Gesture Control of Music Systems

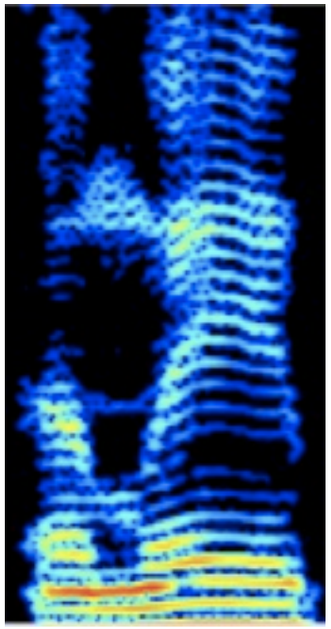
Frédéric Bevilacqua
Ircam
Real Time Musical Interactions Team

Frederic.Bevilacqua@ircam.fr
http://imtr.ircam.fr

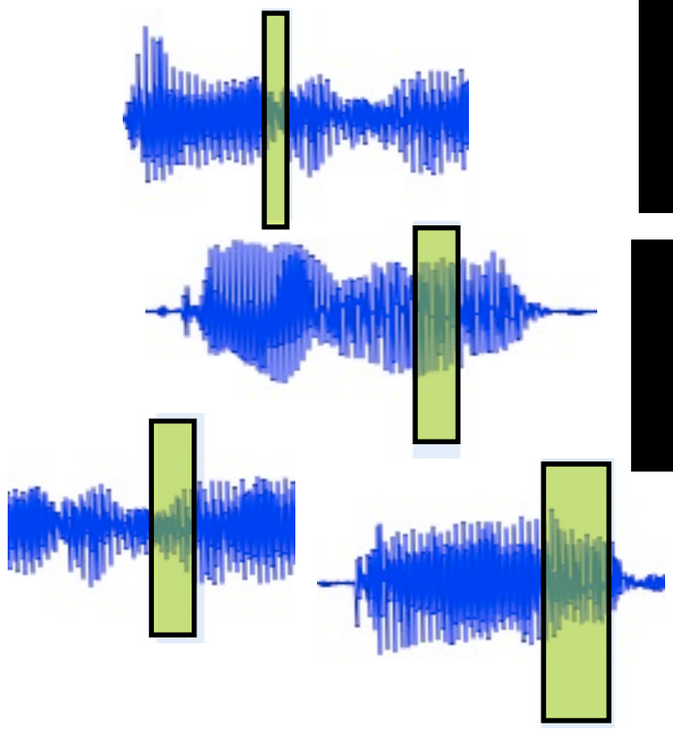
Plan

- Research Context
- Digital Musical Instruments
- Gesture and Music
  - Gesture Analysis/Recognition of Musicians Gestures
- Mapping between Gestures and Sounds
  - Gesture Following and Recognition
- Applications

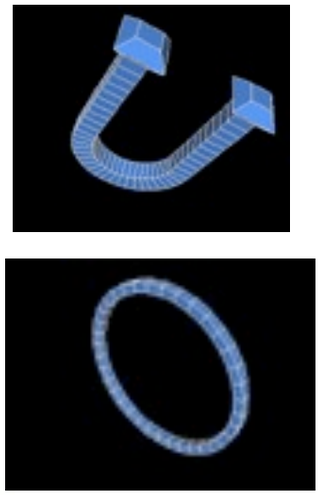
Sound Synthesis



analysis/
synthesis

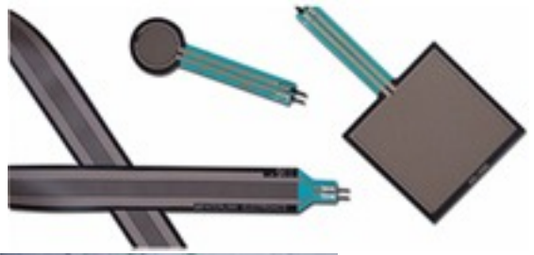


concatenative
synthesis



physical model

Gesture Capture



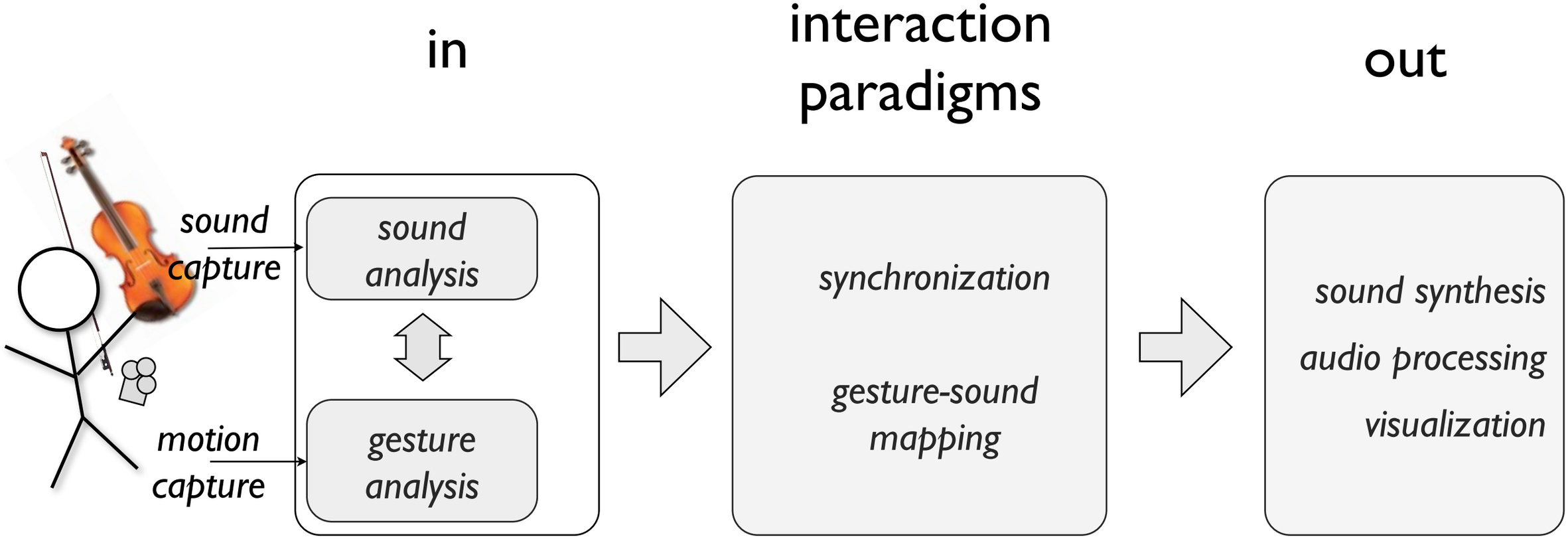
sensors

video

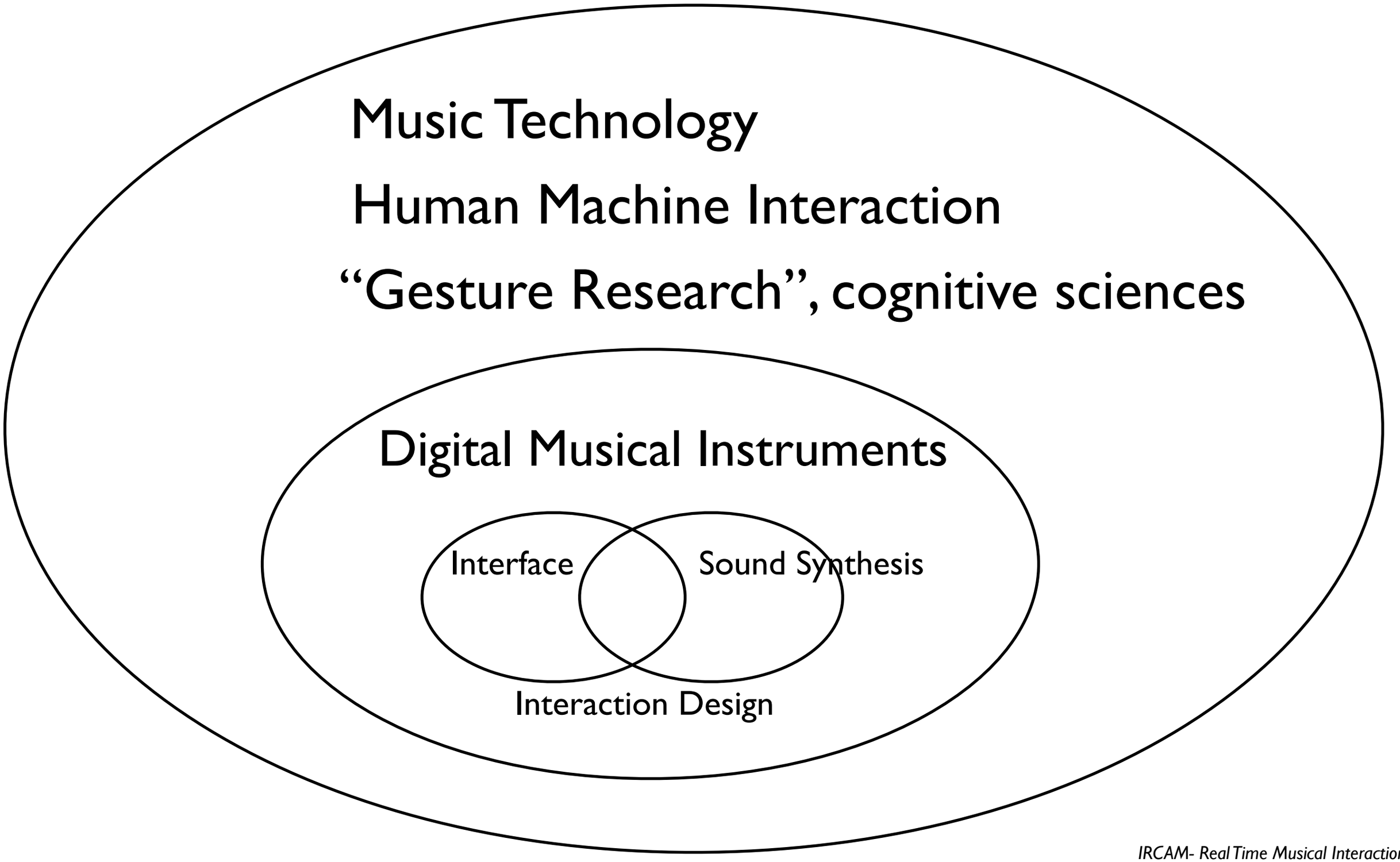
game interfaces

Digital Music Instruments

Musical Digital Instruments



Contexts


Music Technology
Human Machine Interaction
“Gesture Research”, cognitive sciences
Digital Musical Instruments
Interface
Sound Synthesis
Interaction Design

Digital Music Instruments

- *Instrument-like*
  - replicate an acoustic instrument
- *Instrument-inspired*
  - gesture or interface inspired from an acoustic instrument, but the final musical goal is different than the acoustic instrument
- *Extended instrument, Augmented Instrument, Hyper Instrument*
  - Acoustic instrument with additional sensors
- *Alternate controller*
  - New design

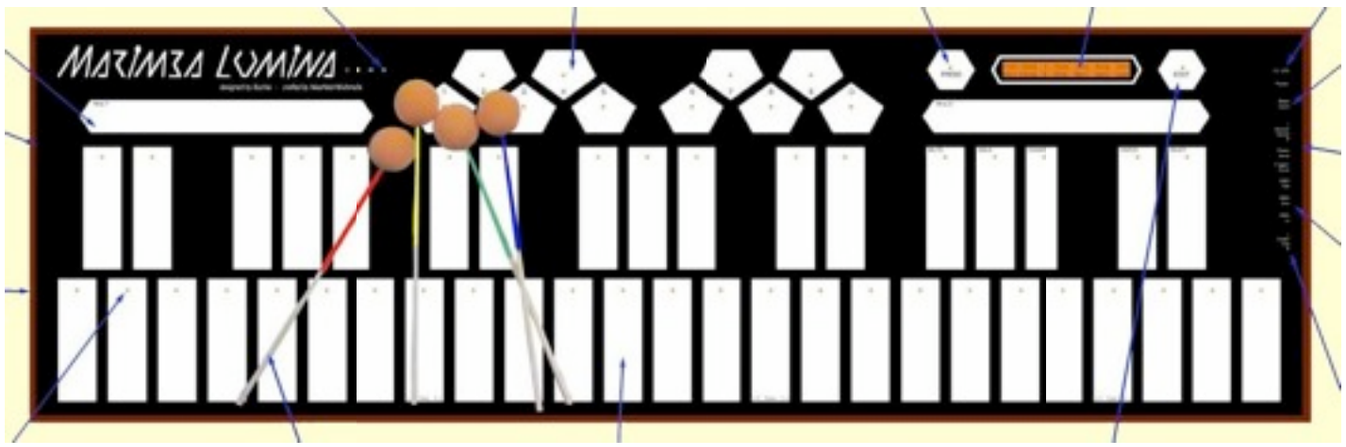
Marcelo M. Wanderley and Philippe Depalle. 2004. **"Gestural Control of Sound Synthesis".** *Proceedings of the IEEE*, vol. 92, No. 4 (April), pp. 632-644

Eduardo R. Miranda and Marcelo M. Wanderley. **New Digital Musical Instruments: Control and Interaction beyond the Keyboard**, A-R Editions, Spring 2006

« Instrument-like »

clavier MIDI Keyboard

EWI Electronic Wind Controller (AKAI)



Marimba Lumina (Buchla)

http://fr.youtube.com/watch?v=FNIKY5kGwLg

« Instrument-inspired »

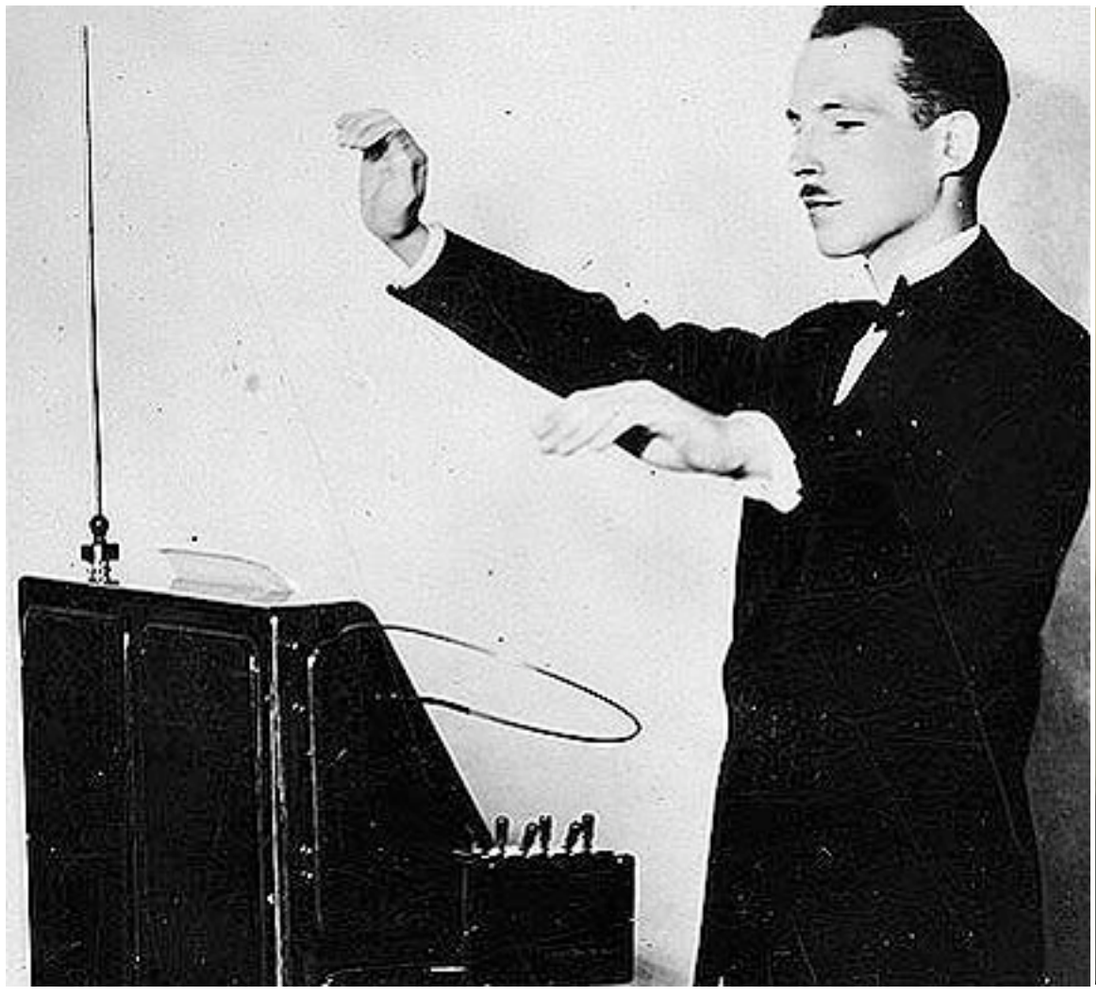
Violon MIDI - Suguru Goto

Augmented Instruments

HyperCello
Tod Machover / Yo-Yo Ma
(1991)

Clarinette & DataGlove,
Butch Rovan

Theremin, 1928



« Alternative controllers »

« The Hands », Michel Waisviz

http://fr.youtube.com/watch?v=U1L-mVGqug4

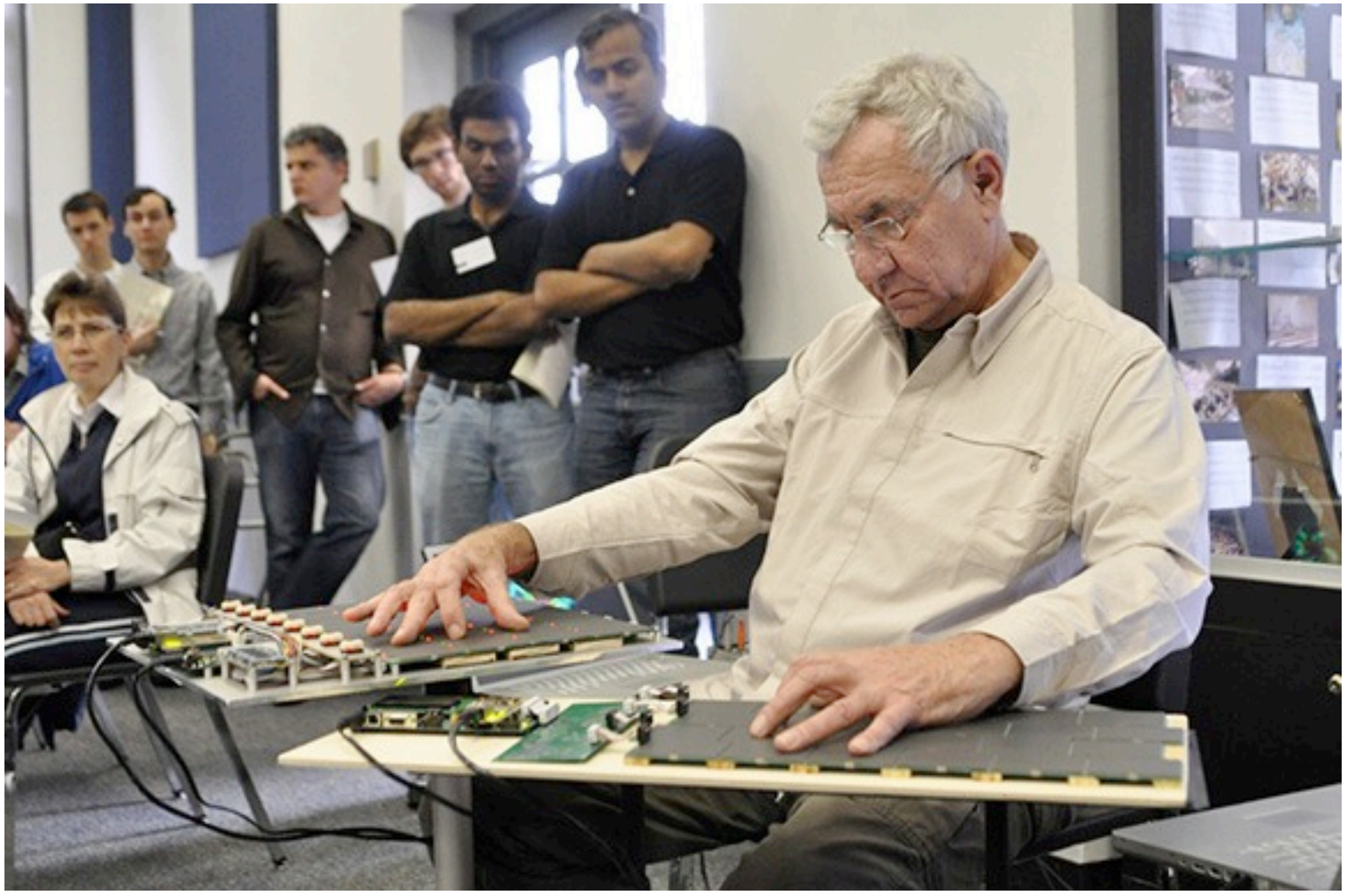
Le Méta-Instrument - Serge de Laubier

« The Hands », Michel Waisviz

Georgia Tech's Guthman Musical Instrument Competition (2009)

Jaime Oliver's Silent Drum

Georgia Tech's Guthman Musical Instrument Competition (2009)



the Slabs, David Wessel (CNMAT, Berkley)

Commercial interfaces

?

ɔctions

Stanford Laptop Orchestra (SLOrk)

Stanford Mobile Phone Orchestra (MoPhO)

"do mobile phones dream of electric orchestras?"







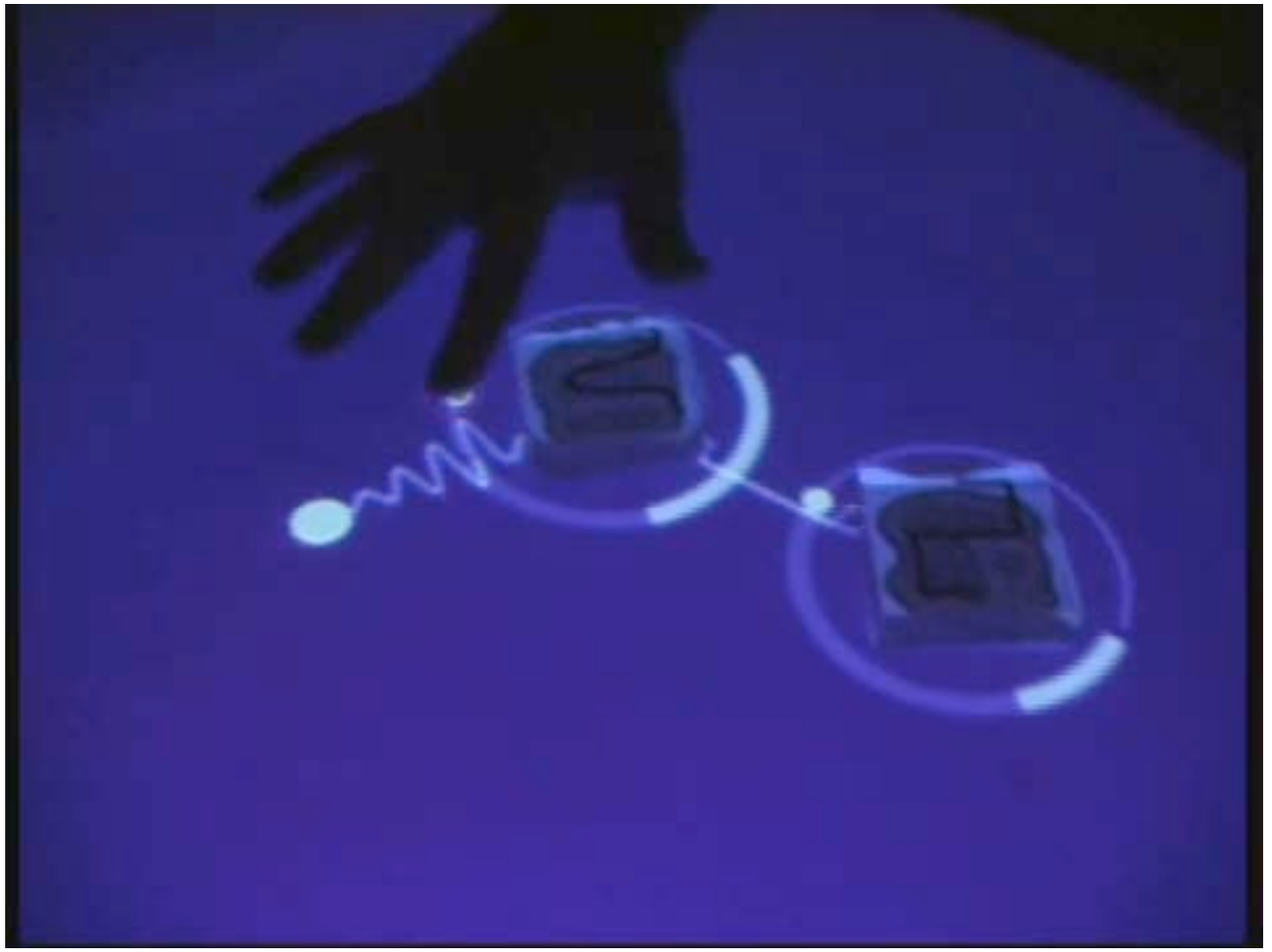
http://mopho.stanford.edu/

Da Fact

reactable

http://www.reactable.com/

reactable



Installation Grainstick

- Cité des Sciences Paris

Pierre Jodlowski Raphaël Thibault

Ircam

Applications

- Music & New Media
  - professional level, music performance, composition
  - music pedagogy
  - music game

- HCI: interaction paradigms using “expressive gestures”

- Rehabilitation (?)
  Sonification of gesture/action (?)

