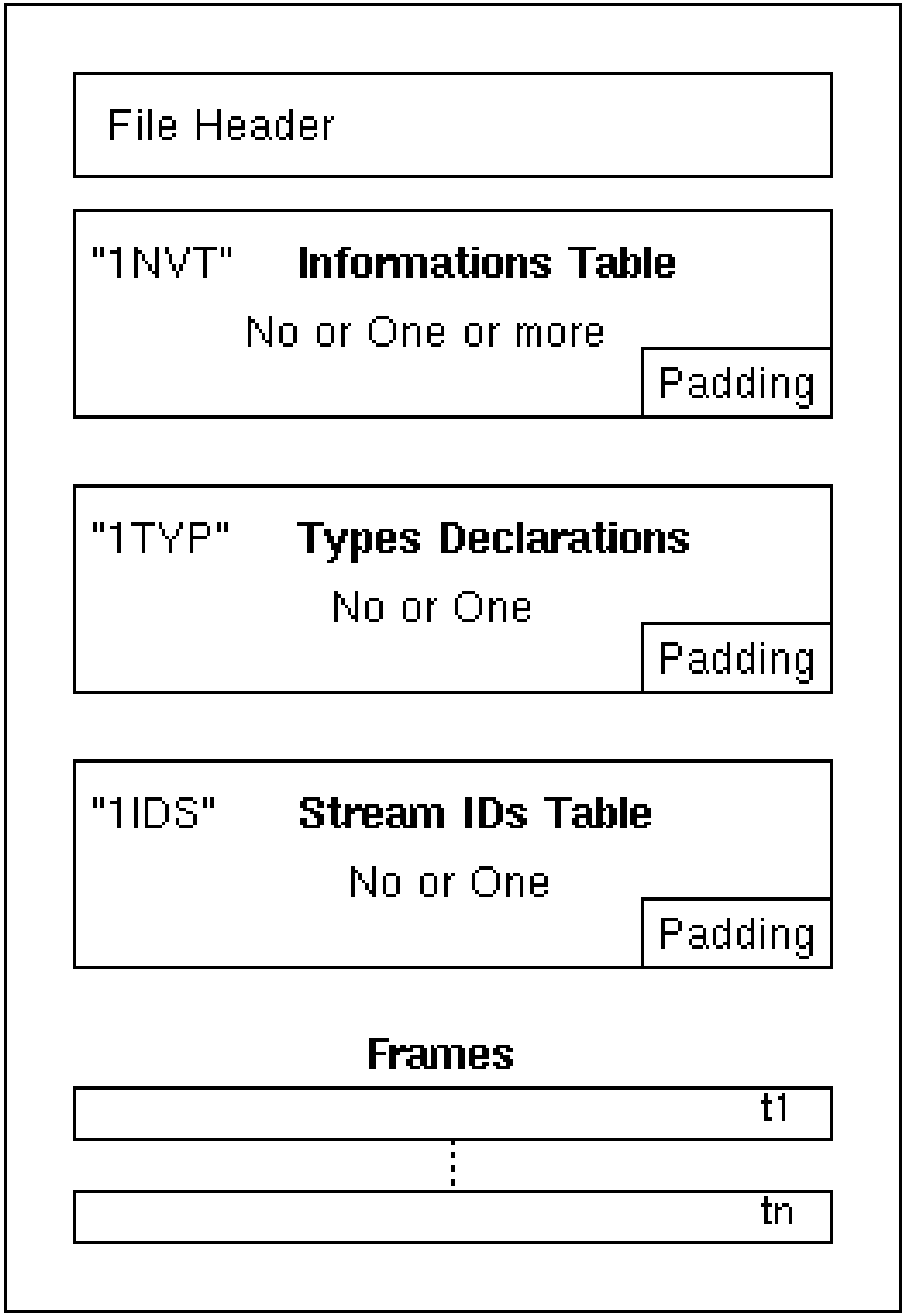
- established standard for the *well-defined and extensible interchange* of a variety of sound descriptions
  - e.g. spectral, sinusoidal, time-domain, descriptors, markers
  - created at the end of the 1990s in collaboration by Ircam, CNMAT, and IUA-UPF
- *Metaformat*: basic data format framework + an extensible set of standard sound descriptions
- Open Source implementations at http://sdif.sourceforge.net (LGPL) and CNMAT http://www.cnmat.berkeley.edu/SDIF

Structure of an SDIF File (bottom-up)



Matrix element types from 1 to 8 bytes; text, integer or floating point

- NVT = Name–Value lookup-table for any context information (date, user, source sound file name, etc.)
- TYP = Type declarations for privately defined types or extended standard types (frame signature, matrix and column names): *obligatory definition, well-defined semantics*



Applications supporting SDIF

- Sound/Music apps
  - Max/MSP (via FTM data structures and CNMAT externals)
  - Analysis/Synthesis software: AudioSculpt, Loris, Spear
  - OpenMusic
- Programming languages
  - C/C++ (SDIF and EaSDIF libraries from Ircam, sdif-lib from CNMAT)
  - Matlab
  - Java, Python, Perl, Ruby, TCL, PHP, SmallTalk... (via SWIG)
- Tools
  - command-line extractors and converters
  - editors, visualisers

Ircam – Centre Pompidou

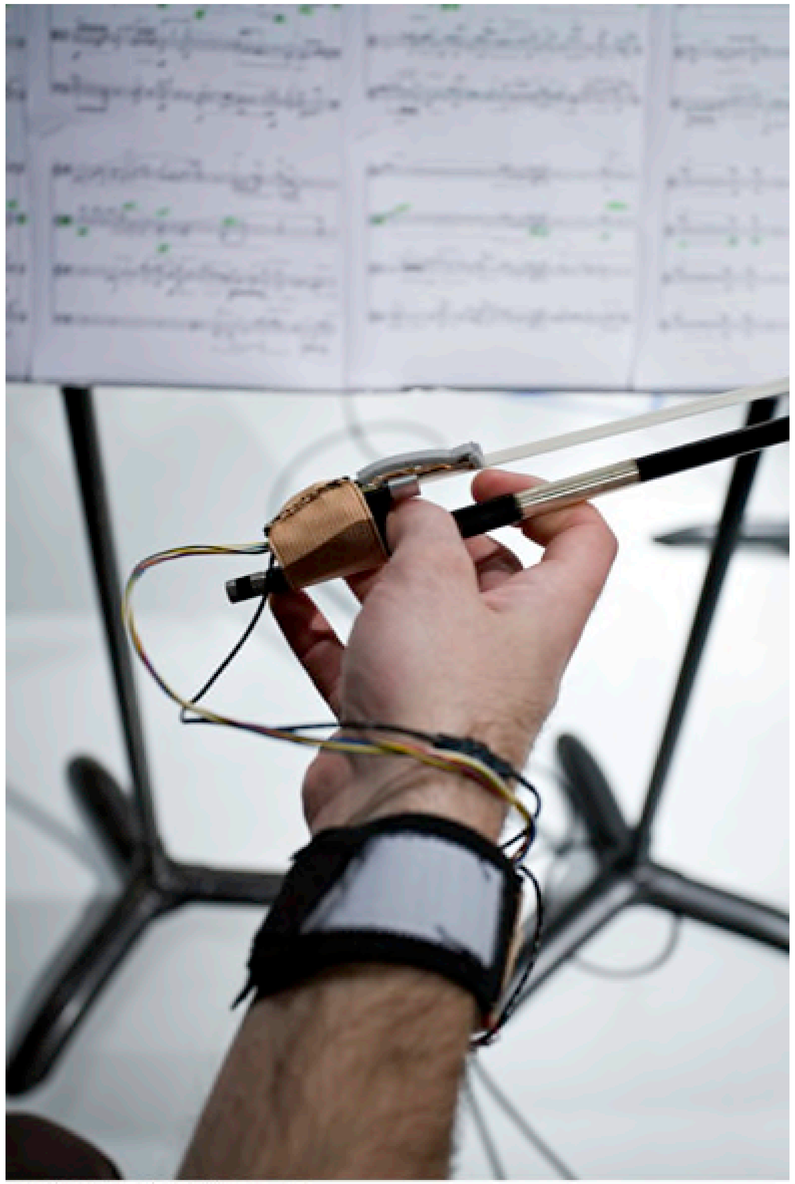
Real-Time Music Interaction Team
***Interaction Musicale Temps-Réel* (IMTR)**

Norbert Schnell, Fréderic Bevilacqua,
Diemo Schwarz, Riccardo Borghesi,
Nicolas Leroy, Nicolas Rasamimanana,
Arshia Cont, Julien Bloit, Jean-Philippe Lambert

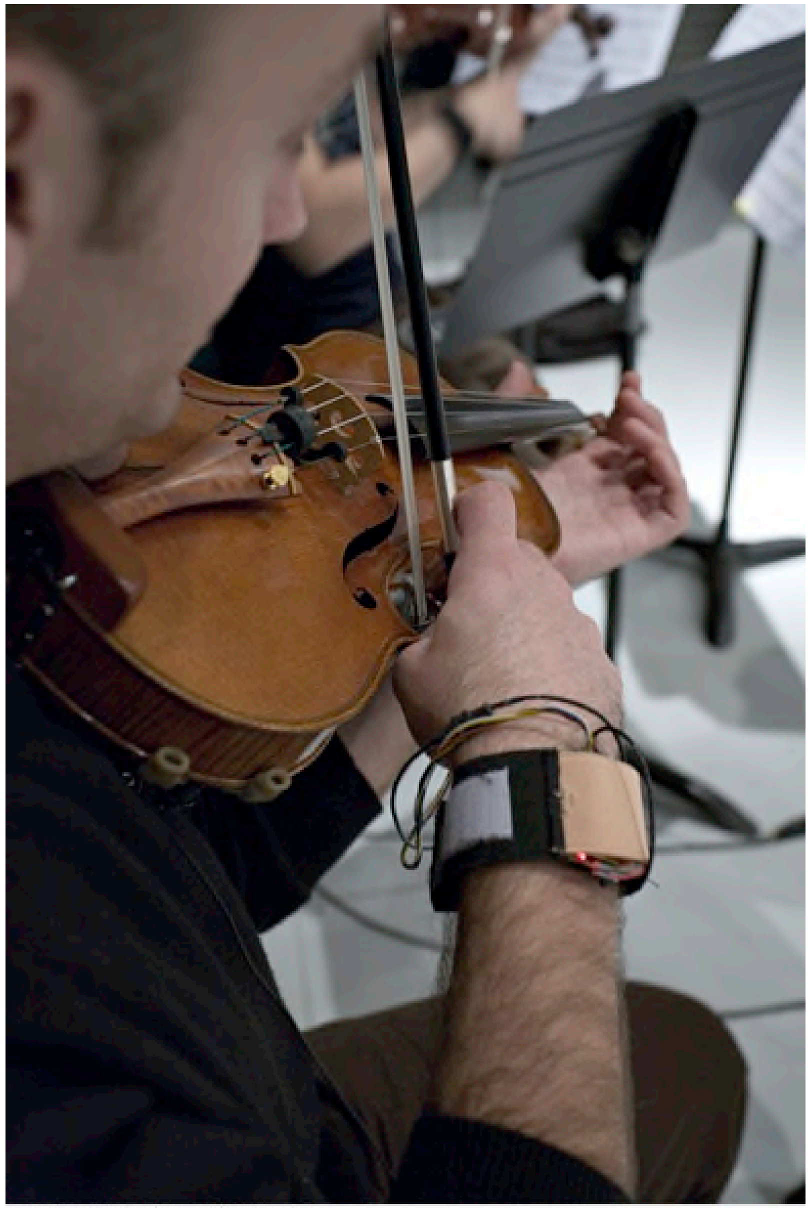
Our Applications (1)