Links to the HCI field

- Notion of *embodied interaction*
  - P. Dourish *Where The Action Is: The Foundations of Embodied Interaction, MIT Press*
  - M. Leman *Embodied Music Cognition and Mediation Technology, MIT Press*
- Tangible interfaces, augmented reality
- Affective computing
- Collaborative and distributed interaction

Bill Buxton

- http://www.billbuxton.com/buxtonIRGVideos.html

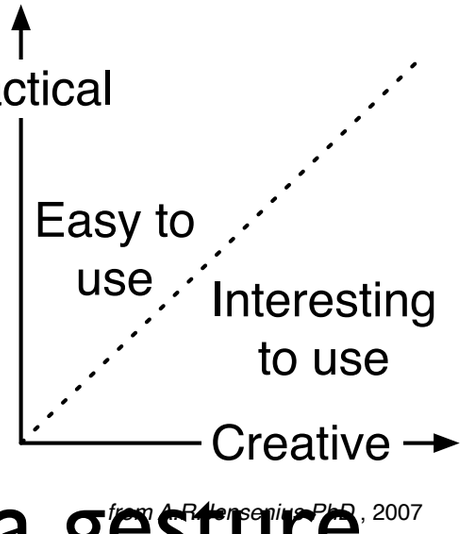
- http://www.youtube.com/watch?v=Arrus9CxUiA

Musical Interfaces

- action-perception loop

- importance of timing and synchronization
  - requirements: low latency (< 10 ms)

- from triggering events...
  to using *continuous gestures*



from A.Ridensemius PhD, 2007

- notion of *expressivity*: measure of "*how*" is a gesture performed

- notions of "goal" and "efficiency" different than in standard HCI

"Clearly, electronic music systems allow much freedom for the performer, because the mappings between control units, on the one hand, and some production units, on the other hand, are not constrained be any biomechanical regularities. (...). However, as most electronic music performers know, it is exactly this freedom of mapping that may disturb the sens of contact and of non-mediation".

"Can we find a way of interacting with machines so that artistic expression can be fully integrated with contemporary technologies? »

Marc Leman, *Embodied Music Cognition and Mediation Technology*, MIT Press.

Gesture and Music

Gesture and Music

Some references:

- Cadoz, C. and M. M. Wanderley, Gesture - Music, in Trends in Gestural Control of Music, M. M. Wanderley and M. Battier, Editors. 2000, Ircam - Centre Pompidou: Paris, France. p. 71--94.

- Jensenius, A. R., M. M. Wanderley, R. I. Godoy and M. Leman (2010). Concepts and Methods in Research on Music-related Gestures. In Godøy, R. I. and M. Leman (Eds.), Musical Gestures: Sound, Movement and Meaning. Routledge.

Types of Musical Gestures

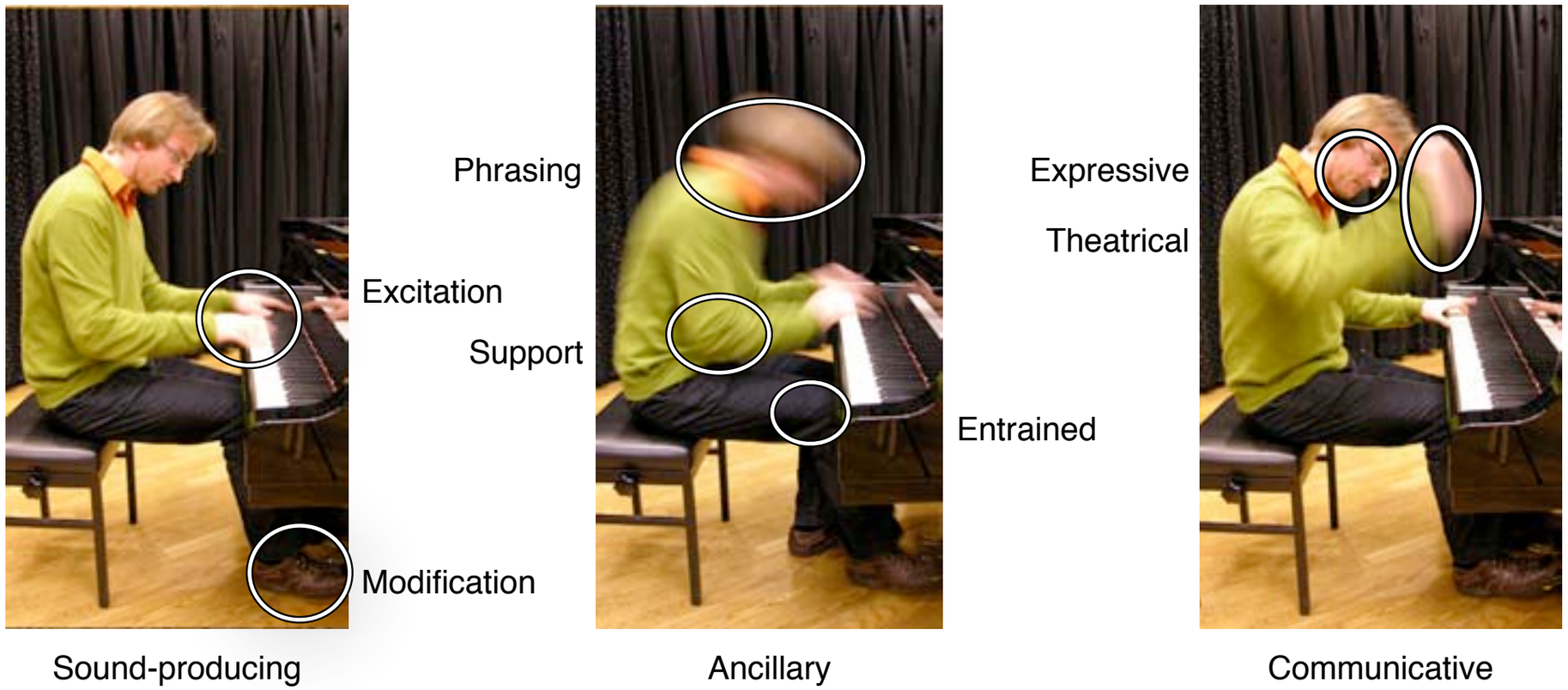
Ancillary,
sound-accompanying,
and communicative

Sound-producing

Sound-modifying

Jensenius, A. R., M. M. Wanderley, R. I. Godoy and M. Leman (2010). Concepts and Methods in Research on Music-related Gestures. In Godøy, R. I. and M. Leman (Eds.), *Musical Gestures: Sound, Movement and Meaning*. Routledge.

Types of Musical Gestures



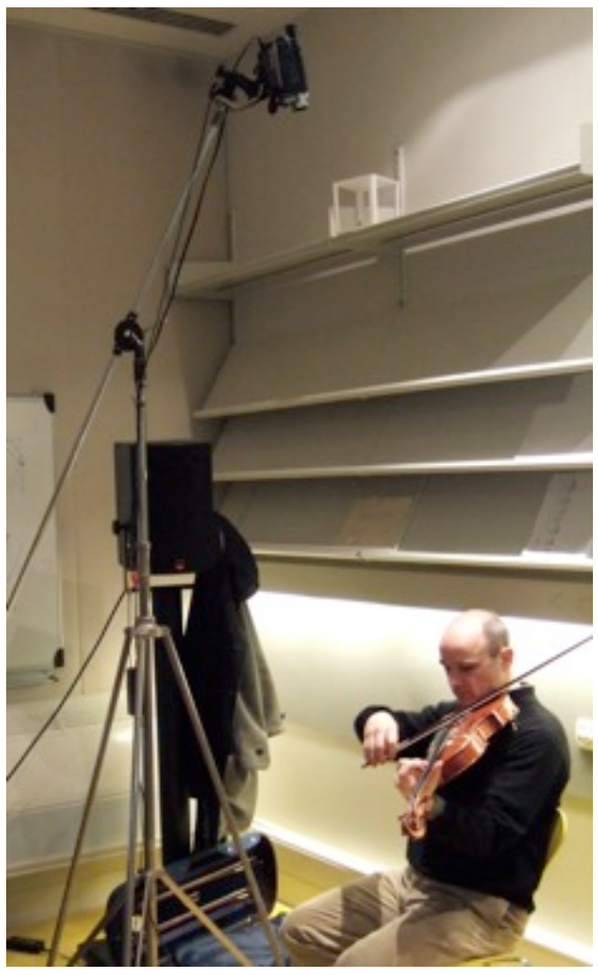
Jensenius, A. R., M. M. Wanderley, R. I. Godoy and M. Leman (2010). Concepts and Methods in Research on Music-related Gestures. In Godøy, R. I. and M. Leman (Eds.), *Musical Gestures: Sound, Movement and Meaning*. Routledge.

Capturing Musician Motion

Violin bowing

sensor attached on the bow

3D optical motion capture



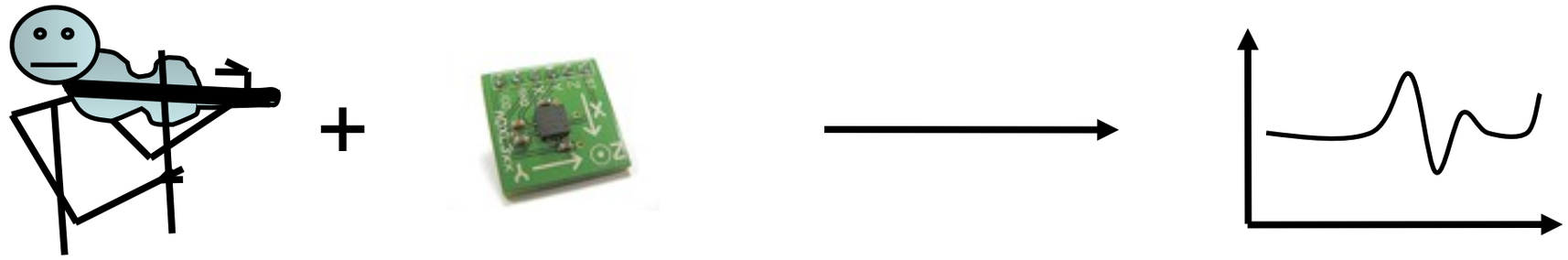
hybrid system

- F. Bevilacqua, N. Rasamimanana, E. Fléty, S. Lemouton, F. Baschet « The augmented violin project: research, composition and performance report » NIME 06

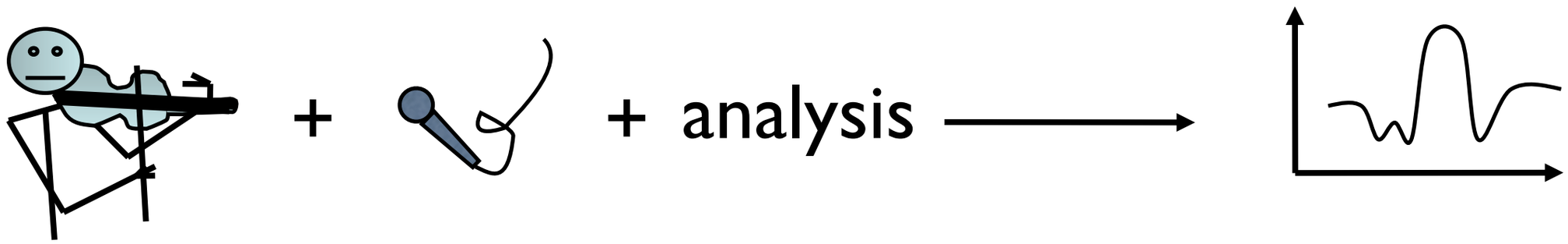
- E. Schoonderwaldt, N. Rasamimanana, F. Bevilacqua « Combining accelerometer and video camera: Reconstruction of bow velocity profiles », NIME 2006

Capturing Musician Gestures

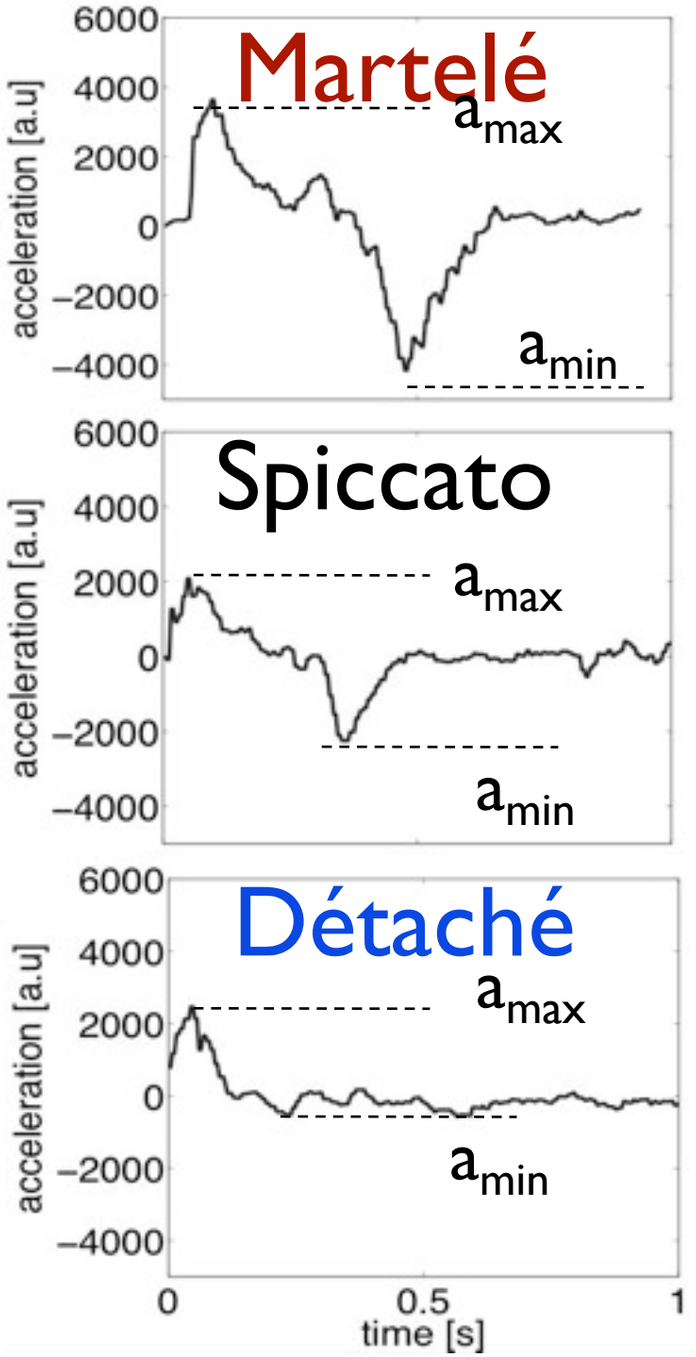
- *Direct* capture of movement, pressure etc using sensors

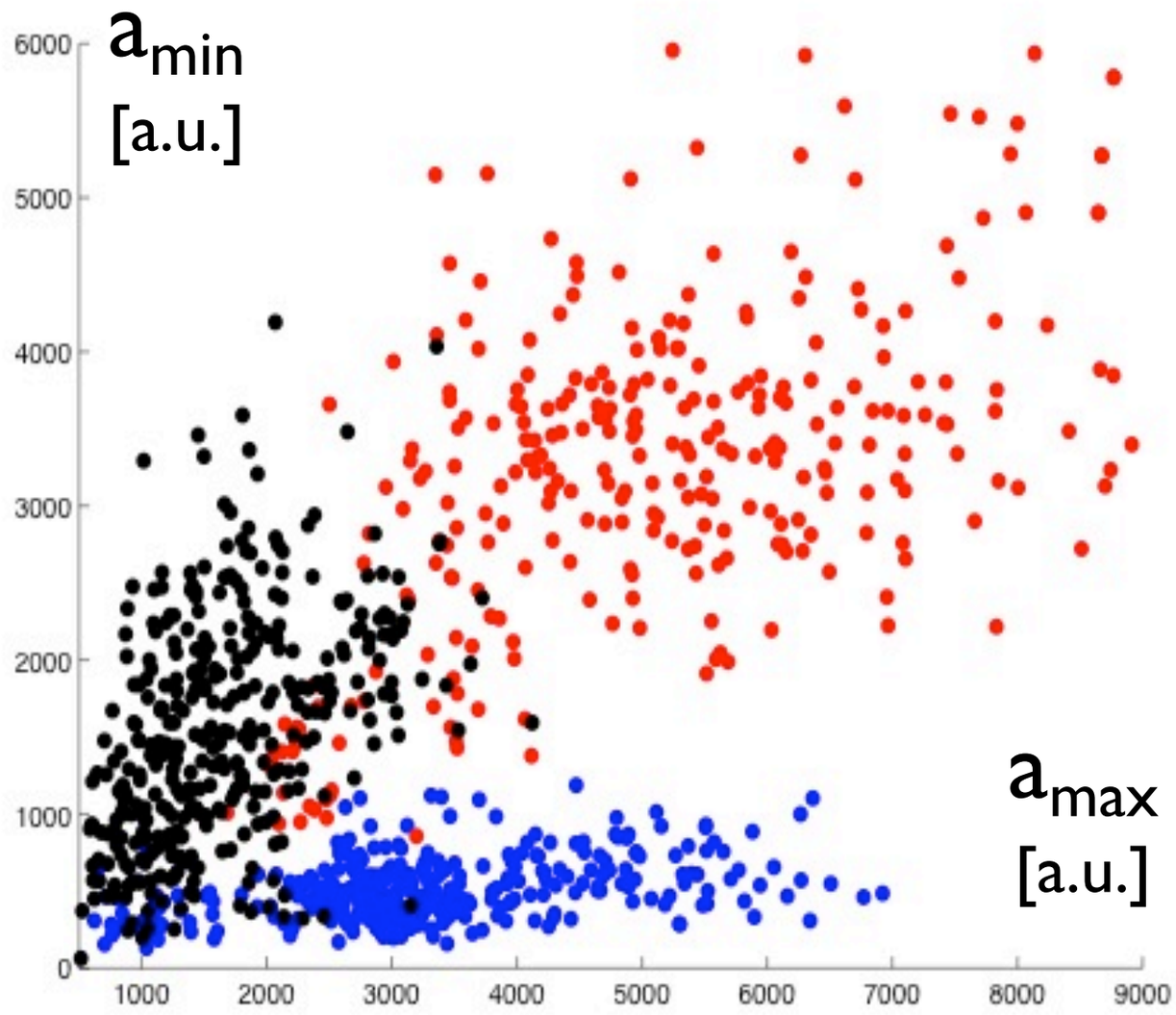


- Indirect capture based on the sound analysis



Bowing styles characterization





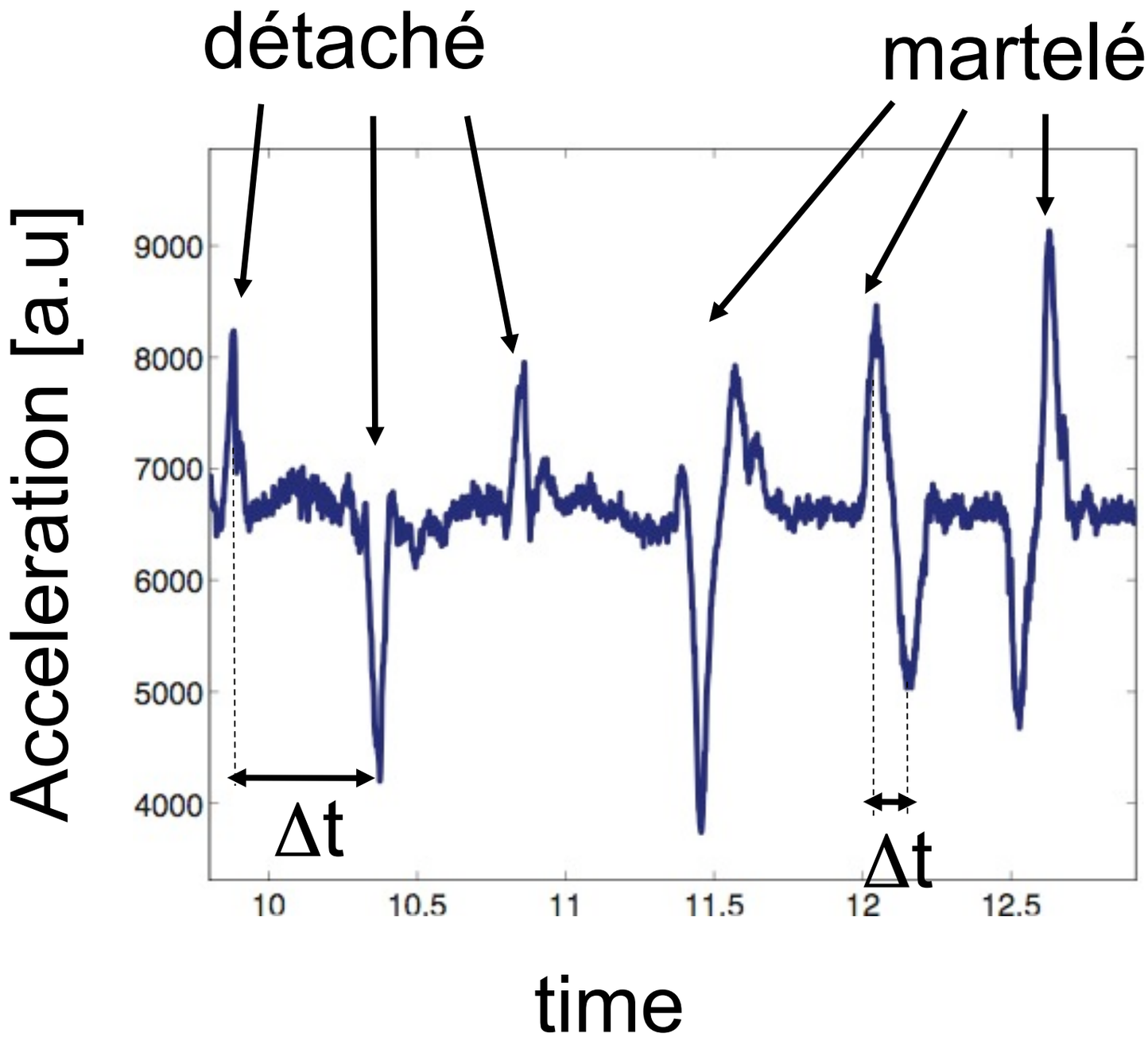
2 violin players, 2 tempi (60 bpm, 120 bpm)
dynamics (pp, mf, ff)

N. Rasamimanana et al., GW 2005, *Lecture Notes in Artificial Intelligence* 3881, pp. 145–155, 2006.

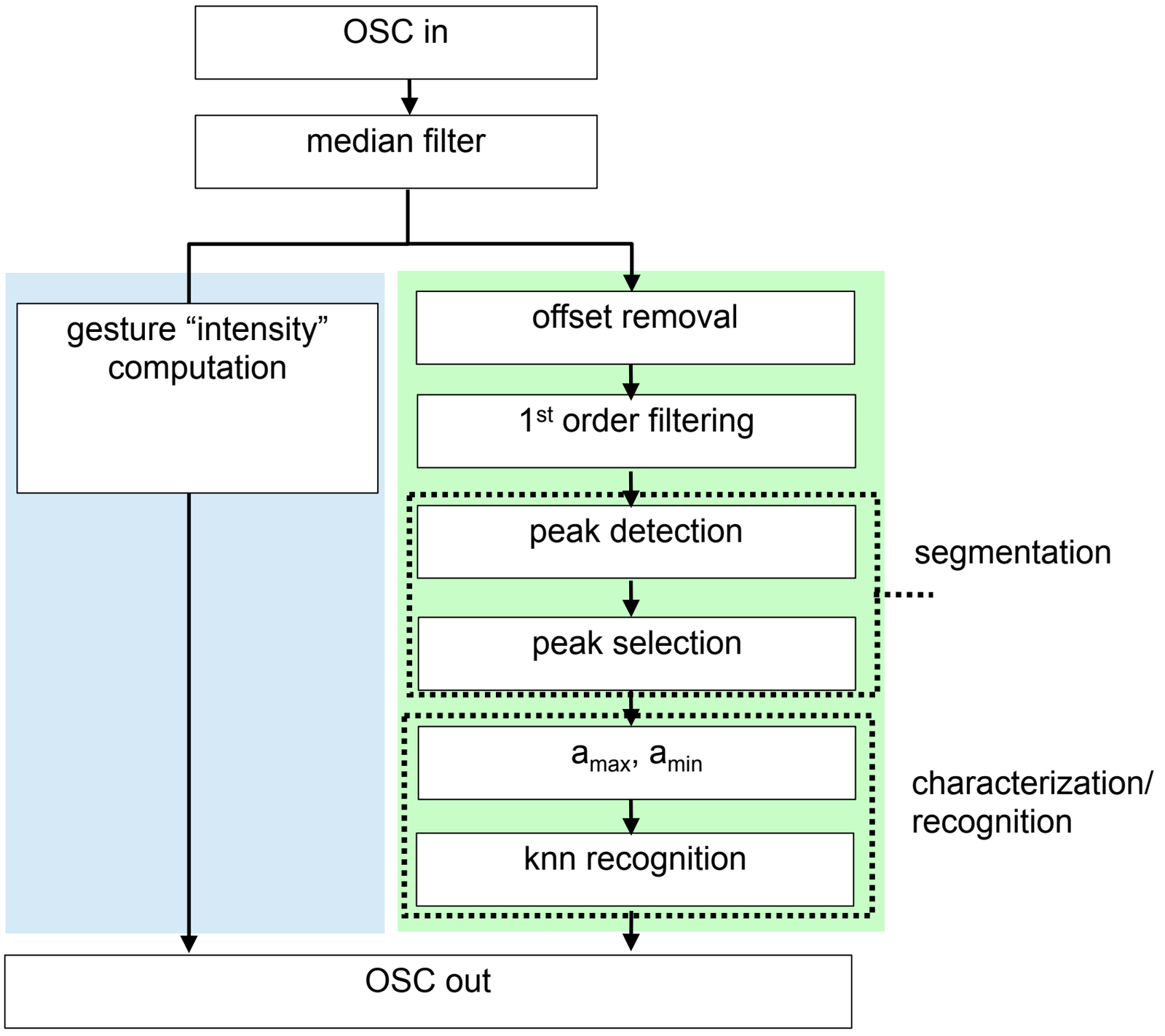
Similar works

- PCA + KNN
  - D. Young. Classification of common violin bowing techniques using gesture data from a playable measurement system. In in NIME 2008 Proceedings, 2009.