- augmented instruments, especially string instruments for:
  - pedagogy (I-Maestro project *http://www.i-maestro.org*)
  - augmented string quartet (Florence Baschet)
- 3D motion capture: study of bowing
- alternative interfaces (using for example the gesture follower and the wii-mote or other wireless interfaces)

Augmenting instruments



mention: Kleinefenn@ifrance.com



mention: Kleinefenn@ifrance.com

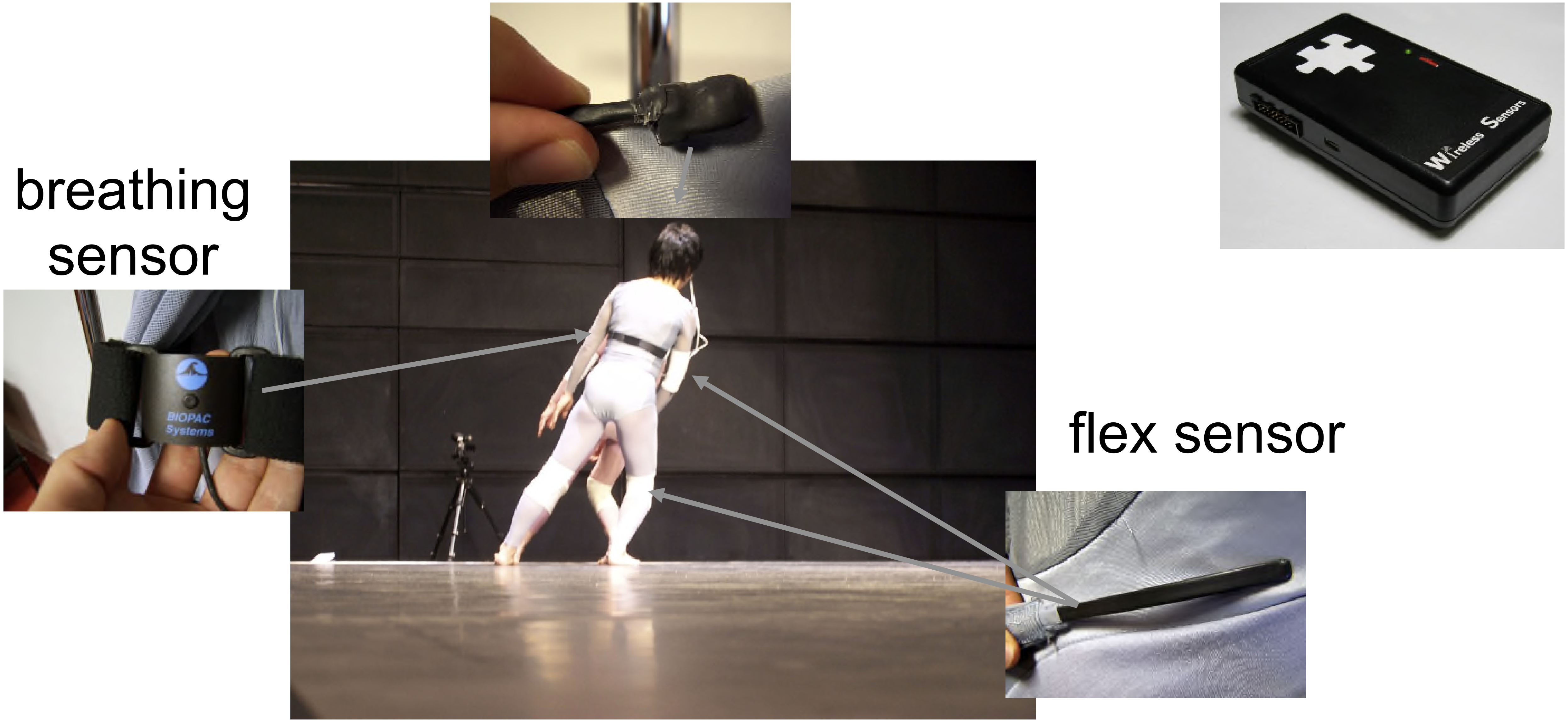
Our Applications (1)

- augmented instruments, especially string instruments for:
  - pedagogy (I-Maestro project *http://www.i-maestro.org*)
  - augmented string quartet (Florence Baschet)
- 3D motion capture: study of bowing
- alternative interfaces (using for example the gesture follower and the wii-mote or other wireless interfaces)

Our Applications (2)

- projects related to dance, using wireless interfaces with accelerometers and/or video capture (Eyesweb, Jitter)
  - for example: documentation/notation project with the Emio Greco company
- virtual reality related projects
  - EarToy, ANR RIAM project on auditory/body/space interaction in immersive audio systems

Sensor technology for Dance



Wireless Interface
(WiseBox)

accelerometer

breathing
sensor

flex sensor

Our Applications (2)

- projects related to dance, using wireless interfaces with accelerometers and/or video capture (Eyesweb, Jitter)
  - for example: documentation/notation project with the Emio Greco company
- virtual reality related projects
  - EarToy, ANR RIAM project on auditory/body/space interaction in immersive audio systems