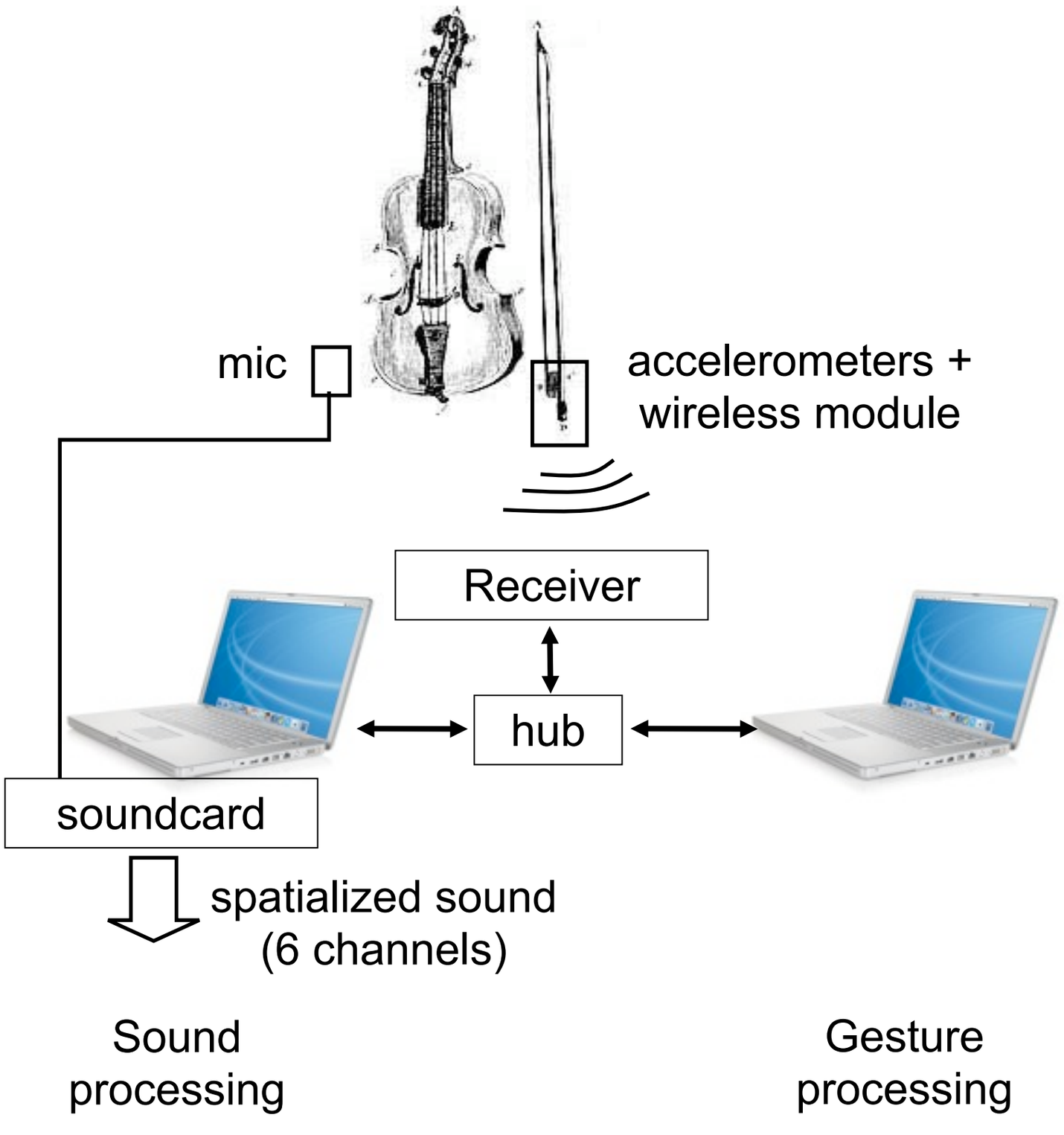
Bowing - Segmentation



Bowing recognition: Real time implementation (Max/MSP)



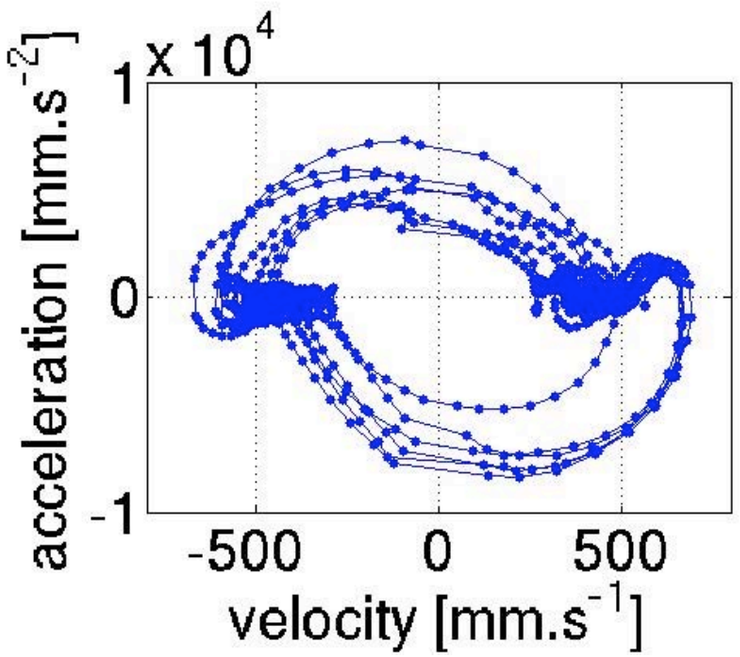
BogenLied -


mic
accelerometers +
wireless module
Receiver
hub
soundcard
spatialized sound
(6 channels)
Sound
processing
Gesture
processing

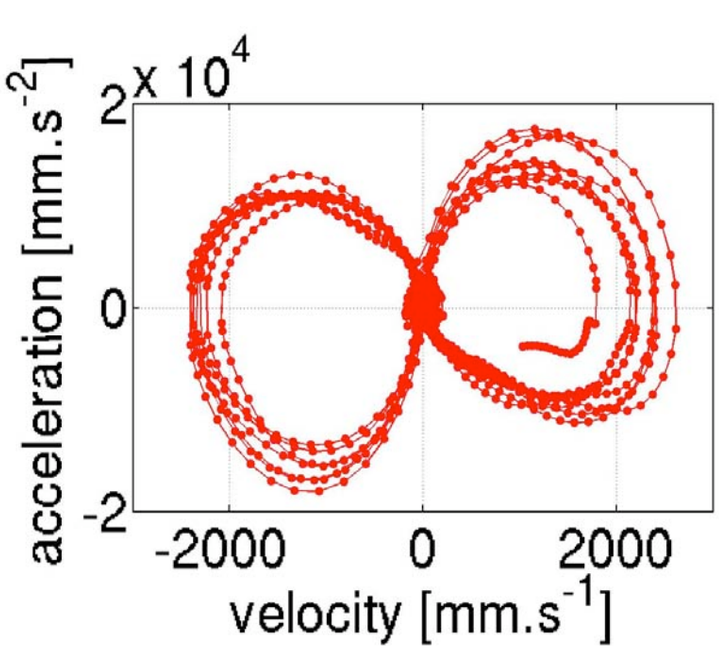
Bowing styles

acceleration vs velocity

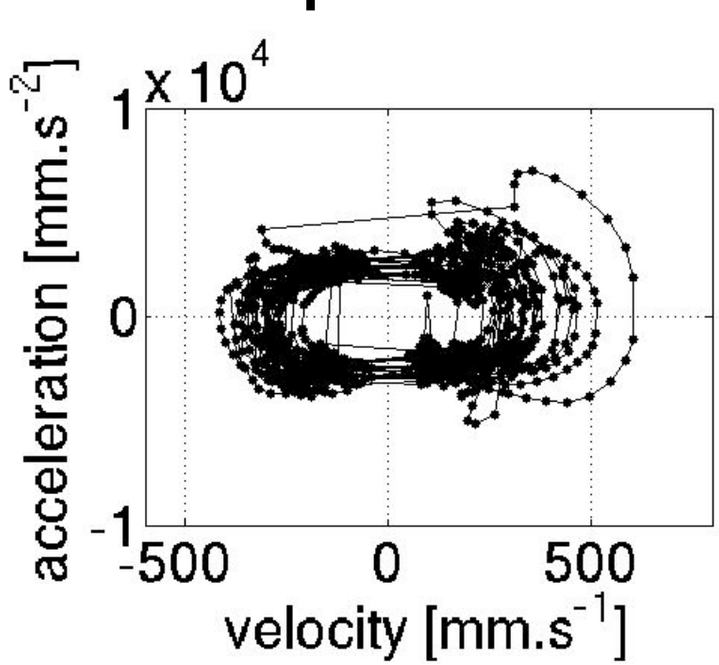
détaché



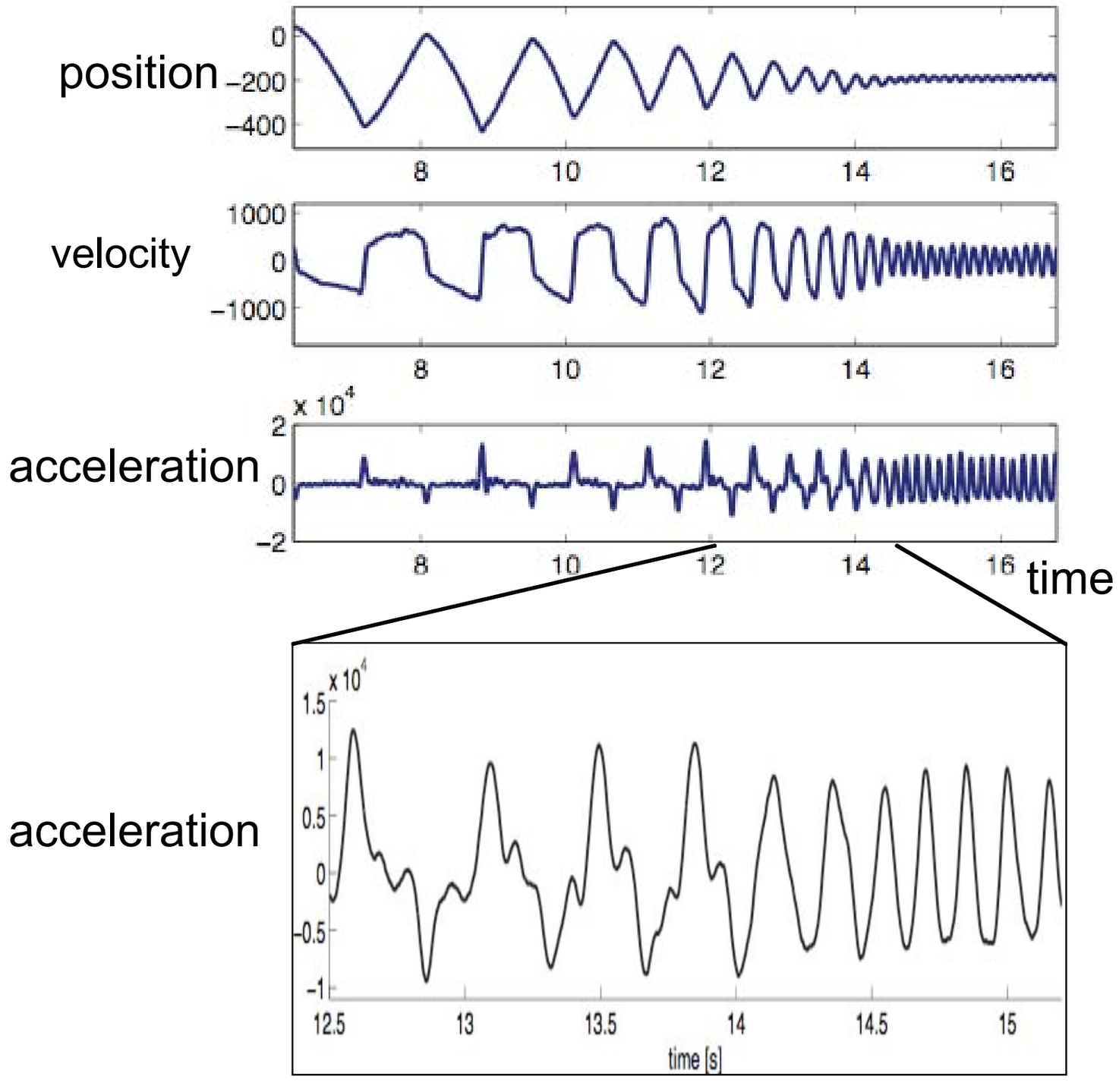
martelé



spiccato



Influence on bowing frequency

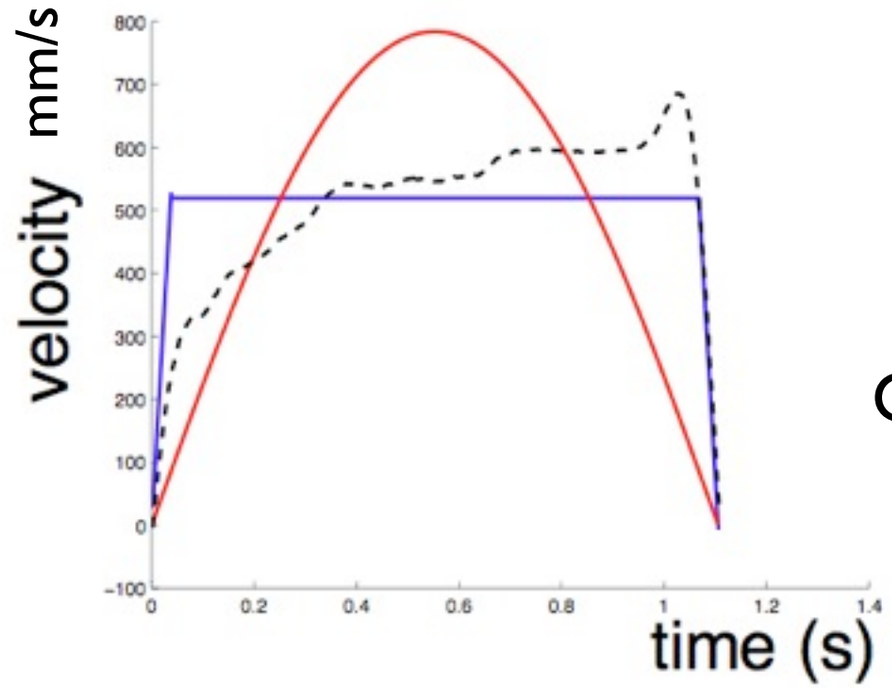


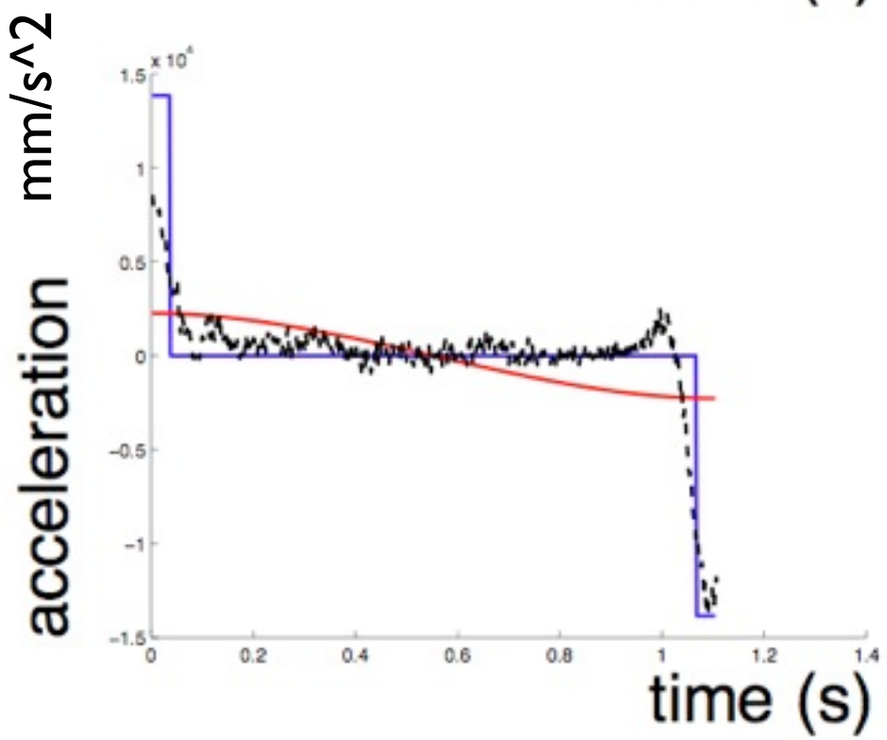
video

N. Rasamimanana et al., GW 2007, *Lecture Notes in Artificial Intelligence*

Bowing model

- Minimizing
  - Minimum impulse : trapezoïdal “continuous control”
  - Minimum jerk (discrete) “balistic control”



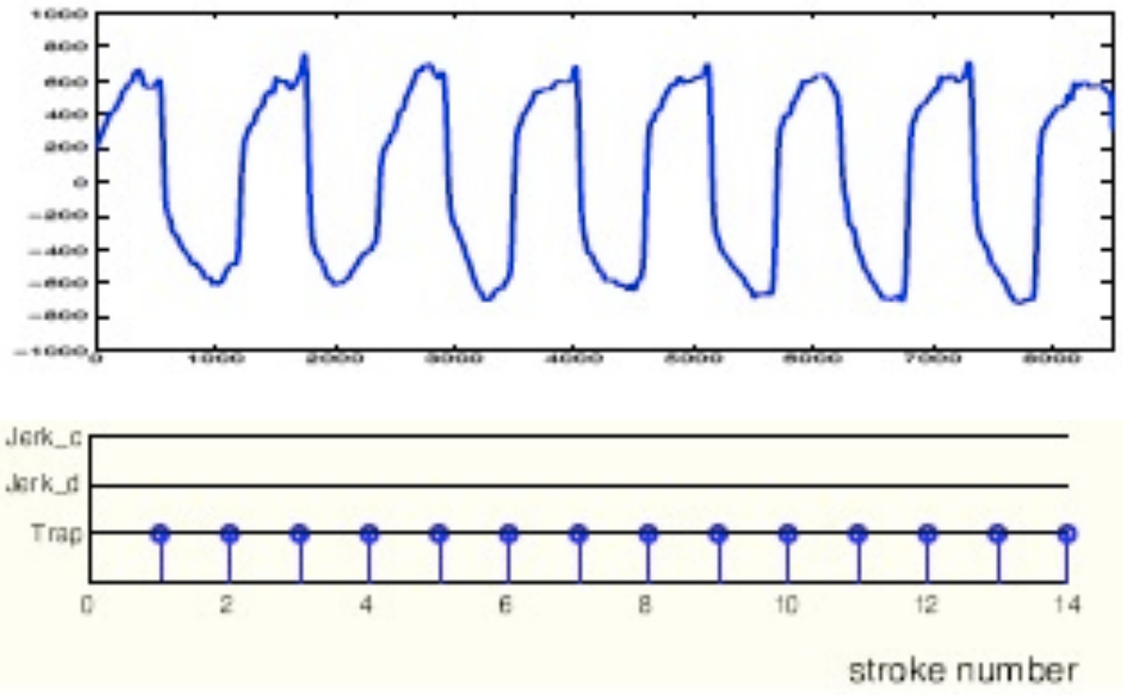


Trap
Jd
Original

Bowing style - scale

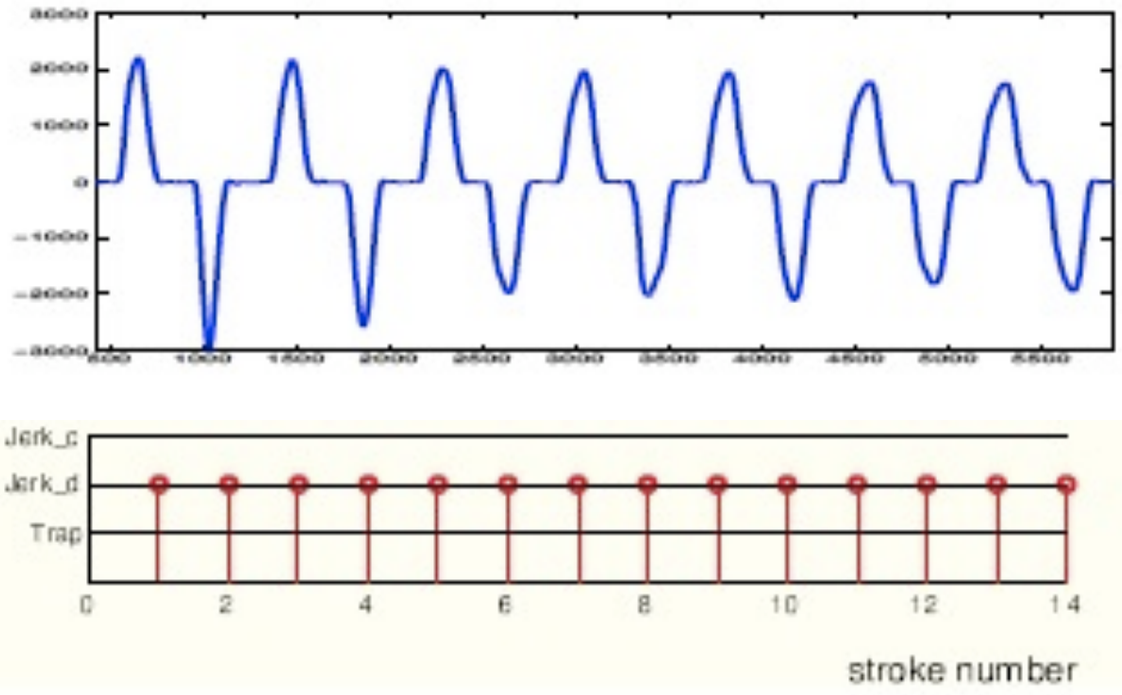
Finding the best model

Détaché



Minimum impulse (Trapezoidal)

Martelé

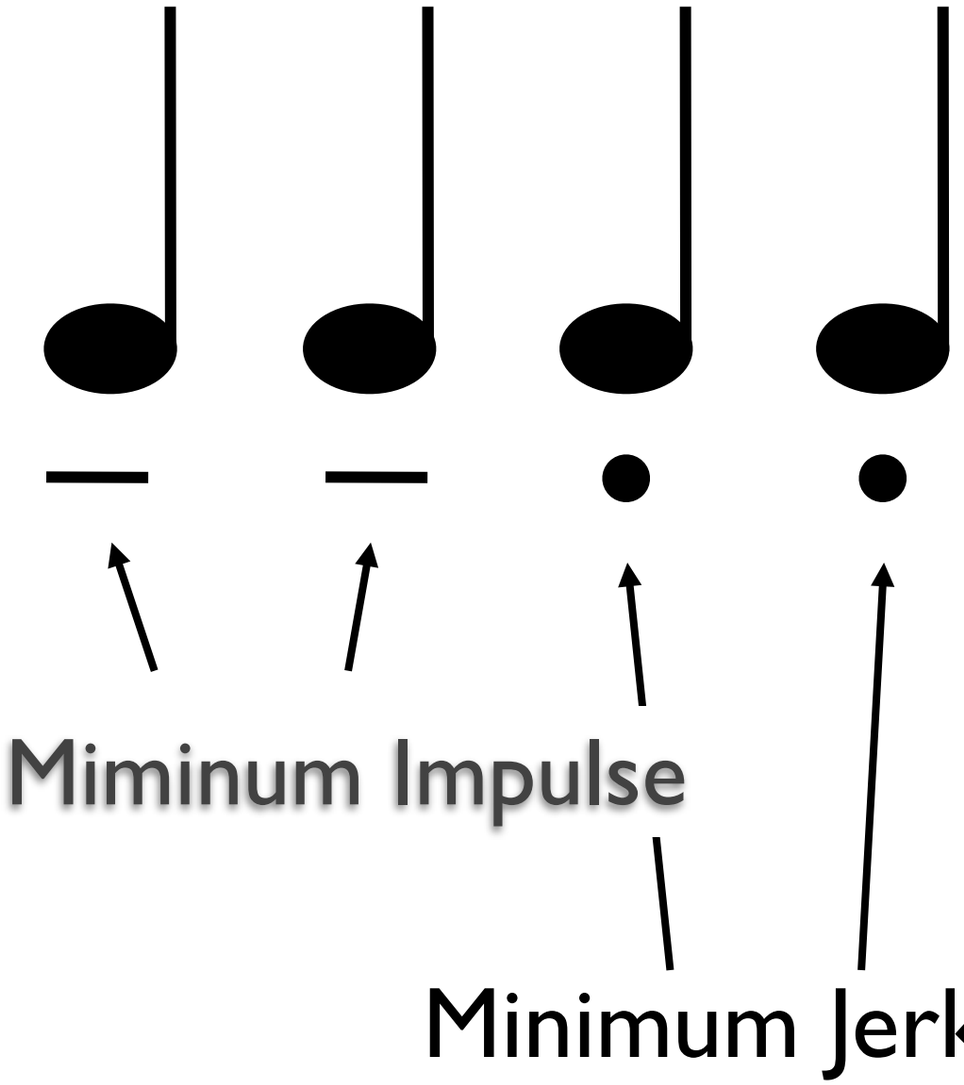


Minimum Jerk

N. Rasamimanana, F. Bevilacqua. « Effort-based analysis of bowing movements: evidence of anticipation effects ». Journal of New Music Research,

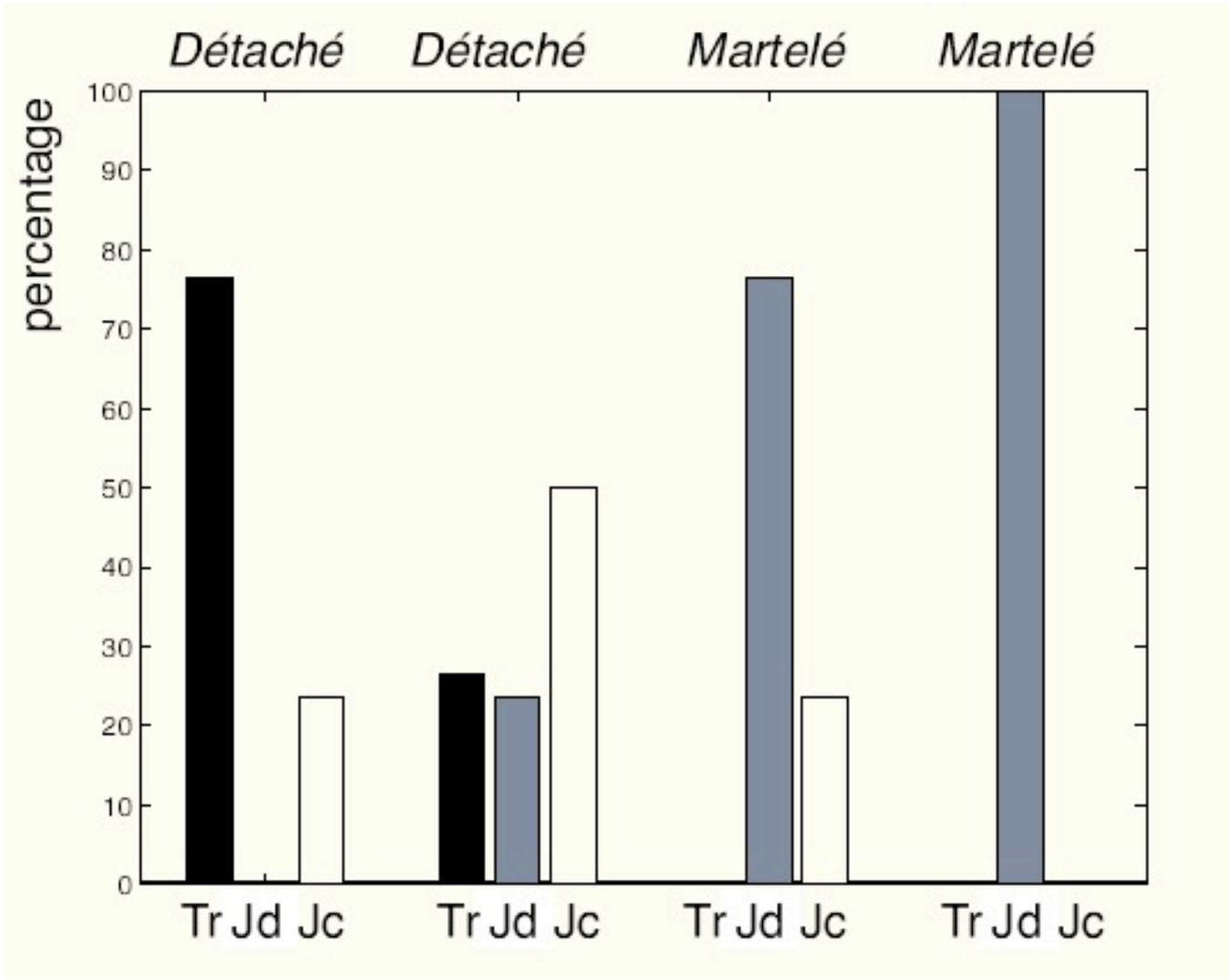
Gestural Co-articulation

detaché *martelé*



Miminum Impulse

Minimum Jerk

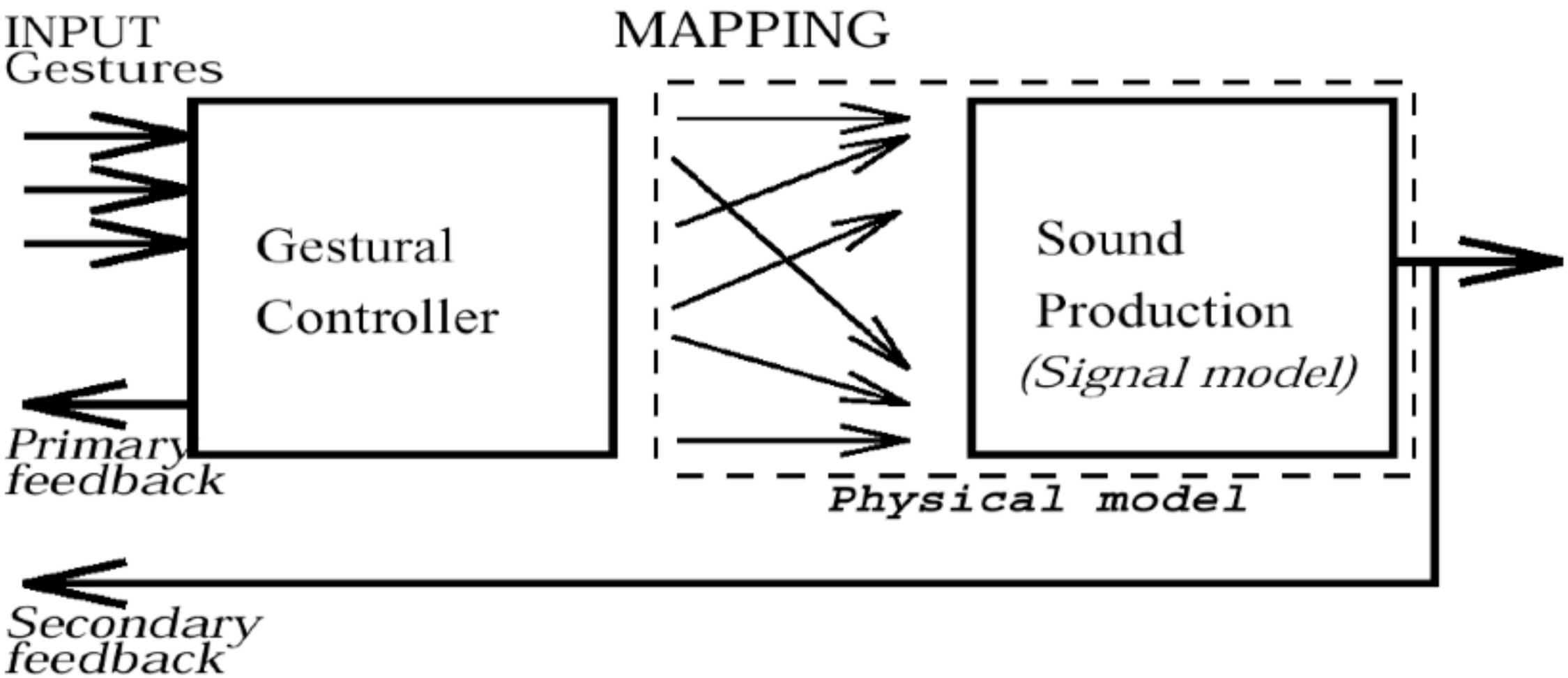


Co-articulation effect

- major difficulty for segmentation and characterization
  - using di-gesture ? (similarly to diphone)
- can be used to anticipate (towards intention ?)
- expressivity links to co-articulation

Gesture to Sound Mapping

Mapping

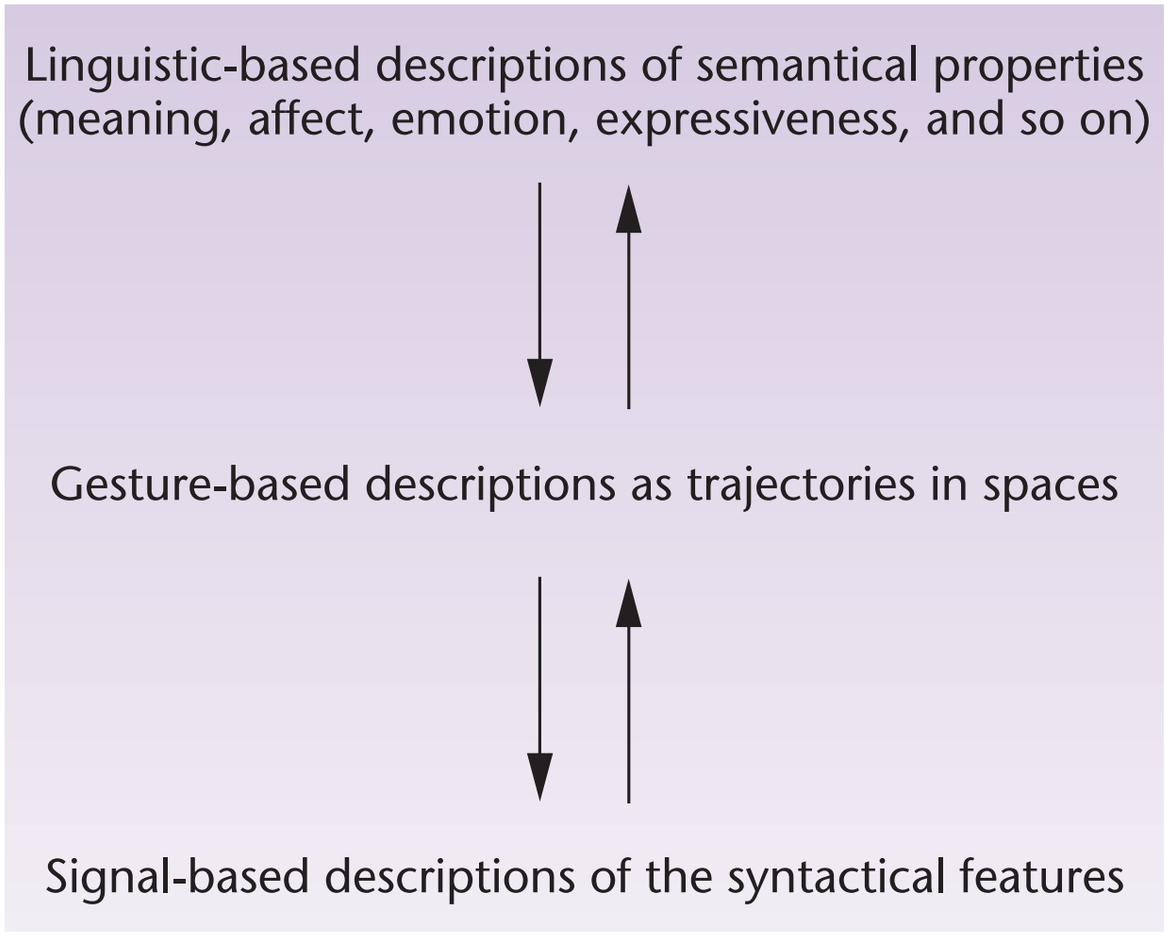


Wanderley, M. 2001. *Performer-Instrument Interaction: Applications to Gestural Control of Music. PhD Thesis. Paris, France: University Pierre et Marie* Curie - Paris VI

See also:
•"Mapping Strategies in Interactive Computer Music." Organised Sound, 7(2), Marcelo Wanderley Ed.
•Wanderley, M and Battier, M -editors. “Trends in Gestural Control of Music”. IRCAM, Centre Pompidou, 2000.

Mapping

- Low Level vs High Level



Antonio Camurri, Gualtiero Volpe, Giovanni De Poli, Marc Leman, "Communicating Expressiveness and Affect in Multimodal Interactive Systems," *IEEE MultiMedia*, vol. 12, no. 1, pp. 43-53, Jan. 2005

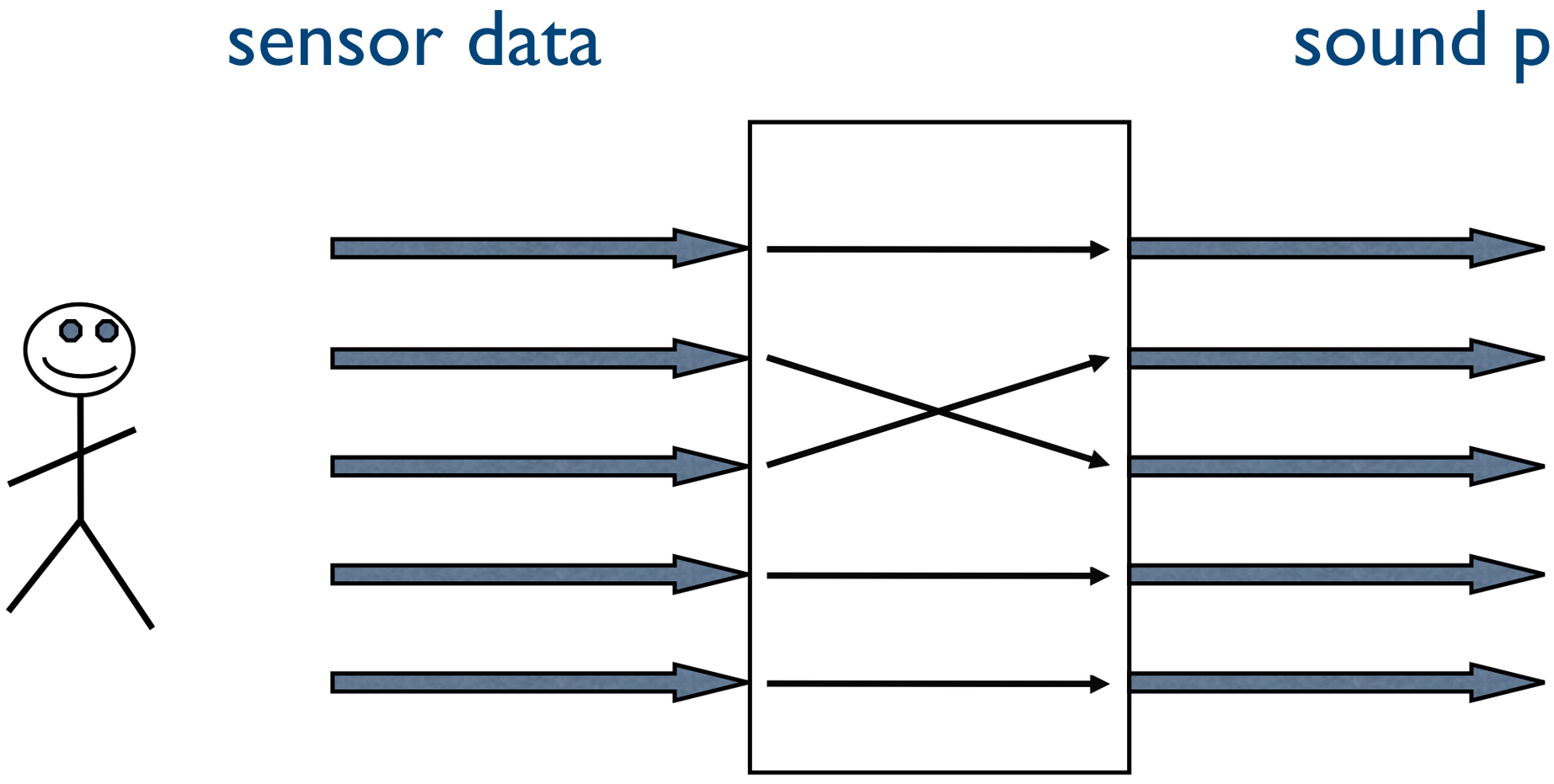
Mapping

- Spatial vs Temporal :
  - « Spatial » : relationship independent of the temporal ordering of data
  - « Temporel » : relationship between temporal processes

- Direct vs Indirect
  - Direct :
    - sensor data directly connected to music parameters
    - relationship “manually” set
  - Indirect
    - uses machine learning techniques to set the relationship

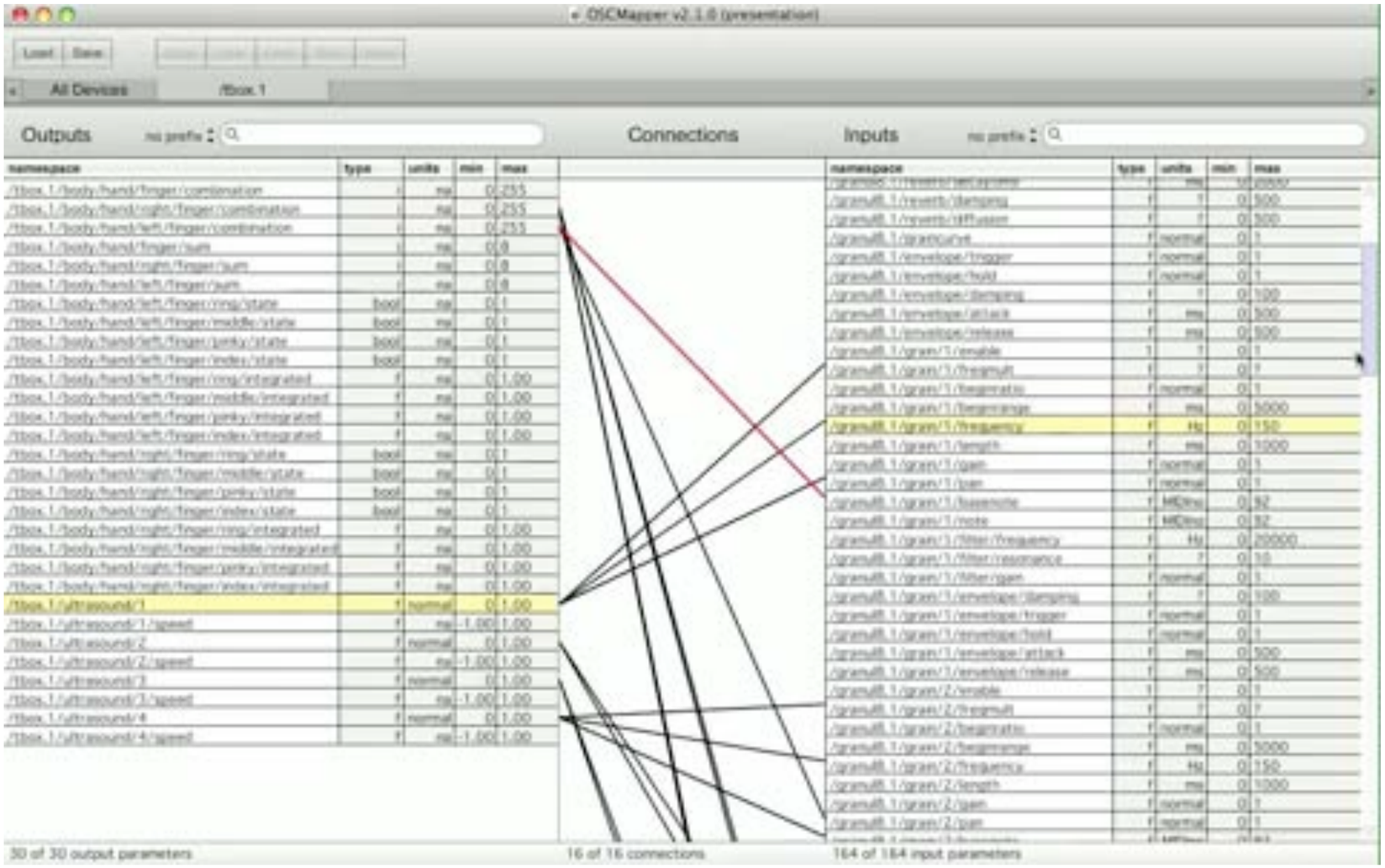
Mapping (Spatial)