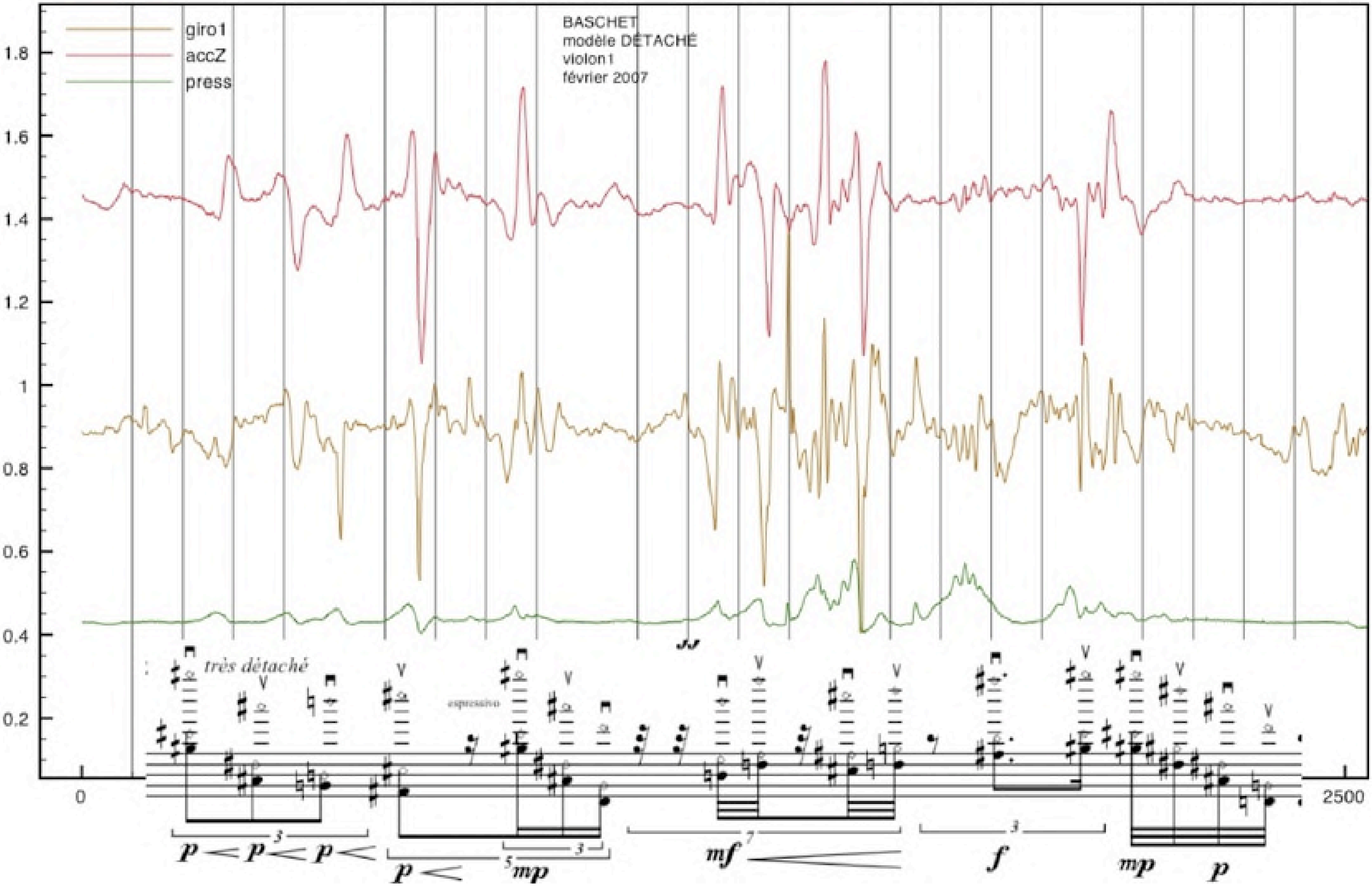
How do you currently work with music-related movement and gesture data?

- record gesture data synchronously with sound, replay for study
- "gesture data" sometimes based on sound properties!
  - e.g. loudness envelope, spectral centroid, etc.

How do you currently work with music-related movement and gesture data?

- Once we have that, we can and have to:
  - create relationships between different sets of gesture data corresponding to different gestures or different performances of the same gesture
  - store and manage libraries of gestures
  - study the relationship between sensor and motion capture data and audio features
  - align and recognise different gesture data sets (or simply "gestures")
  - align gesture data to audio data and to symbolic music representation

Alignment Score/Gesture Data



What are your needs of formats and standards?

- export easily the gesture to various applications (e.g. Matlab).
- the gesture can be of many formats, we need a header describing what the data is (accelerometer, from video, etc)
- interoperability between different applications (on different platforms)
- unified storage of sensor, motion capture, audio, and sound descriptor data
- variable rate data, high precision (time and data)

What are your suggestions for future development?

- use of a format enabling *multimodality*: audio, sound descriptions, and gesture
- support of segmentation and annotation
- support of the relationship between gesture data and symbolic data
- **use of SDIF**
  - *[the **S** could also stand for **Signal**]*
  - needs definition of a set of (non-exclusive) standard types
- implementation suggestions:
  - platform independent visualization components

Panel on Gesture Standards: Antonio Camurri

- EyesWeb XMI:
  - Stable & robust version, publicly released before 15 Sept 07;
  - New Tools: EyesWeb-Mobius in collaboration between Antonio Camurri (UGDist) and Ben Knapp (QUB and TRIL Centre) to design high level GUIs controlling distributed EyesWeb patches; running also on mobiles and palmtops.
  - MOBIUS Blocks and Bio-Tools for standards-based physiological data processing
- We intend to collaborate with the research community to improve communication of EyesWeb XMI with other tools, and to support research projects: include new datatypes, import contributes in order to adhere to emerging standards in the research community. How?
  - **EyesWeb Week, February 2008, Casa Paganini, Genoa**: in collaboration with at UGDist. A session will be dedicated to emerging gesture standards for music research.
  - **Proposal of a continuation of this ICMC panel to NIME 2008** (hosted by InfoMus Lab – Casa Paganini)