- one-to-one



Mapping Musical Instruments

IDMIL lab, Mc Gill



Mapping

- one-to-many

sensor data

sound parameters

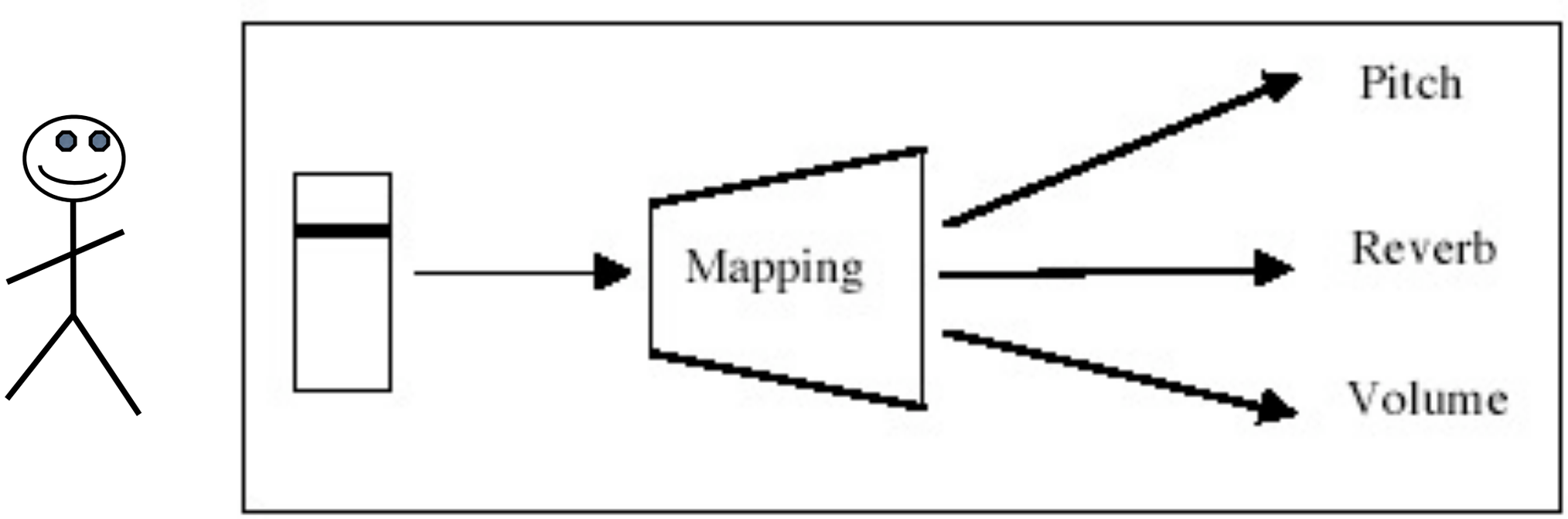
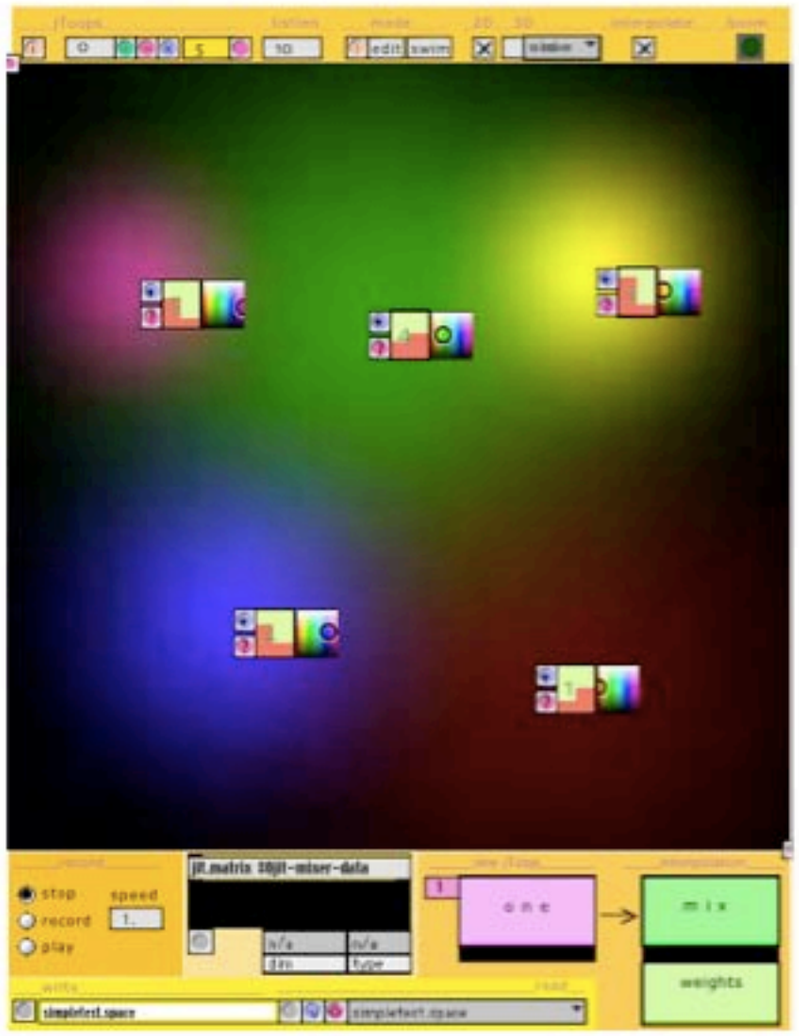
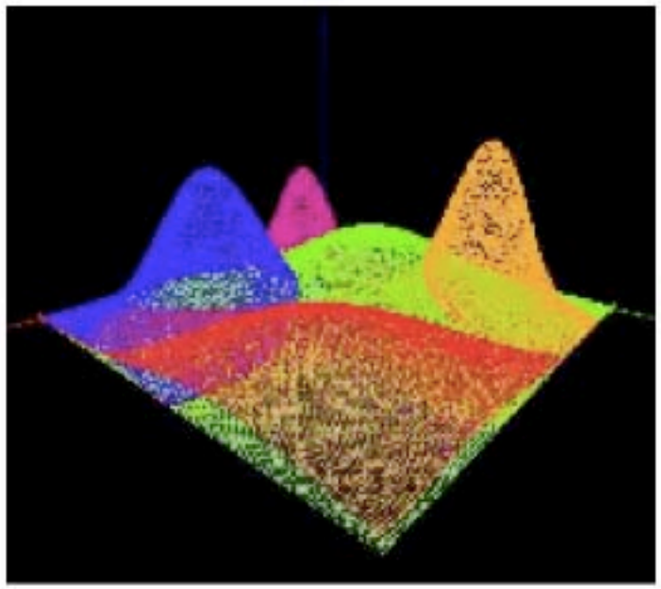


Fig. 2. Divergent Mapping: One control operates many parameters.

Interpolation





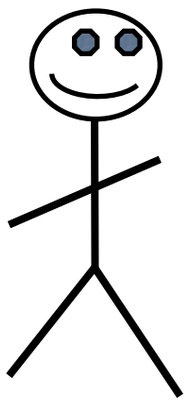
Ali Momeni, David Wessel: Characterizing and Controlling Musical Material Intuitively with Geometric Models. NIME 2003

Mapping

- many-to-one

sensor data

sound parameters



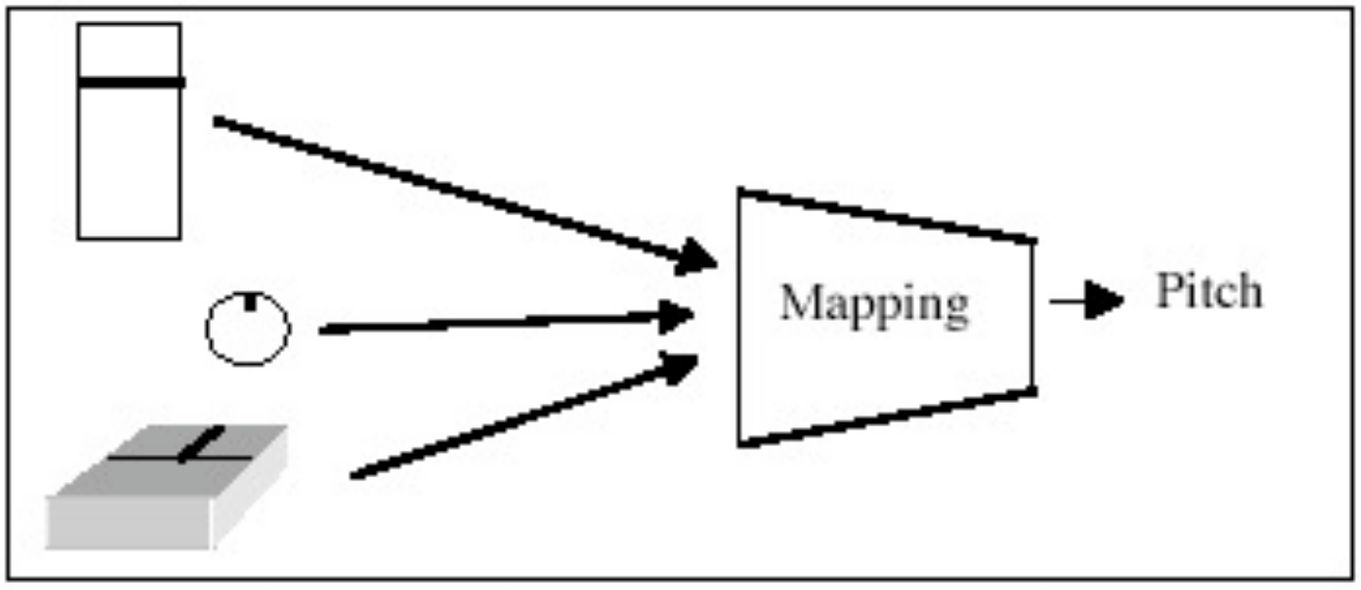
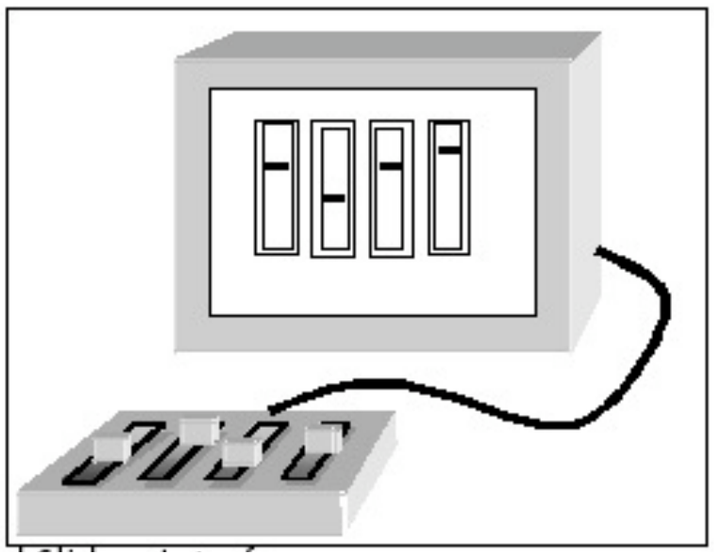


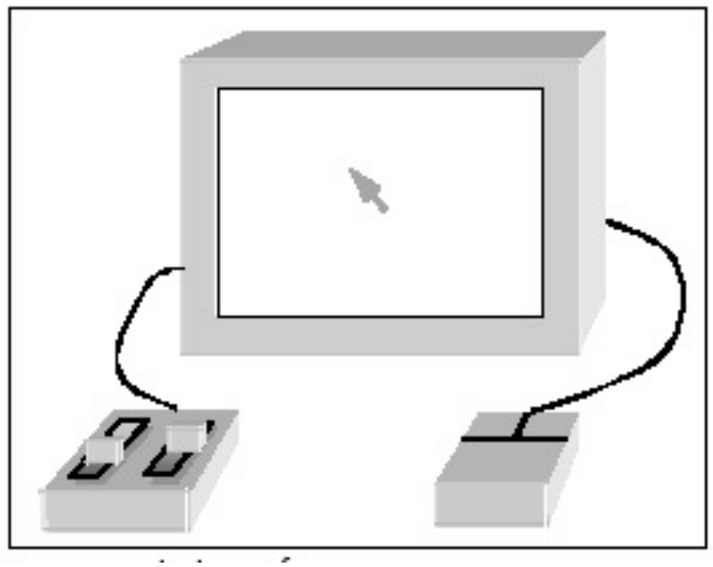
Fig. 1. Convergent Mapping; Many controls operate one parameter.

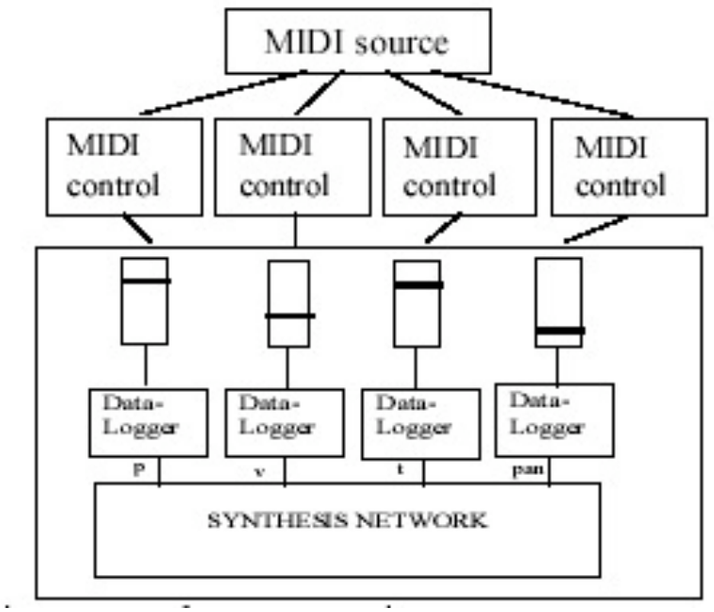
Simple ou complexe mapping ?

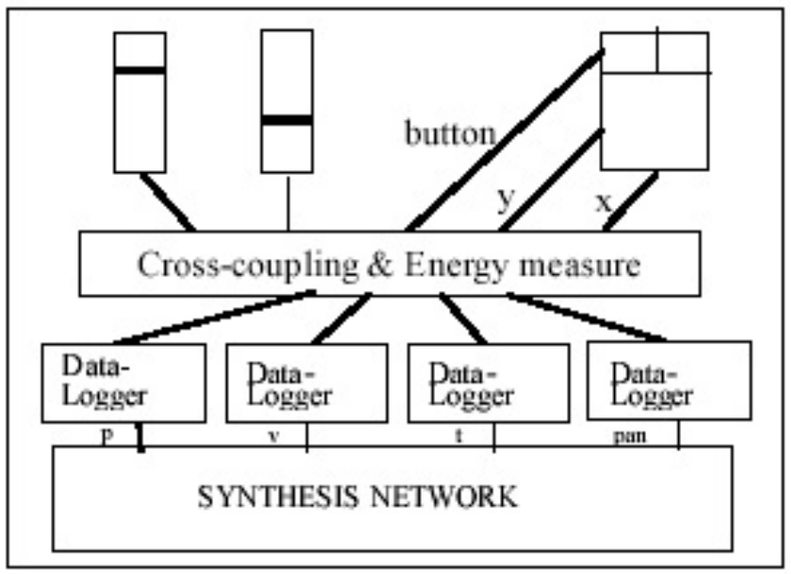
Hunt, A., and Kirk, R. 2000. "Mapping Strategies for Musical Performance."

In M. Wanderley and M. Battier, eds. Trends in Gestural Control of Music. Ircam, Centre Pompidou.









The Physical Sliders Interface

The Multiparametric Interface

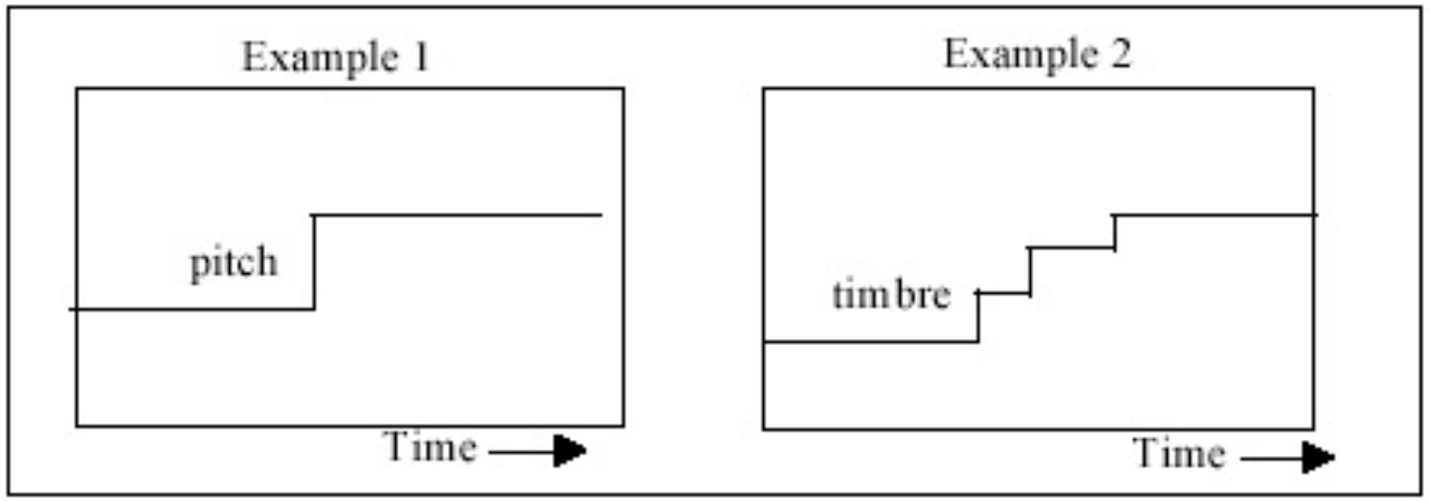


Fig. 16. Group A sounds: stepwise uni-parameter changes.

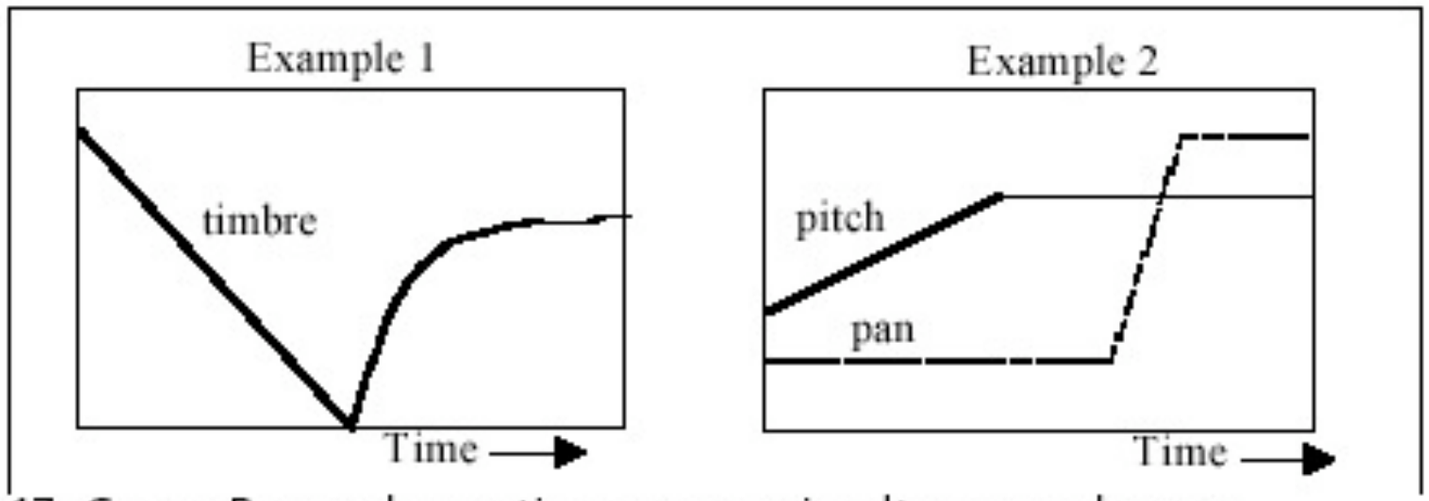


Fig. 17. Group B sounds: continuous non-simultaneous changes.

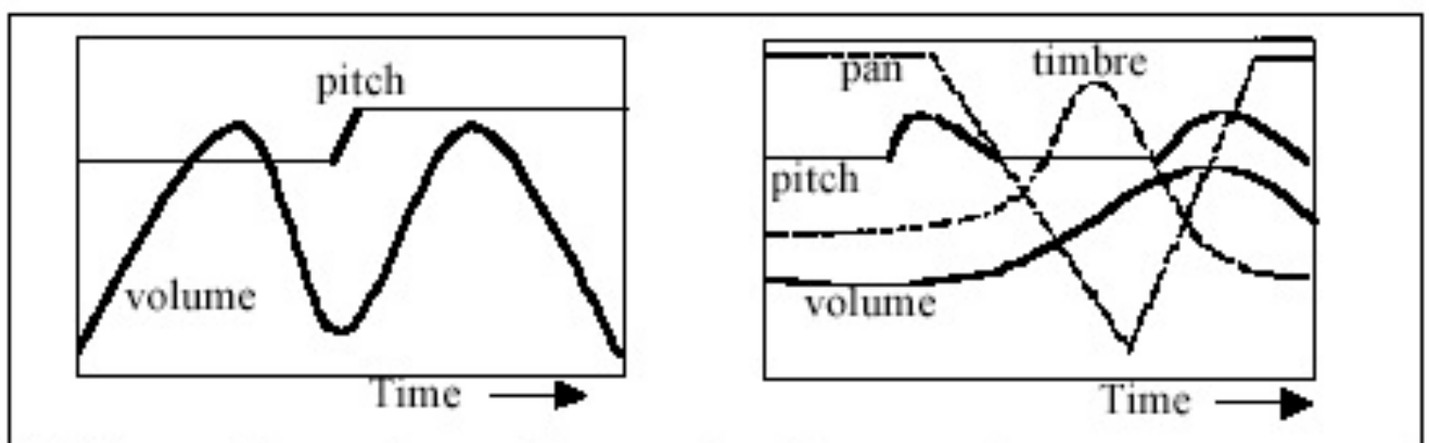


Fig. 18. Group C sounds: continuous simultaneous changes.

Conclusions of Hunt and Kirk study

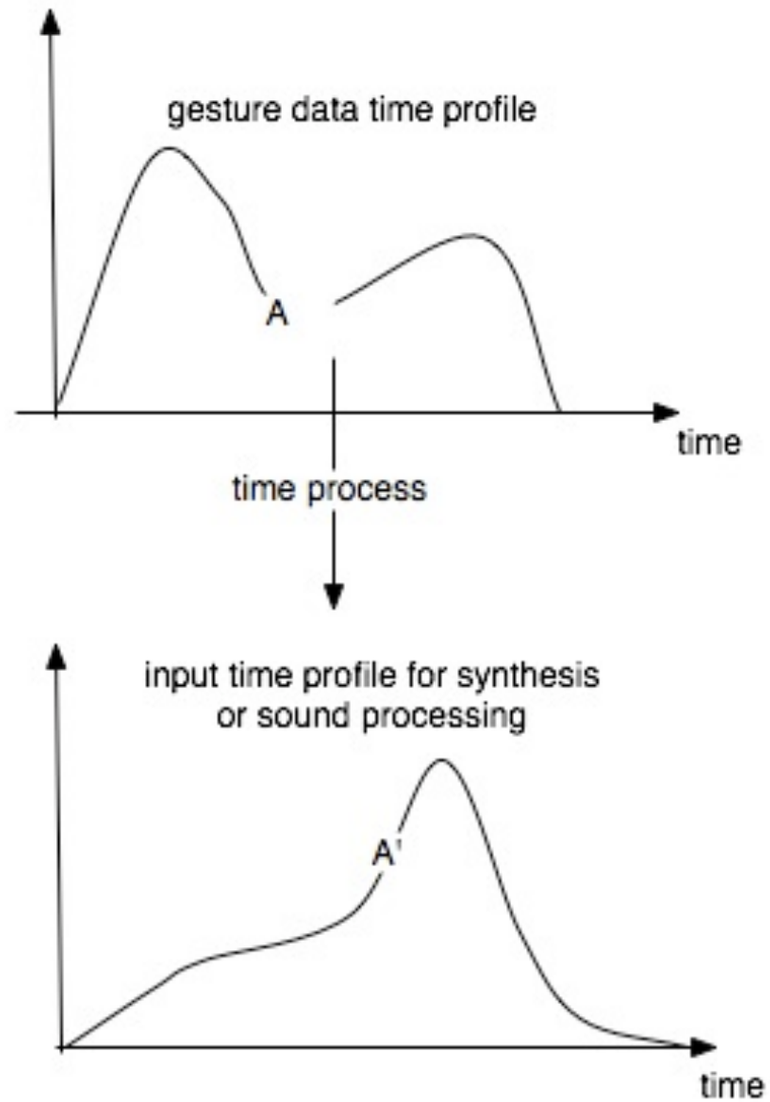
- The multiparametric interface allowed people to think gesturally, or to mentally rehearse sounds as shapes.

- The majority of users felt that the multiparametric interface had the most long-term potential.

- Several users reported that the multiparametric interface was “fun”.