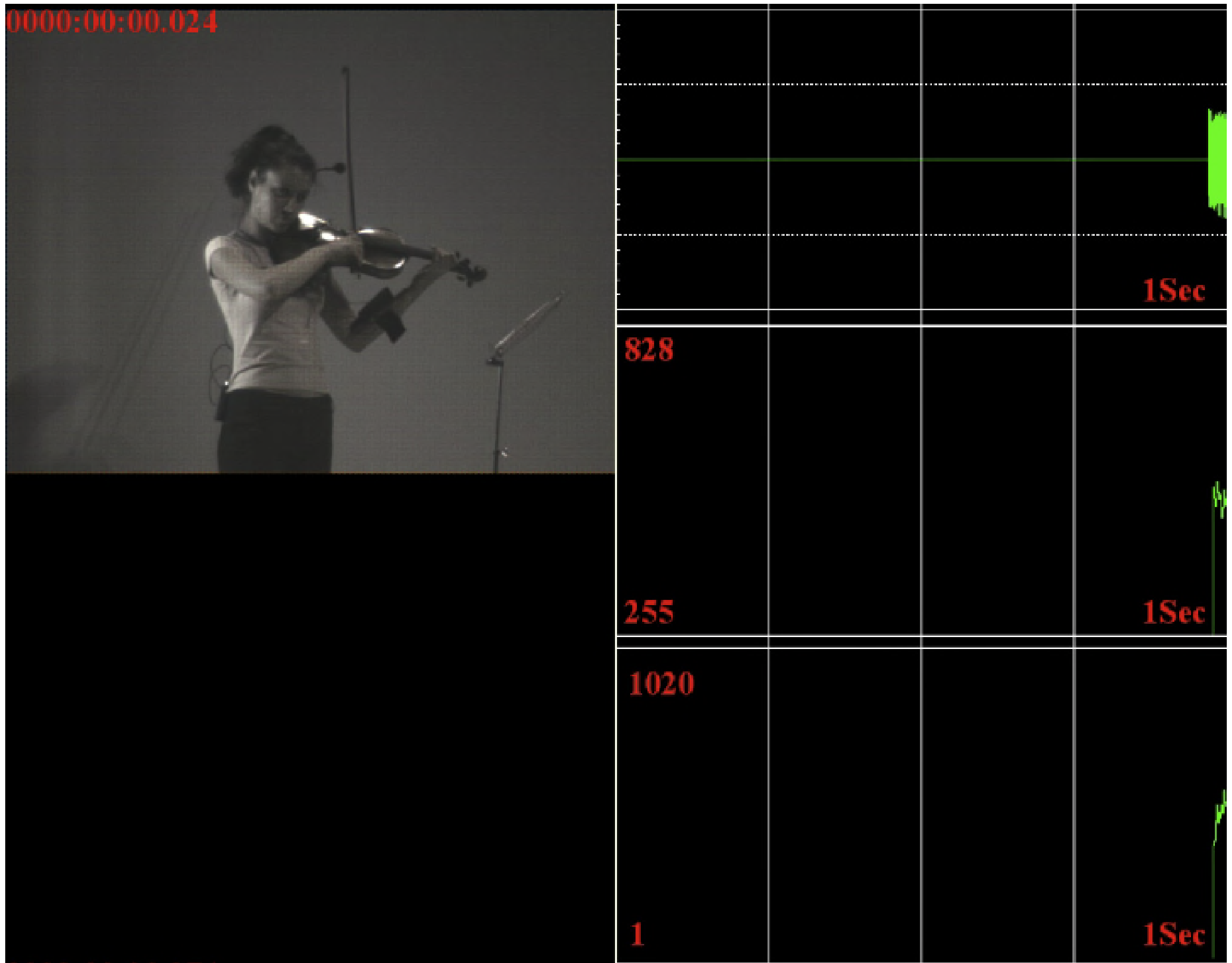
Issues on gesture data representation in EyesWeb XMI

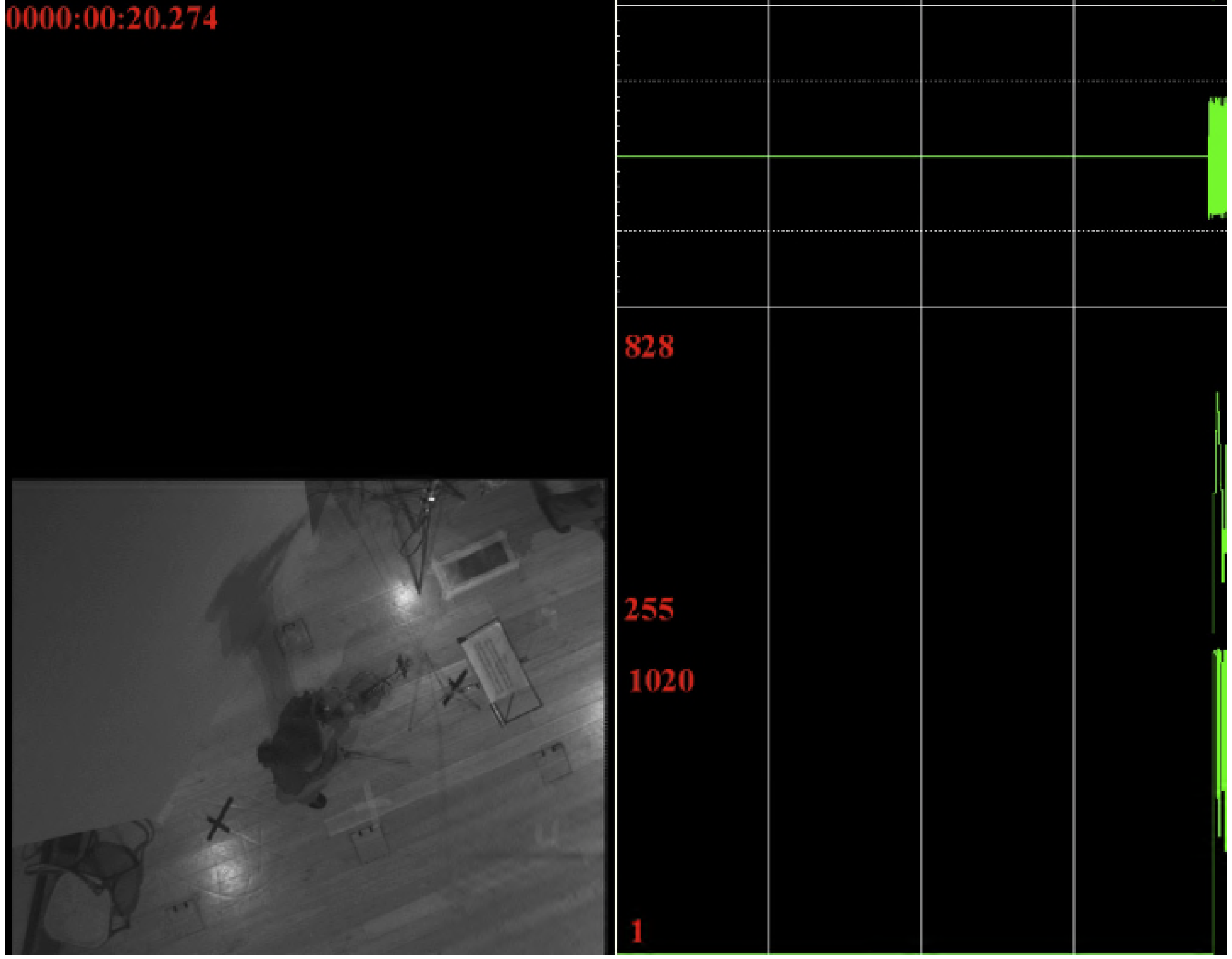
- What you want to do with tracked movement data? flexibility
  - Real-time processing: at each instant a motion tracking block generates corresponding values (e.g. Position of a joint, value of a sensor; “snapshot” of the human skeleton with time stamp)
  - Data analysis: generate a trajectory f(t) for each joint; buffered; separate streams, one for each extracted cue, time stamps.
  - Two approaches, two different representations of data: support both views: sw modules to translate from one to the other
- General guidelines: generality
  - A skeleton model (eg H-ANIM) is not always needed: different approaches should be supported (eg approaches not based on joints tracking, expressive cues).
  - Gesture cues as *low varying signals* (wrt sampling rates of sensors, audio etc):
  - Support to multimodality and multisensory processing: gesture in a wider perspective
  - Expressive gesture data – need for a layered standard (different level of abstraction, e.g. (www.megaproject.org)