Mapping

- Spatial vs Temporal :
  - « Spatial » : relationship independent of the temporal ordering of data
  - « Temporel » : relationship between temporal processes



Indirect Mapping using Machine Learning Techniques

- Neural Network
  - Mostly static postures

- Principal Component Analysis
  - Data dimension reduction

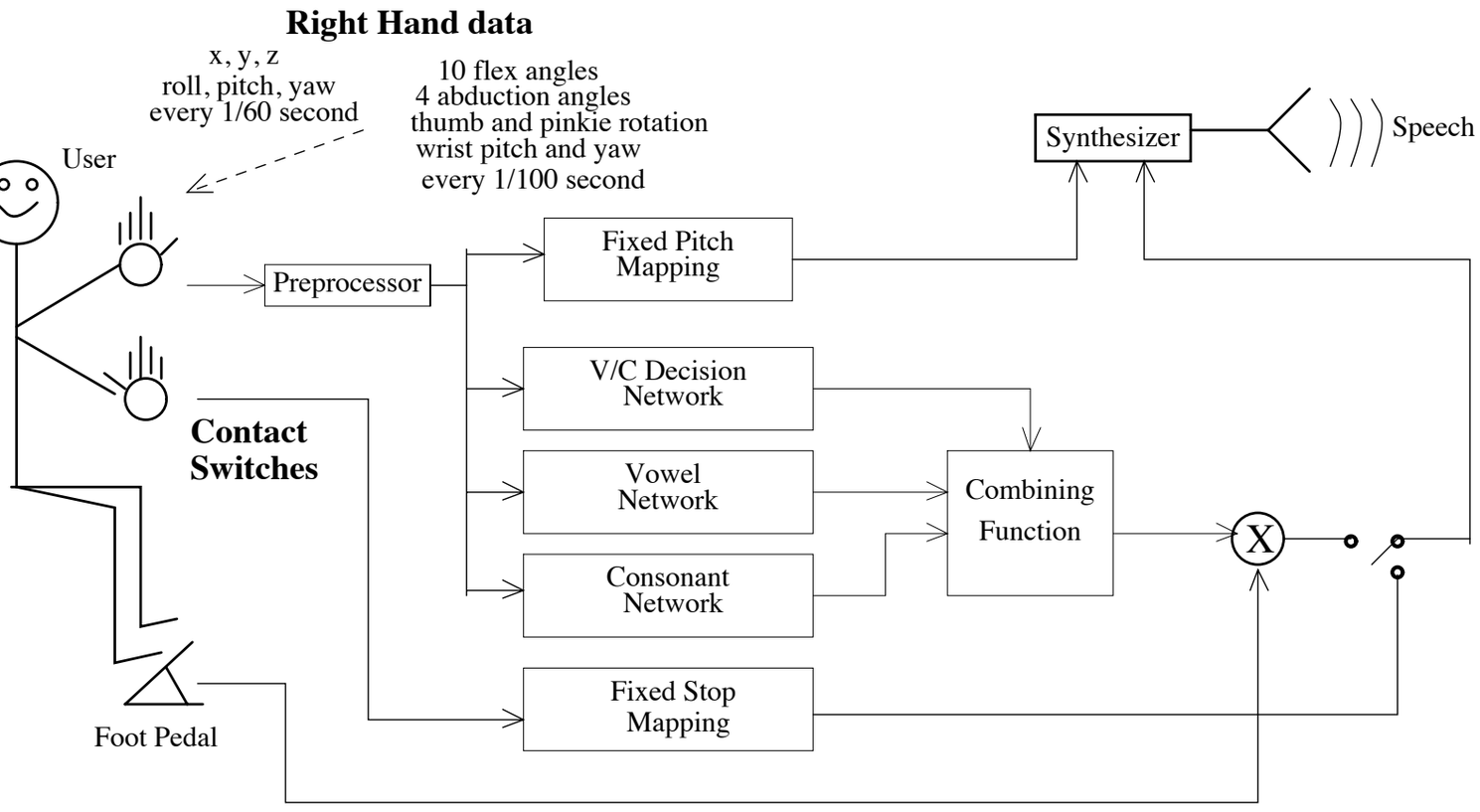
- Finite State Machine
  - Modeling sequences of postures

- DTW, HMM methods
  - Recognition of temporal profiles

Synchronization and recognition

Sydney Fels : Glove-TalkII

- adaptive Interface that Maps Hand Gestures to Speech
- using neural network



Fels, S. S. and Hinton, G. E. Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. IEEE Trans. On Neural Networks, vol. 4, No. 1, 1993.

Conducting gestures

- Several works on conducting gestures
  - Study of professional conducting gesture
  - Beat detections, tempo, anticipation
  - Public Installation
  - Music Pedagogy

Figure 1: ***Maestro!*****, an interactive conducting exhibit for children that we developed, at the Betty Brinn Children's Museum in Milwaukee, USA.** ***Photo appears courtesy of the Betty Brinn Children's Museum in Milwaukee, WI, USA.***

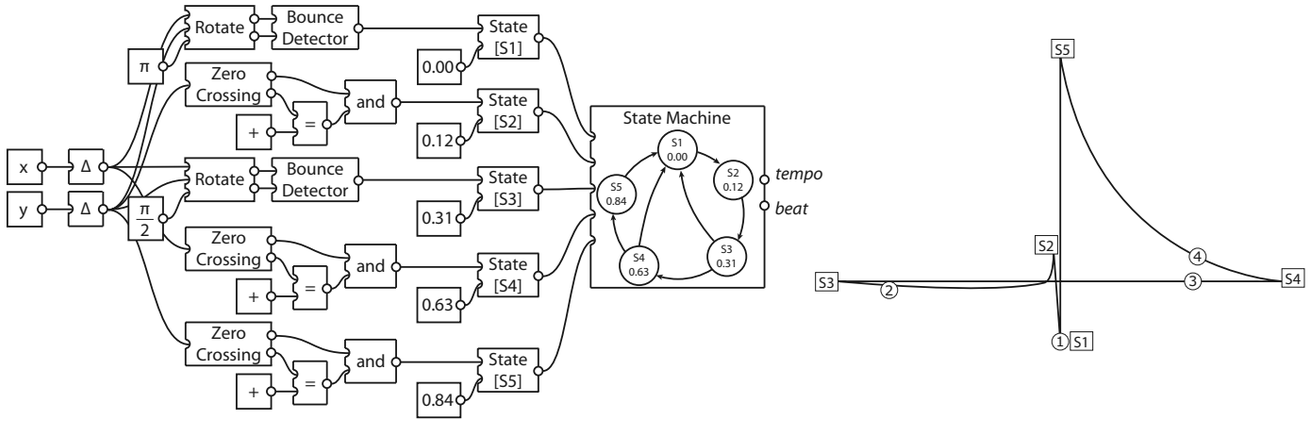
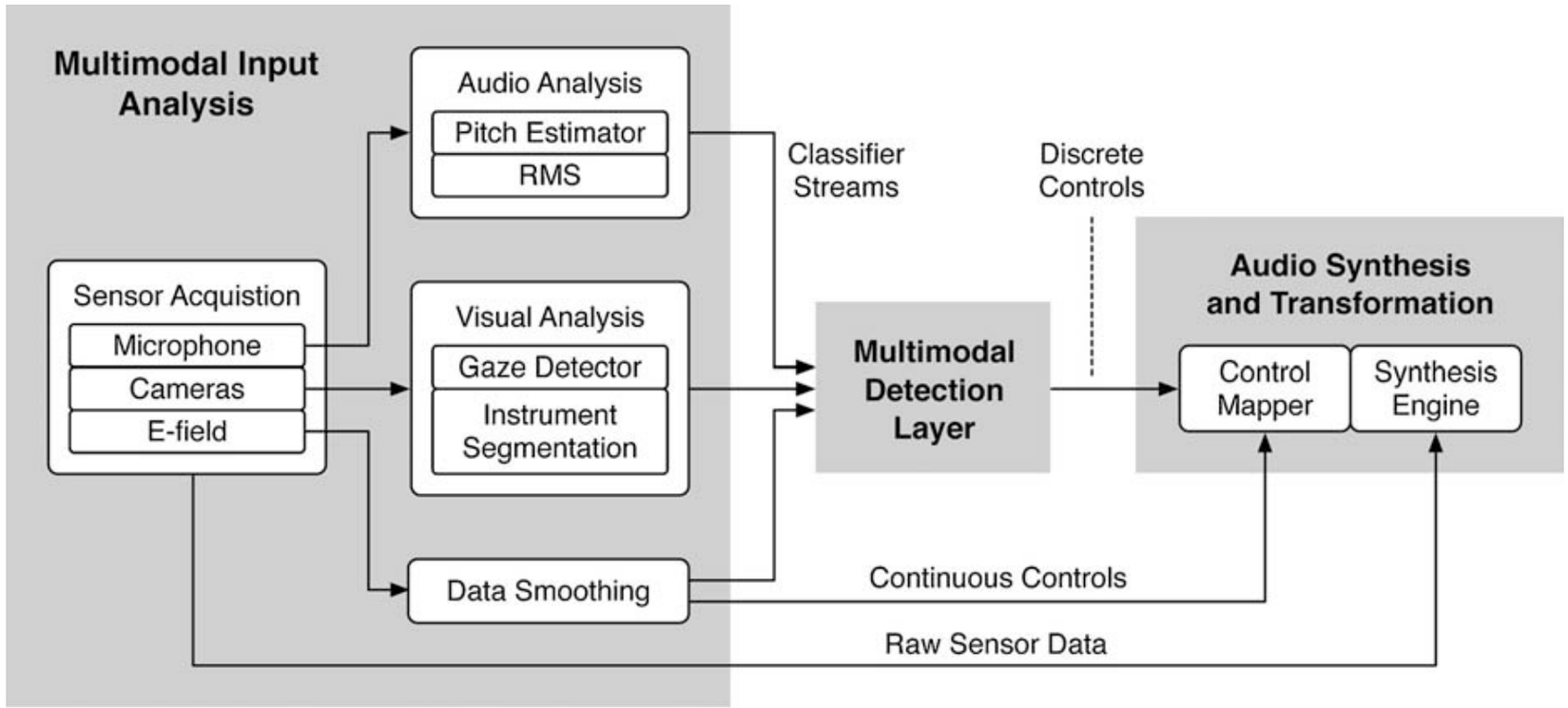


Figure 6: The left figure shows the *conga* graph for the Four-Beat Neutral-Legato gesture profile. Five features are detected, which are used to trigger the progress of a state machine that also acts as a beat predictor. The input to the state machine is the current progress (0 to 1) of the baton as it moves through one complete cycle of the gesture, starting at the first beat. The right figure shows the corresponding beat pattern that is tracked; numbered circles indicate beats, squared labels indicate the features that are tracked and the state that they correspond to.

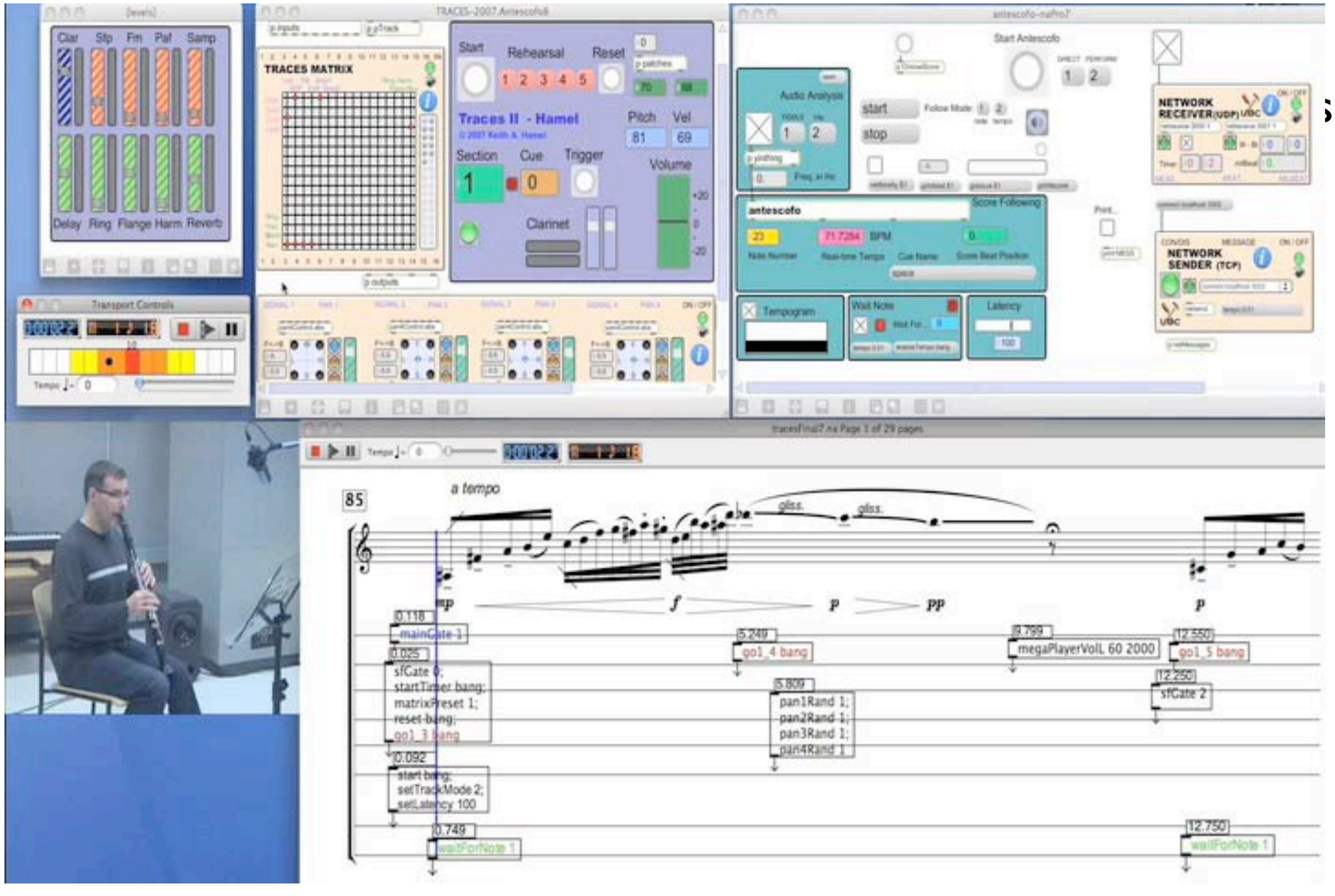
E. Lee, I. Grüll, H. Kiel, and J. Borchers. conga: a framework for adaptive conducting gesture analysis. In *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, pages 260–265, Paris, France,

“Multimodal Music Stand”



Overholt, D., Thompson, J., Putnam, L., Bell, B., Kleban, J., Sturm, B., and Kuchera-Morin, J. 2009. A multimodal system for gesture recognition in interactive music performance. *Comput. Music J.* 33, 4 (Dec. 2009), 69-82

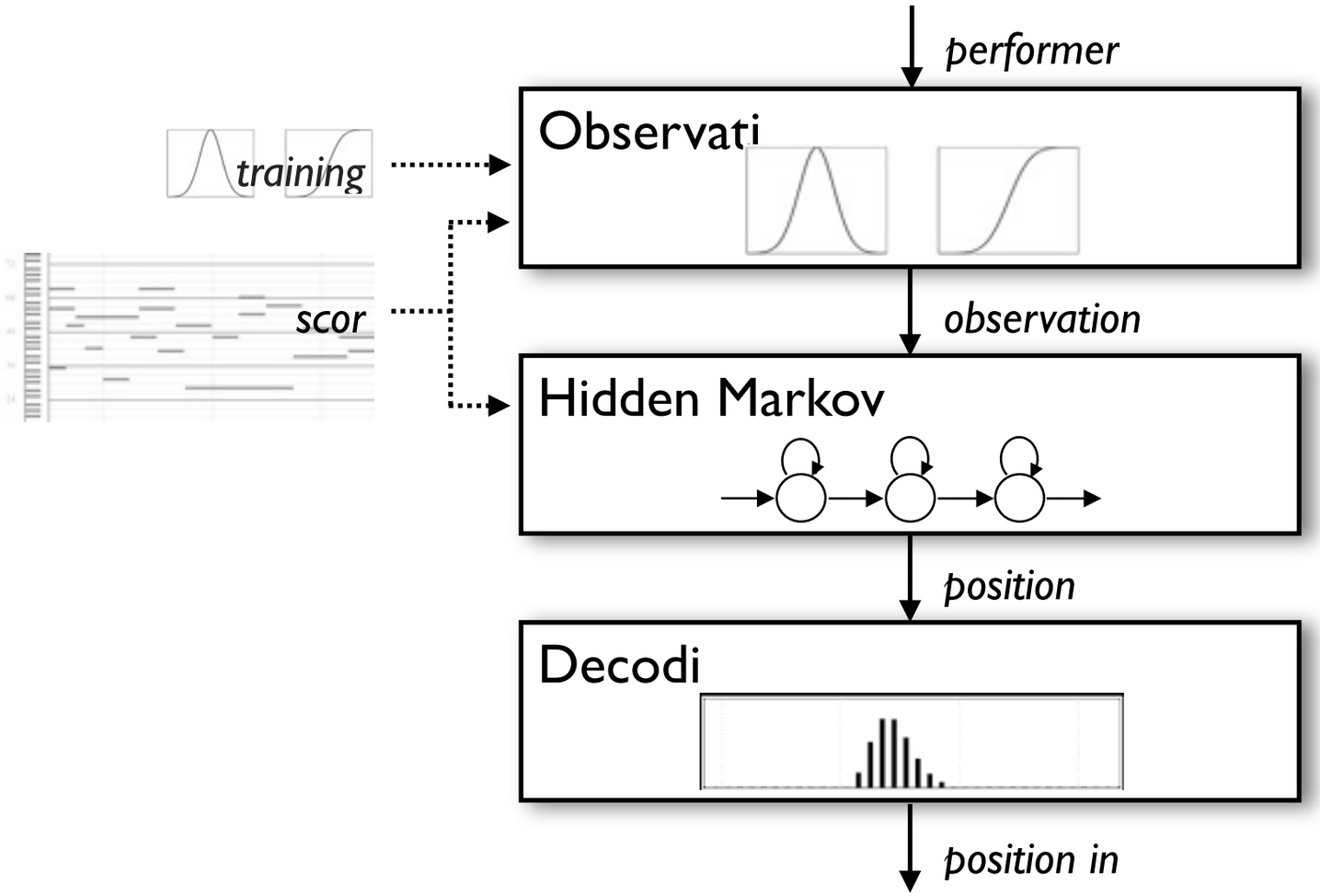
Score Following - Antescofo~



A. Cont « ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music », International Computer Music Conference, North Irland, 2008

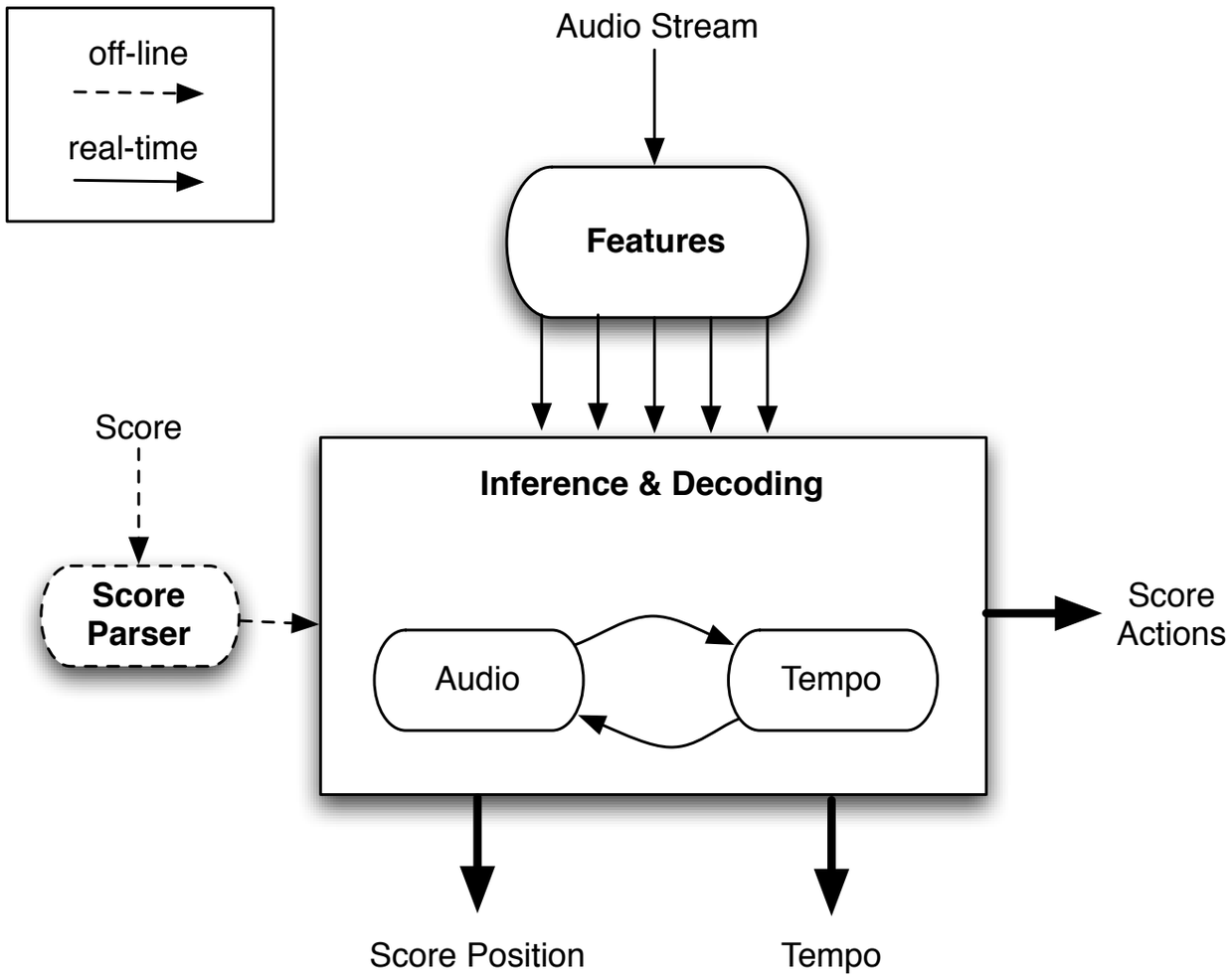
Score following

- For Score Following References: http://cosmal.ucsd.edu/arshia/index.php?n=Main.Scofobib
- http://imtr.ircam.fr/imtr/Score_Following_History
- Best systems use Markov/Semi-Markov modelling of musical events



Score following

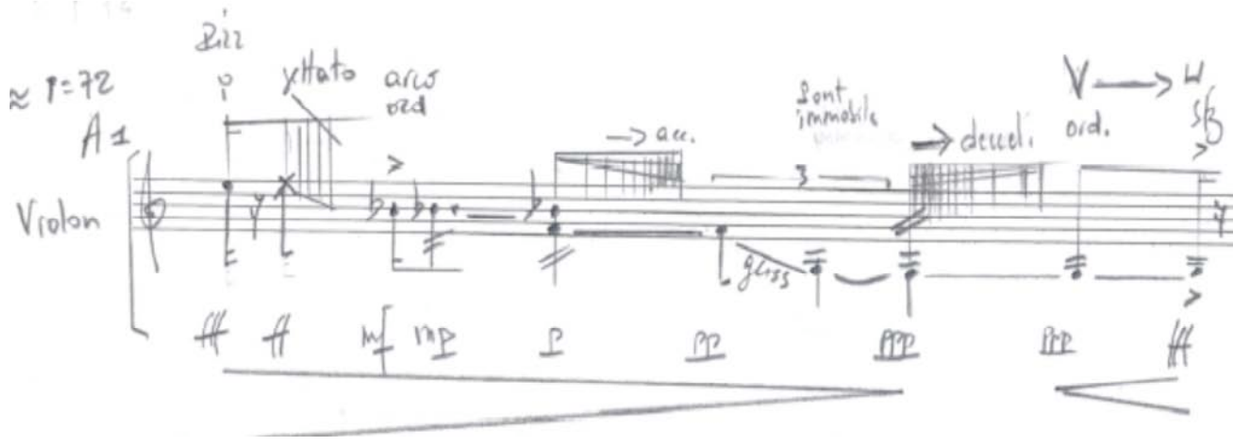
- Antescofo (Anticipatory Score Follower)



Arshia Cont. A coupled duration-focused architecture for realtime music to score alignment, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009 (in press).

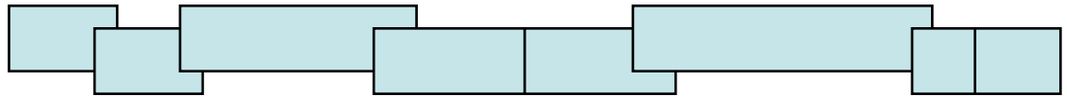
Score Following / Gesture Follower

score



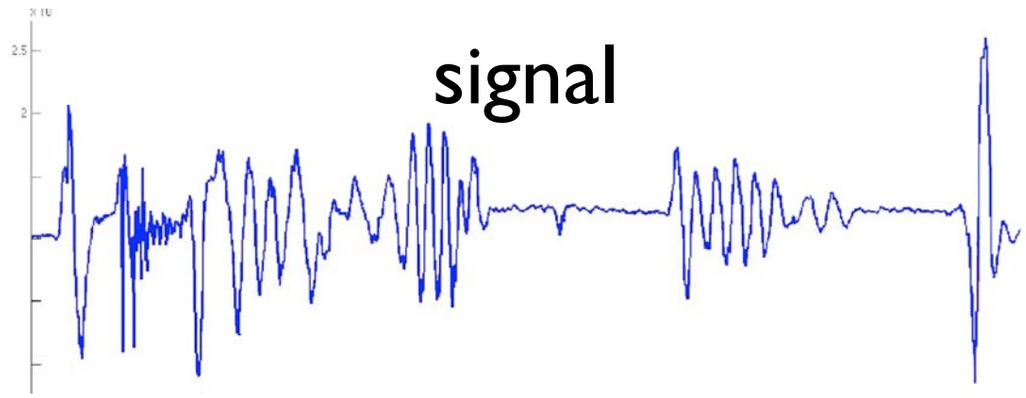
Modeling (HMM)

symbols



Modeling (HMM)

signal



gesture follower @ Ircam

http://imtr.ircam.fr/imtr/Gesture_Follower

- Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., Rasamimanana, N. « Continuous realtime gesture following and recognition », accepted in Lecture Notes in Computer Science (LNCS), Gesture in Embodied Communication and Human-Computer Interaction, Springer Verlag. 2009
- F. Bevilacqua, F. Guédy, N. Schnell, E. Fléty, N. Leroy, " Wireless sensor interface and gesture-follower for music pedagogy", Proc. of the International Conference of New Interfaces for Musical Expression (NIME 07), p 124-129, 2007

Goals

- *Hyp:* Gesture « *meaning* » is in *temporal* evolutions

- Real-time gesture analysis :
  - *gesture following*: **time progression** of the performed gesture
  - *recognition/characterization:* **similarity** of the performed gesture to prerecorded gestures

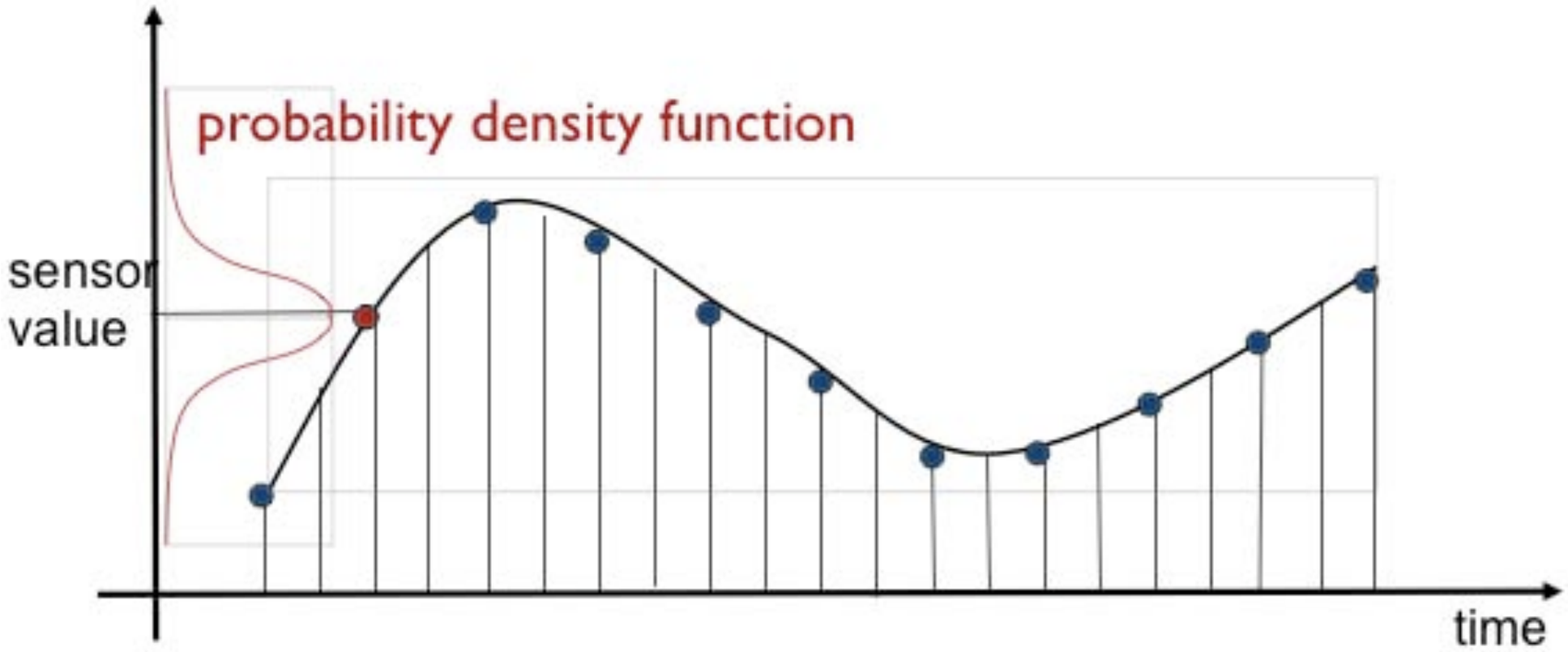
- Requirements
  - simple learning procedure, with a single example
  - adaptation to the user idiosyncrasies
  - continuous analysis from the beginning of the gestures

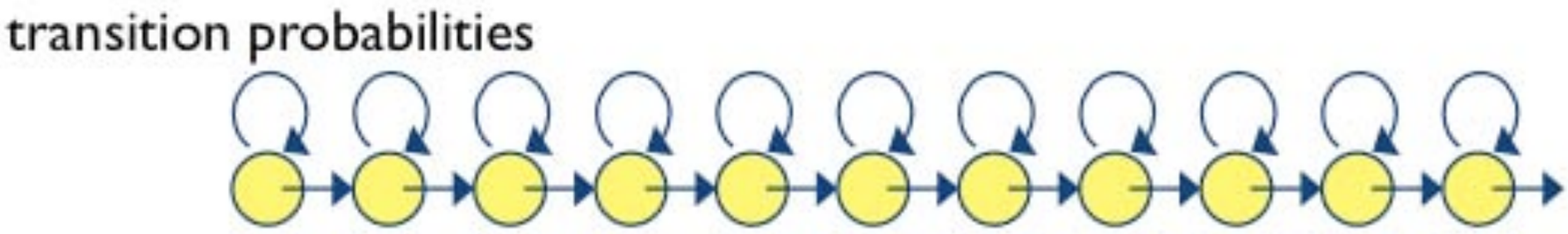
Gesture ?

- Any continuous datastream of parameters
- typically 0.1 to 1000 Hz
- from motion capture systems:
  - image descriptors
  - accelerometers, gyroscope, magnetometers
- from sound descriptors
  - pitch, loudness
  - mfccs, ...
- multimodal data

Time Profile Modeling: HMM

Markov Models

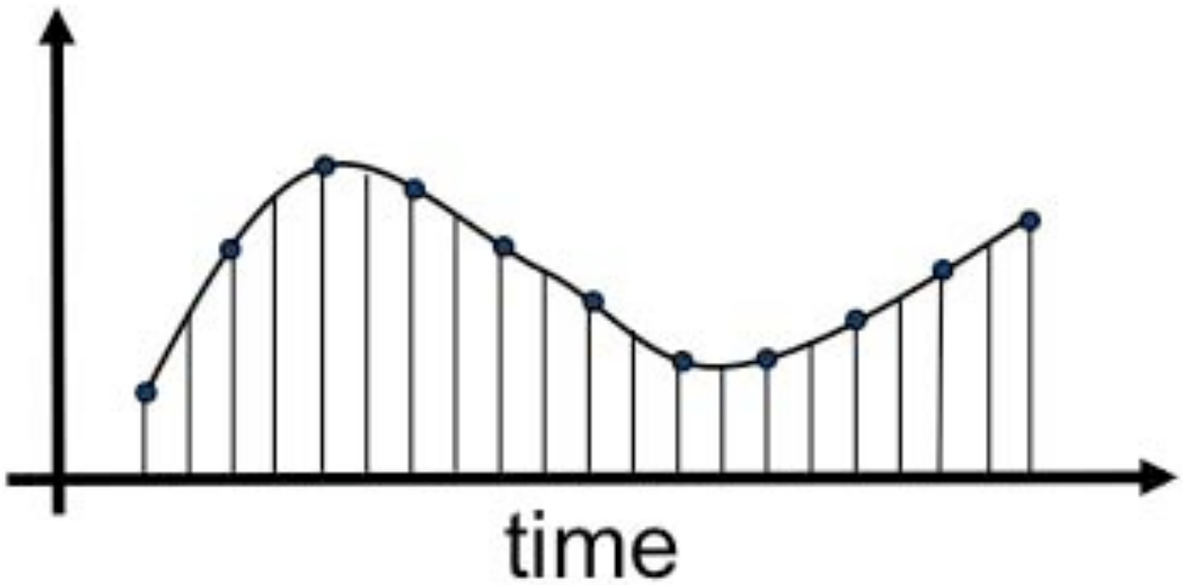




Markov Chains

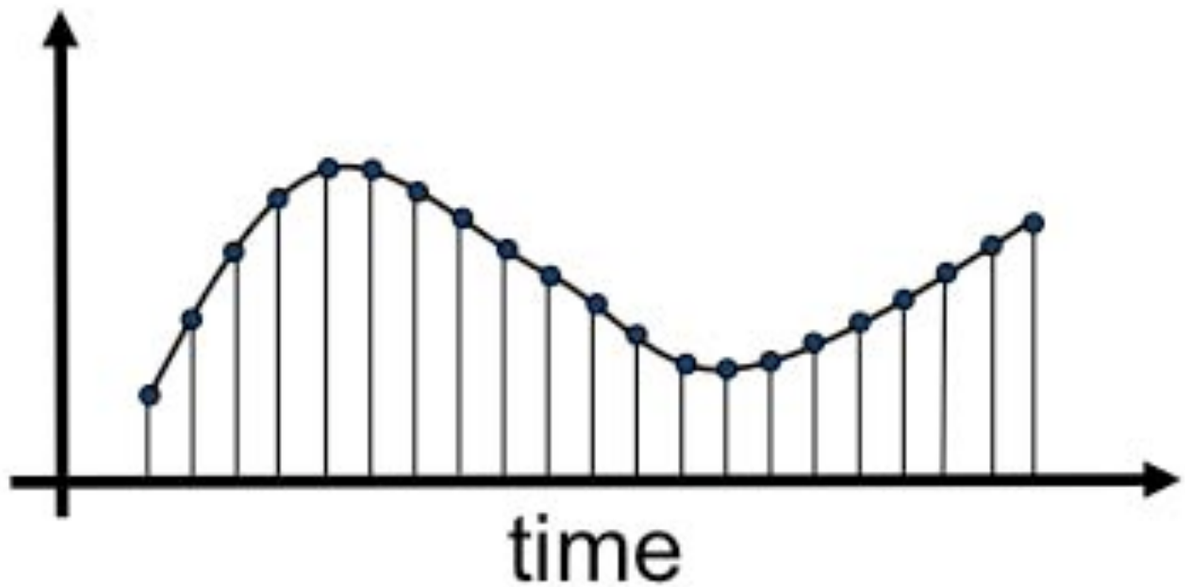
HMM structures

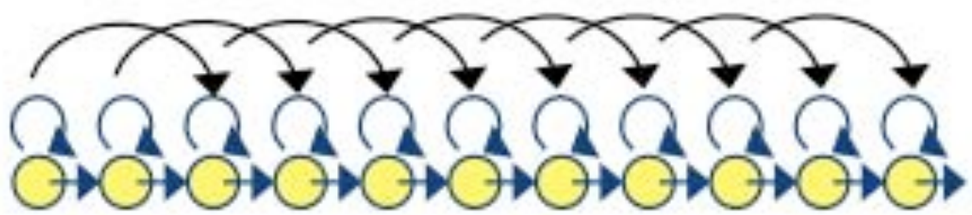
one state every
two samples



maximum relative speed = 2

one state
every sample





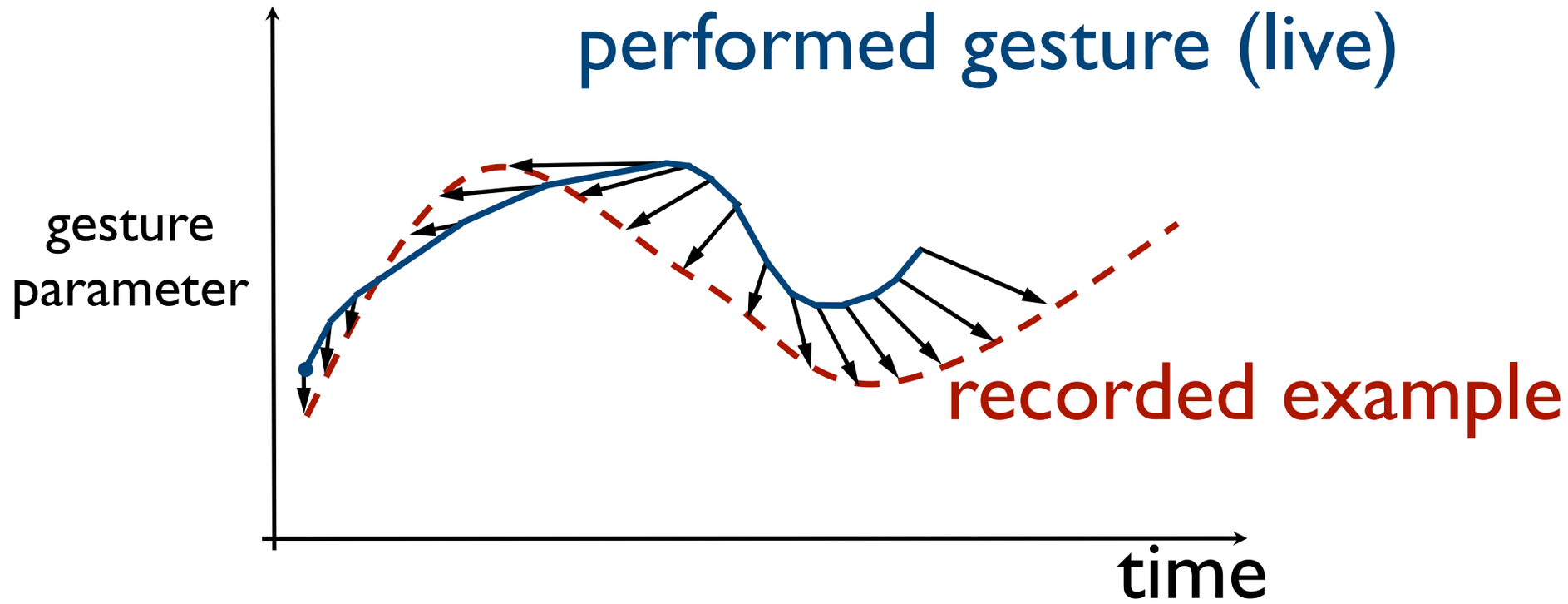
maximum relative speed = 2

Hybrid Approach

- Hybrid between:
  - Template based - Dynamic Time Warping
  - Linear Dynamics Model
  - HMM

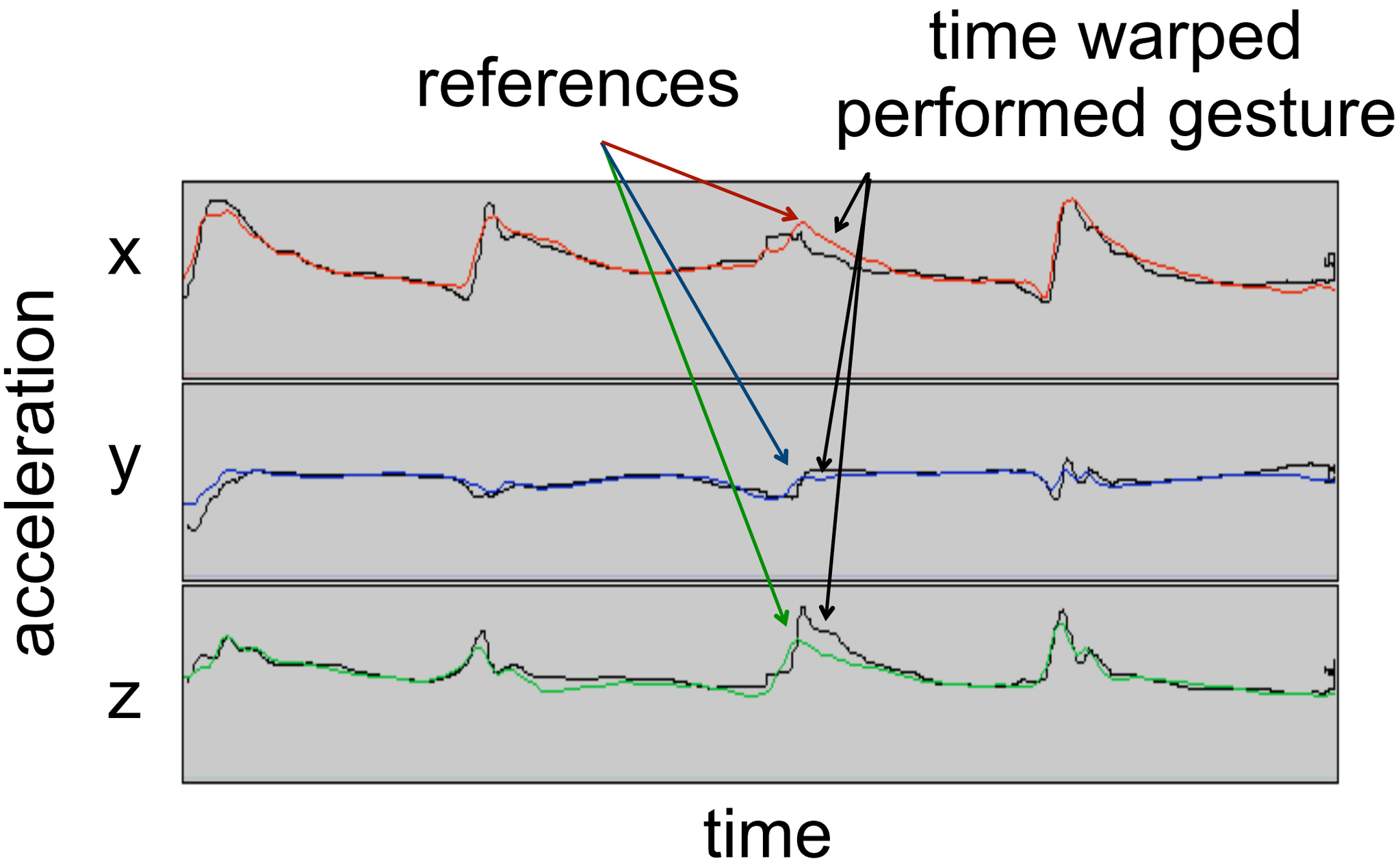
- Similar to S. Rajko et al. (ASU), also developed in an artistic context
  - G. Qian, T. Ingalls and J. James, Real-time Gesture Recognition with Minimal Training Requirements and On-line Learning, to appear in IEEE Conference on Computer Vision and Pattern Recognition, 2007.
  - S. Rajko and G. Qian, A Hybrid HMM/DPA Adaptive Gesture Recognition Method, ISVC 2005,p 227-234.

Real-time time warping



- Synchronization/following
- Recognition
- Anticipation (prediction)

Time warping



Learning phase

- Transition matrix
  - left-to-right Markov chain
  - states regularly spaced in time

    ⇒ transition matrix set by the sampling rate

    ⇒ direct relationship between state number i and time

    (T= 1/1-a, where a is the self transition prob)

- Emission probabilities

$$e^{-\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2}$$

values from the time profile

calculated or set by user

Forward Calculation

State probability for given observation $O(t_n) = b$

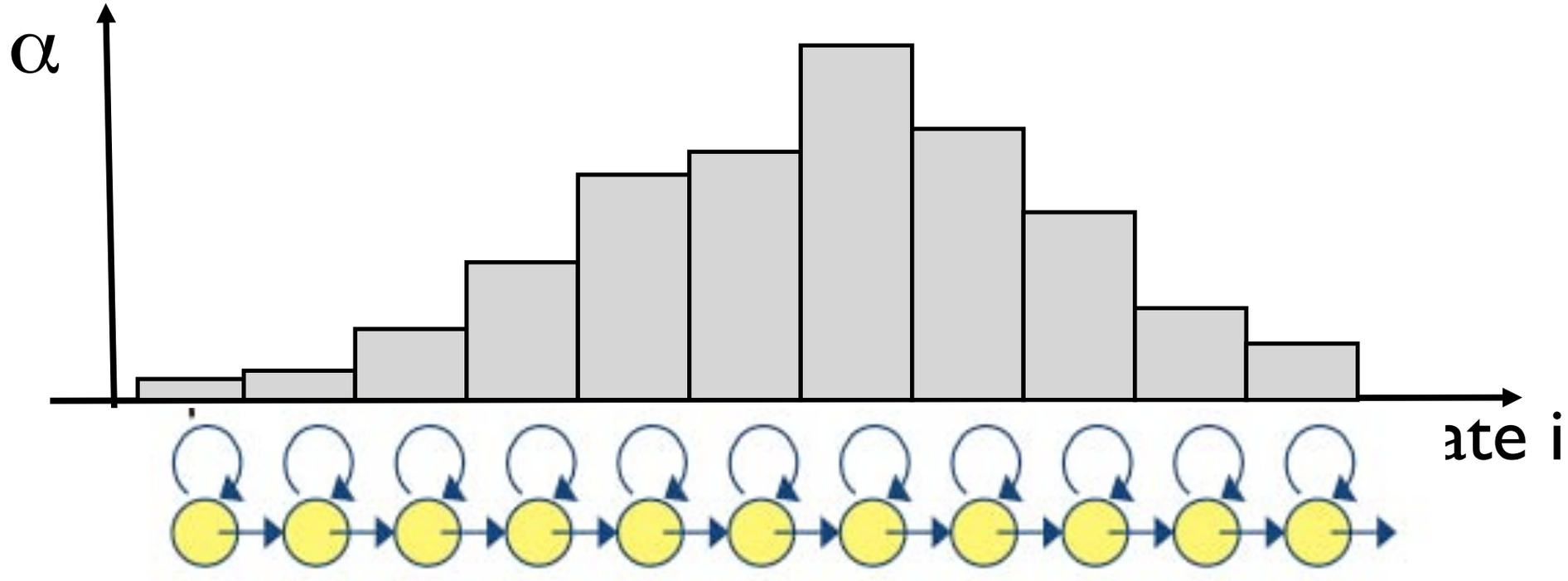
Transition Matrix

$$\alpha(t_{n+1}) = \overset{\searrow}{A}\left[\alpha(t_n) \cdot \overset{\vee}{b}\right]$$

state probability at $t = t_{n+1}$

Decoding phase

- Using the *forward computation* [Rabiner 89] (causal !)
- Compute the probability α of being at state i