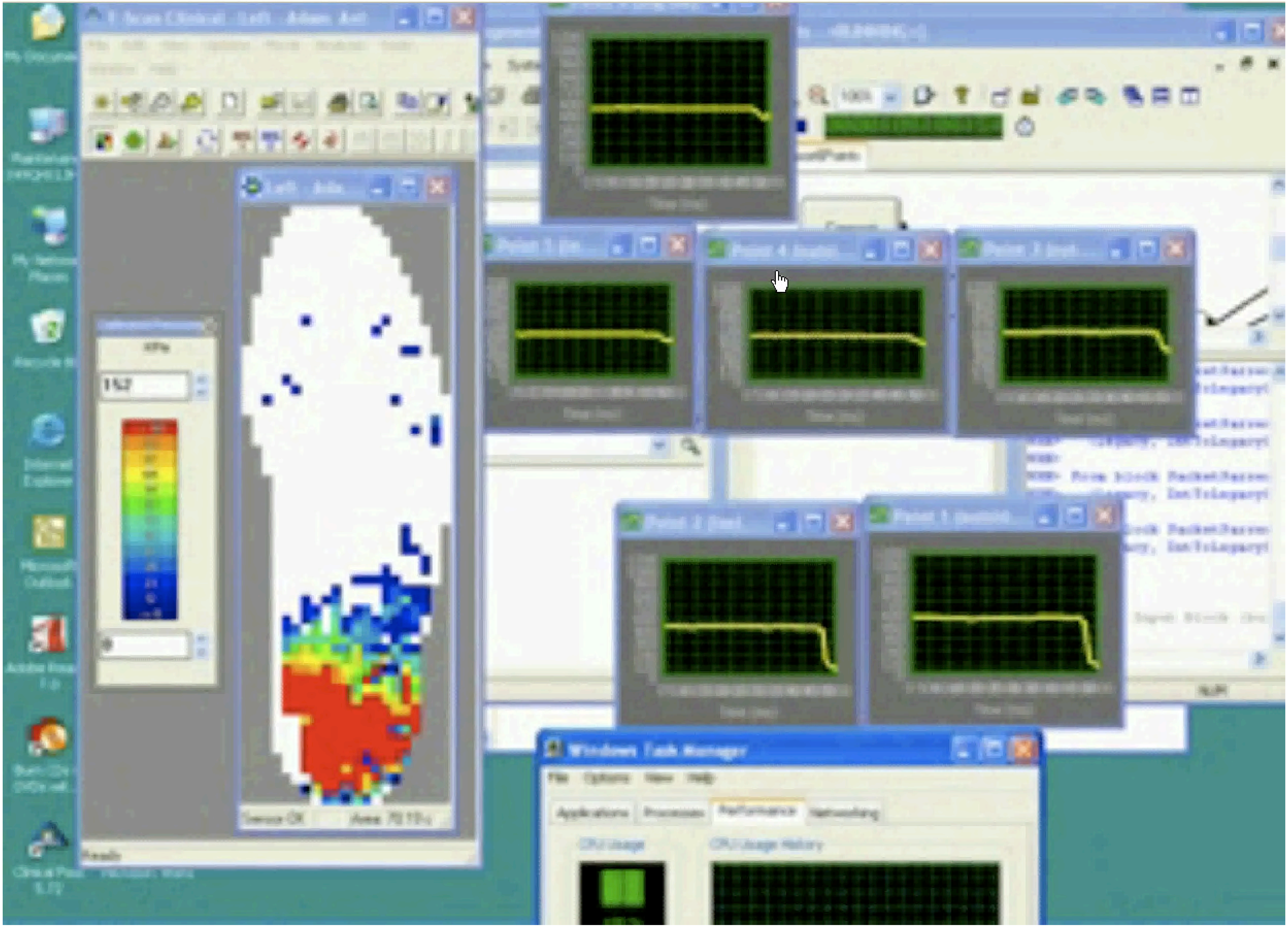
EyesWeb - Mobius

An example of an EyesWeb-Mobius application, developed by UGDIST and TRIL Centre: a palmtop runs a high-level control interface of two EyesWeb XMI applications running in the two laptops in the background. The palmtop shows a video window showing the result of image analysis (executed in the laptop on the right), slider and buttons to control both remote applications, and widget to show the state of the EyesWeb applications.

EyesWeb-Mobius include a development environment to generate user interface layouts to be run on desktop as well as in palmtops and mobile phones.


0000:00:00.024
1Sec
828
255
1Sec
1020
1
1Sec


0000:00:20.274
828
255
1020
1


Left - Jolo
KPa
152
Sensor OK
Area: 70.1
Ready
Point 1
Point 4
Point 3
Point 2
Point 1
Time (sec)
Windows Task Manager
File Options View Help
Applications Processes Performance Networking
CPU Usage
CPU Usage History

ICMC 2007

Stuart Pullinger
(on behalf of Douglas McGilvray),
Centre for Music Technology,
University of Glasgow

Ashitaka

- An 'audiovisual' instrument
- Using motion to connect audio & visuals: 'synchresis'
- Motion & gestures mapped to audio and visual transformations

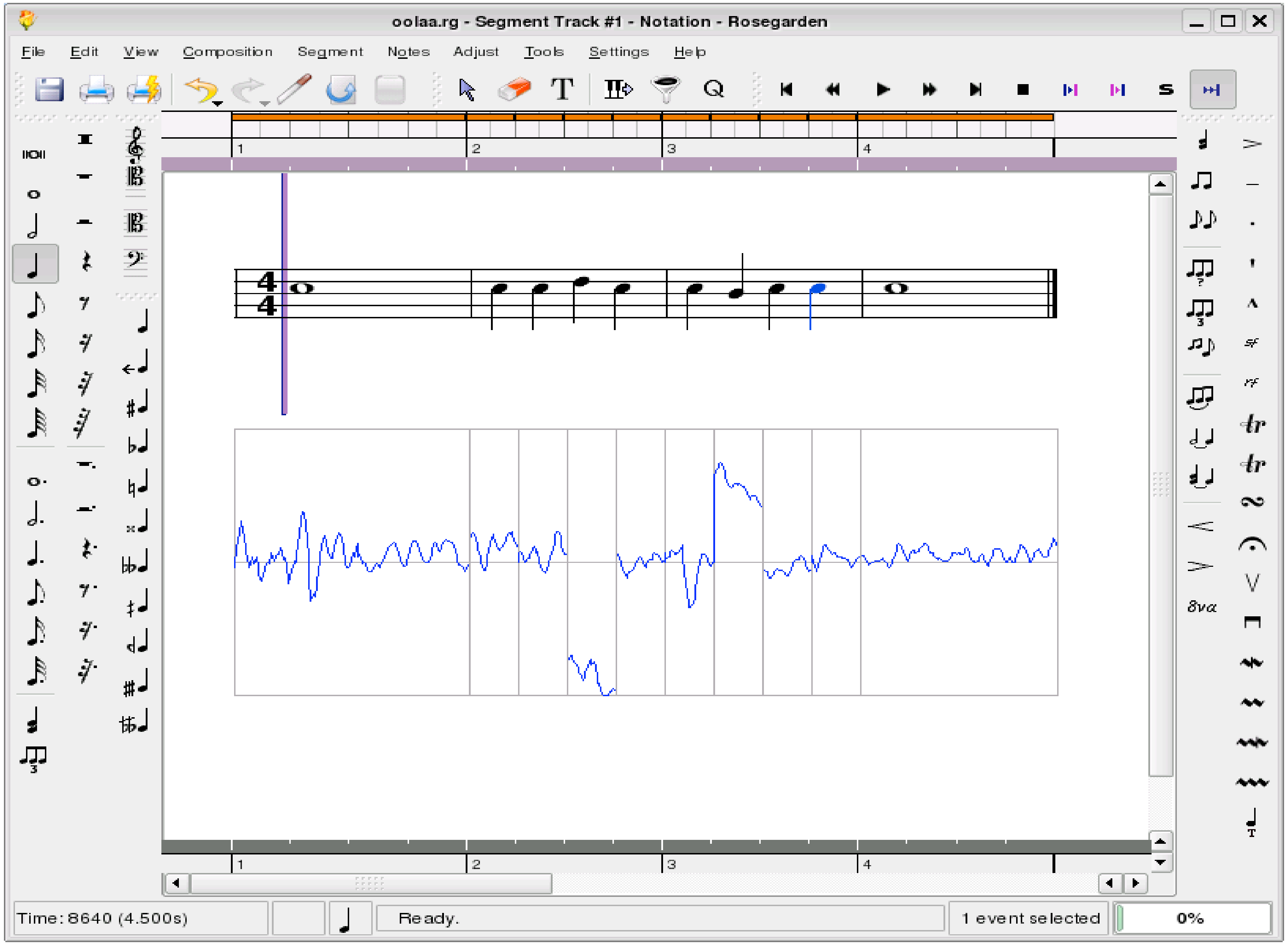
Multi-modal analysis of piano performance

- Video tracking of finger position and shape
- MIDI gestural information captured using a Moog piano bar
- Score, Video, Audio, Gestural, Analysis

PML

- Performance Markup Language (PML)
- XML based specification for the representation for the analysis of performance issues
- Score, performance & analysis information in separate, overlapping hierarchies.
- “Building the camera while shooting the film”

Microtonal Pitchtracker



oolaa.rg - Segment Track #1 - Notation - Rosegarden

Time: 8640 (4.500s)

Ready.

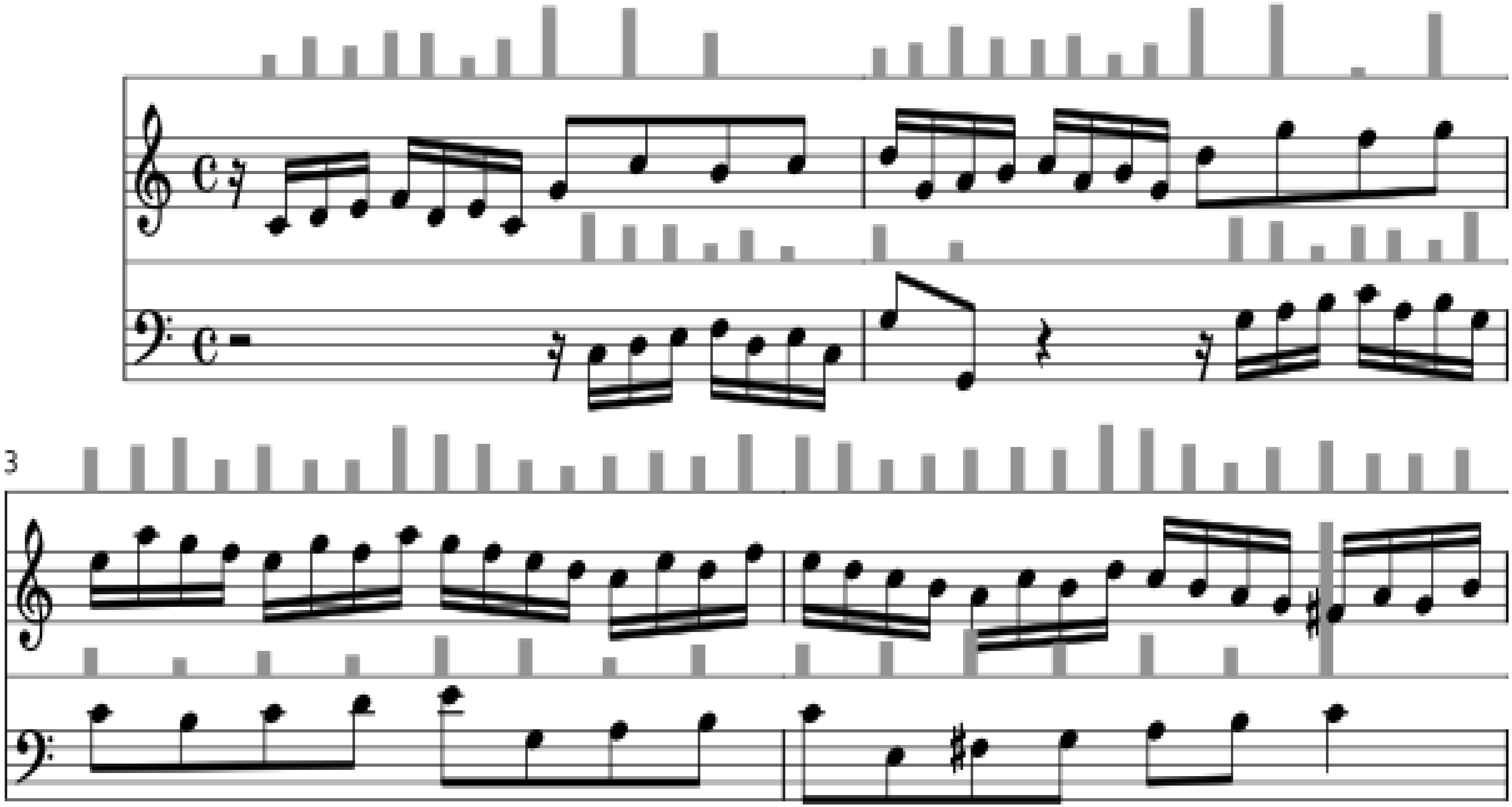
1 event selected

0%

Music and Performance Database

- Store score, audio, video and performance data.
- Add functions to DB to ease analysis – create a data model.
- Create presentation software to display results in the context of the score.

Like this...



Our Needs

- Open and Free (as in beer and speech)
- Must come with tools and programming interfaces
- Must be widely supported/compatible with existing systems
- Need to stay focused on analysis and not try to describe all aspects of music

Future Developments

- Engineers, composers, musicians and performers must collaborate to create a taxonomy of gesture.
- Could be formalised in RDF/OWL
- ... and incorporated into XML format
- Modularity
- Gestural Scenes/Scenarios

1. How do you currently work with music-related movement and gesture data?

2. What are your needs of formats and standards?

3. What are your suggestions for future development?