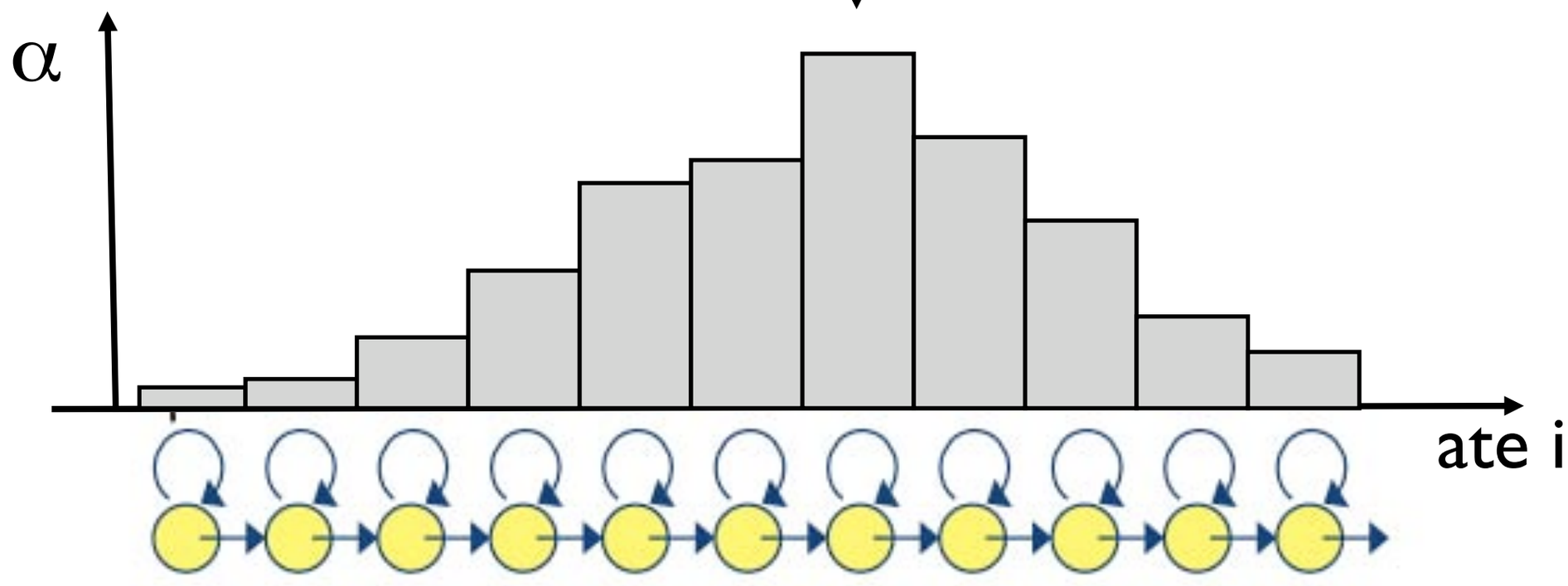


Decoding phase

State with maximum probability at time *t*

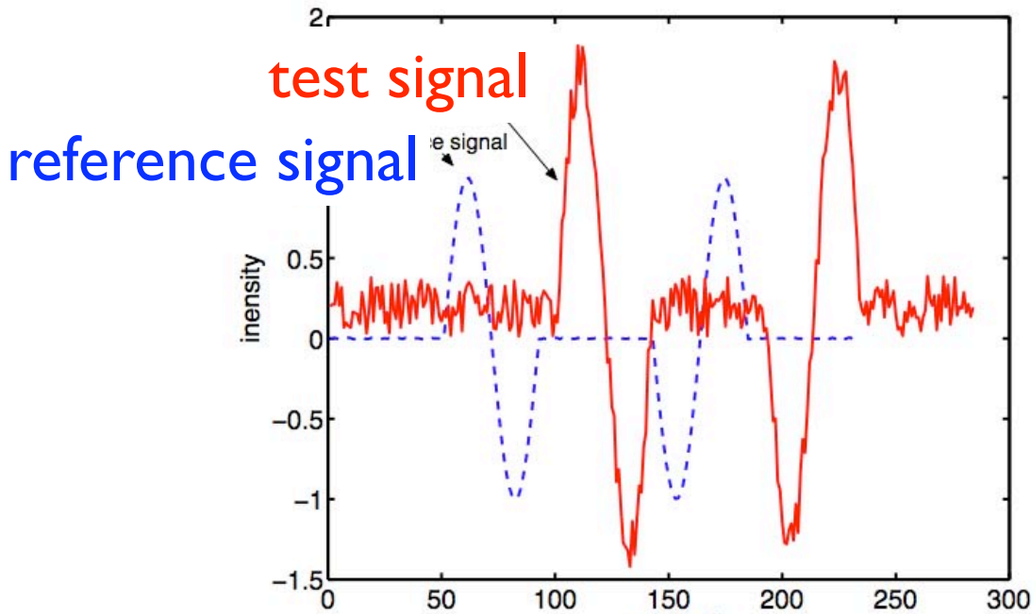
→ time progression

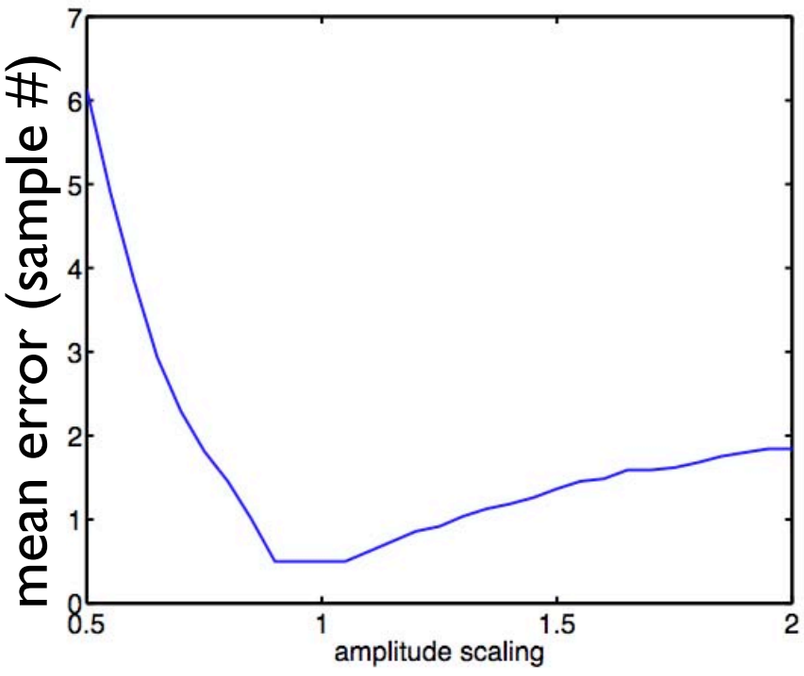
↓



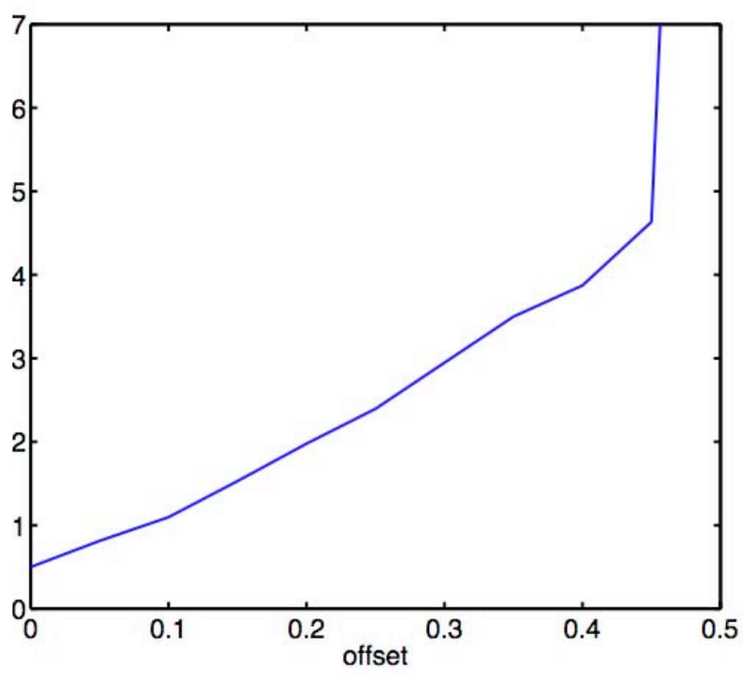
$\sum \alpha_i$ = likelihood at time *t*

Evaluation with synthesized signals

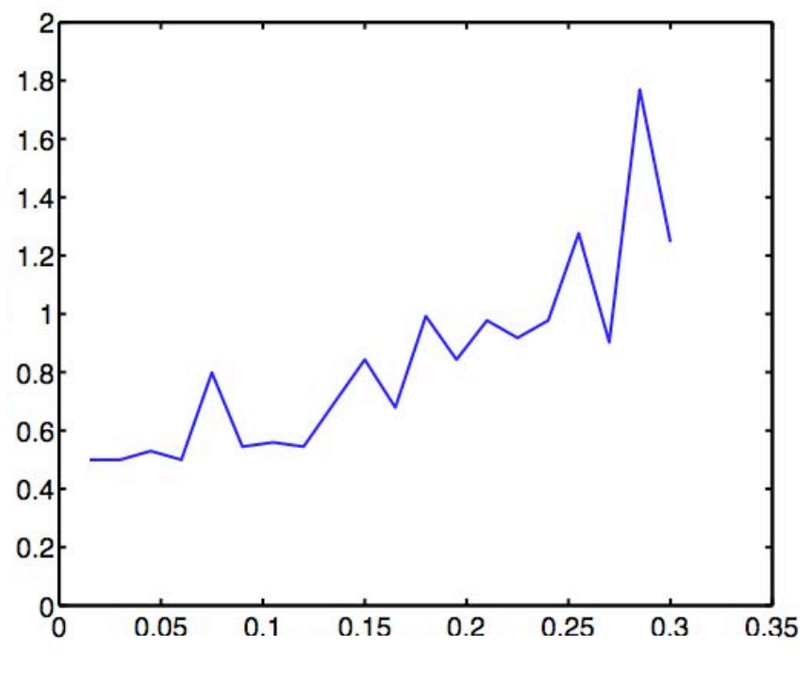




scaling

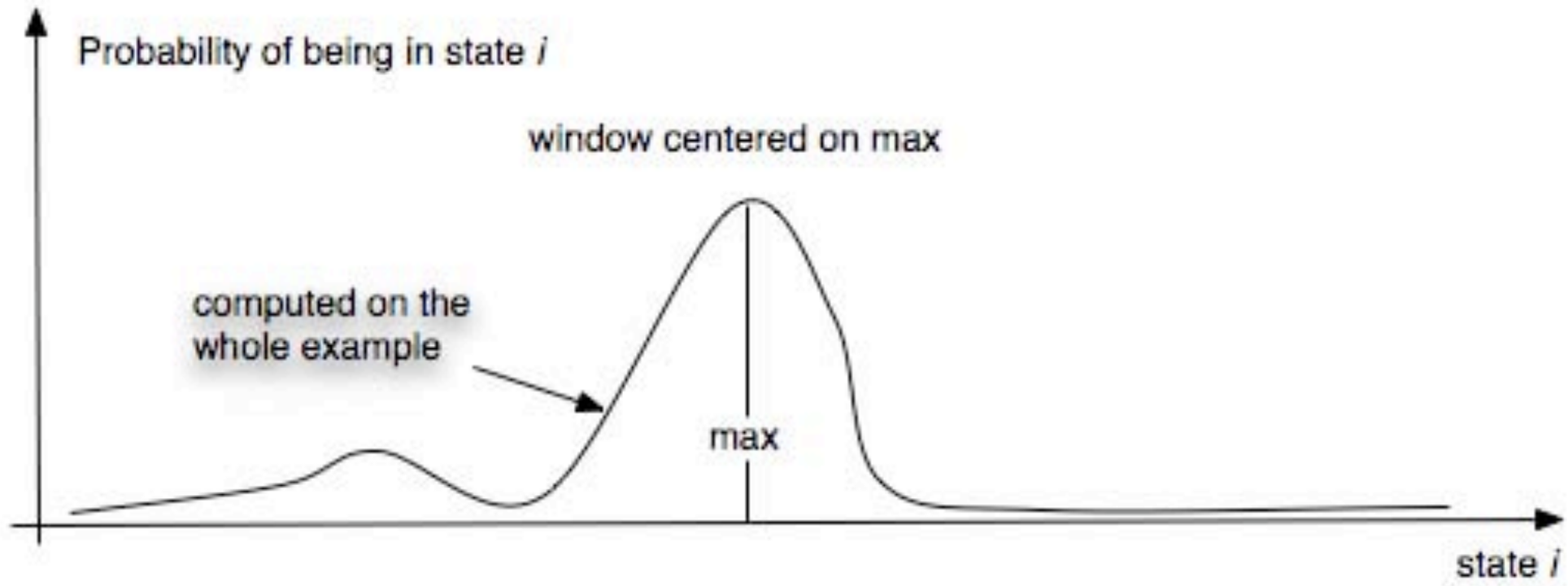


offset

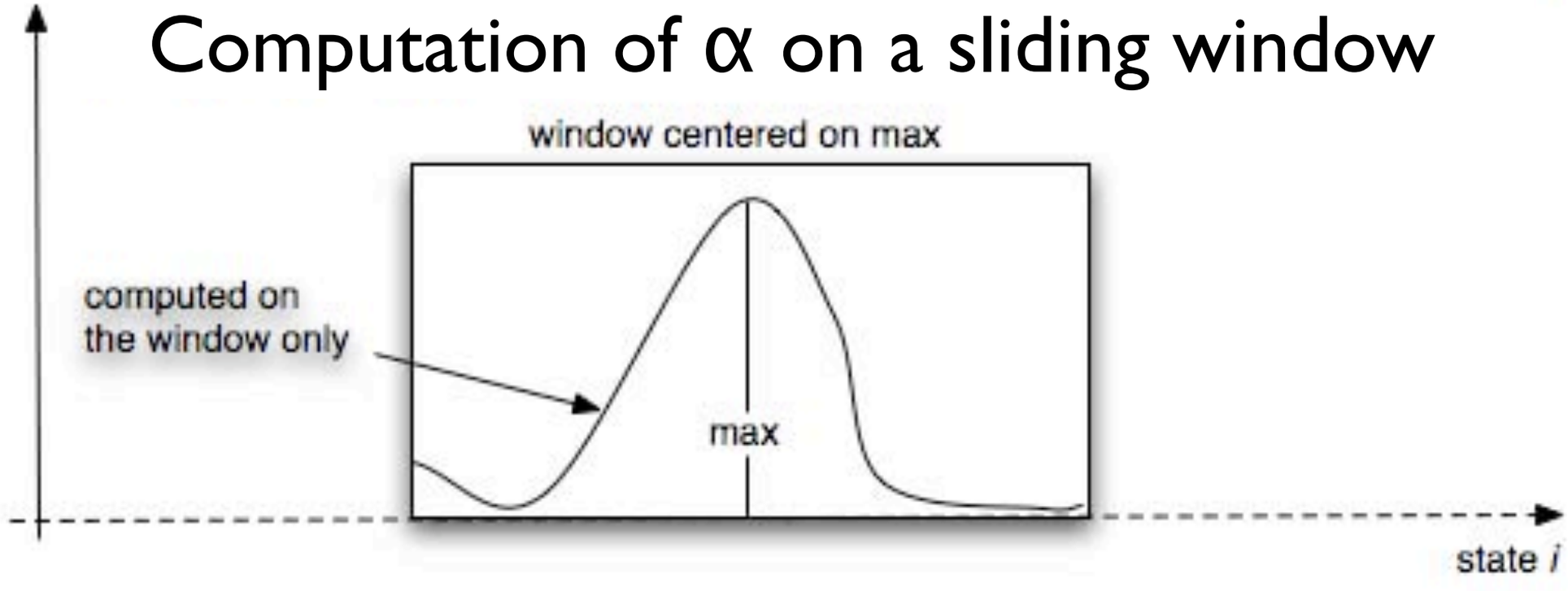


noise

Following long sequences



Computation of α on a sliding window



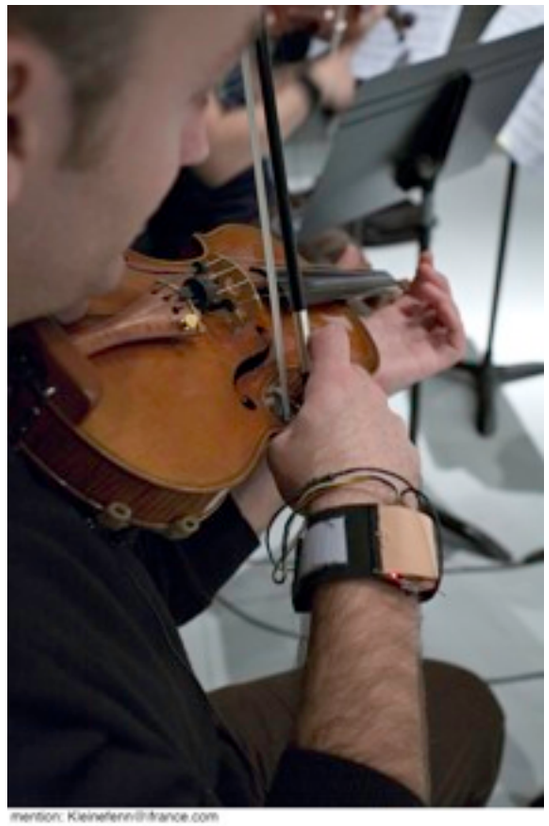
Gesture Follower - Context

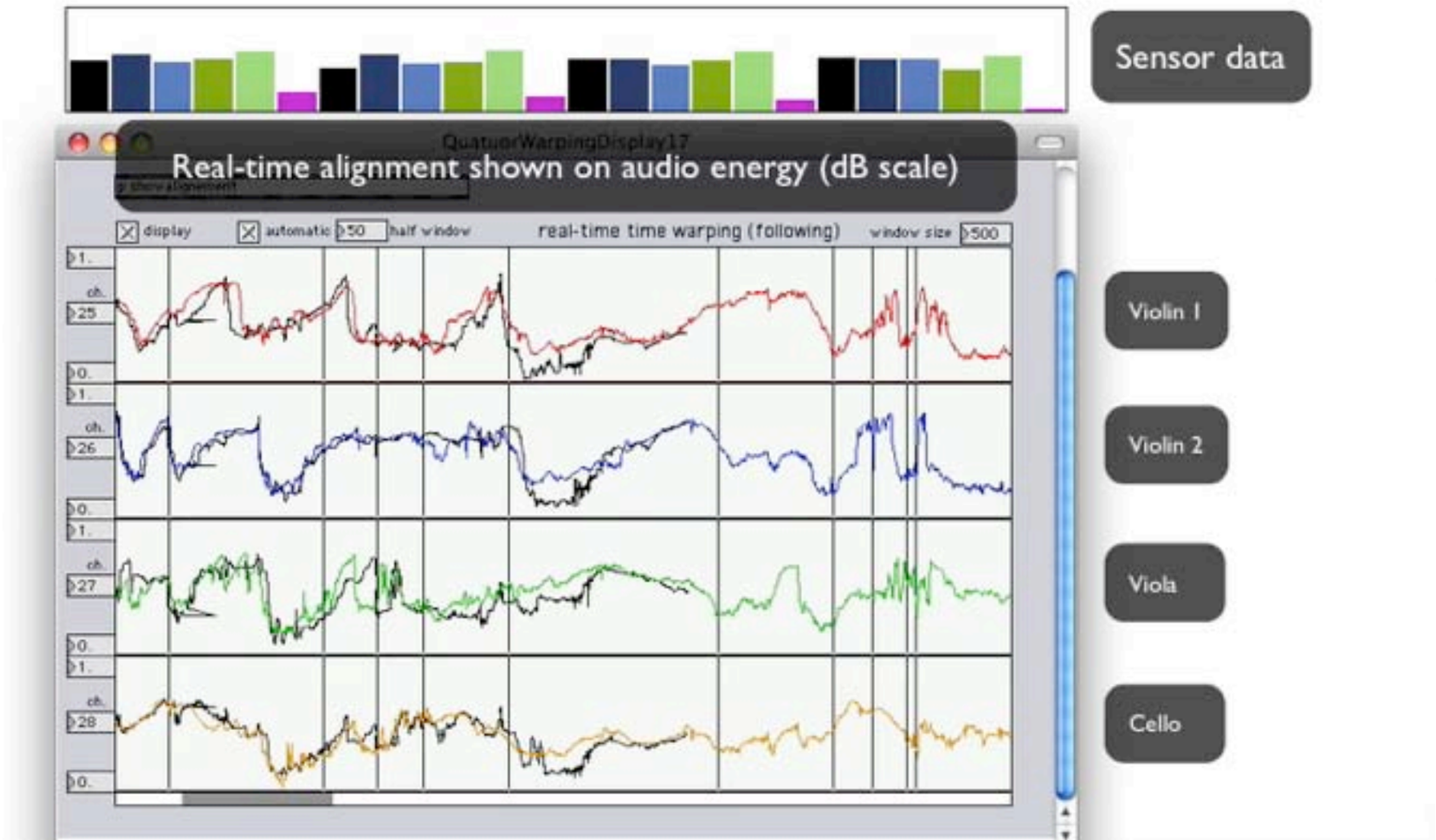
dance
(performance
and installation)

music pedagogy

music performance

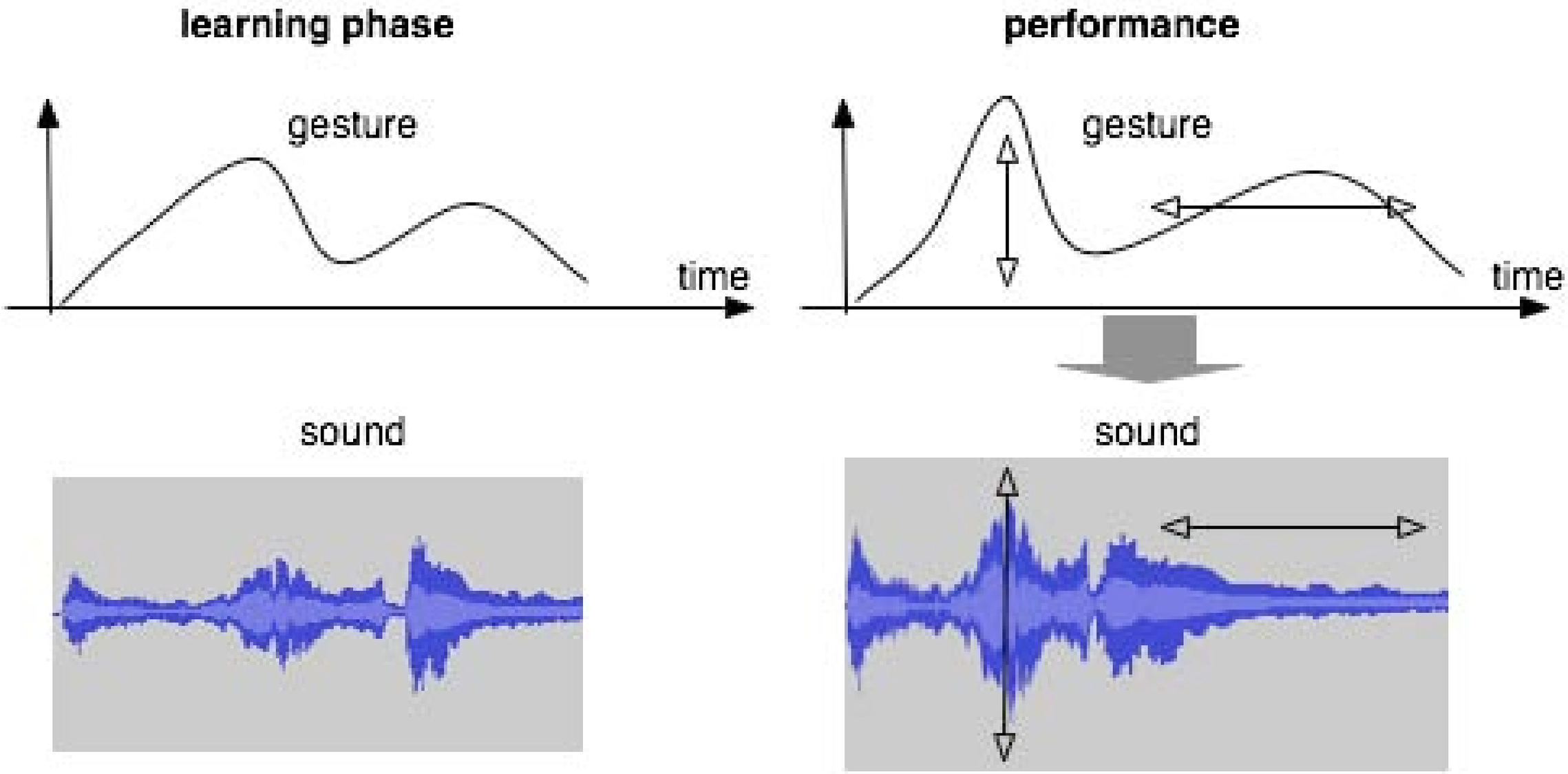
StreicherKreis - Florence Baschet





gesture" =
acceleration
angular velocity
pressure
audio energy

Synchronizing Sound to Gesture



Music Pedagogy applications

- Conducting

Atelier des Feuillantines Fabrice Guédy

Homo Ludens (Richard Siegal - The Bakery)

2:55

Recognizing movement qualities

Sarah Fdili Alaoui (PhD work)
Collaboration with the dance company Emio Greco I PC

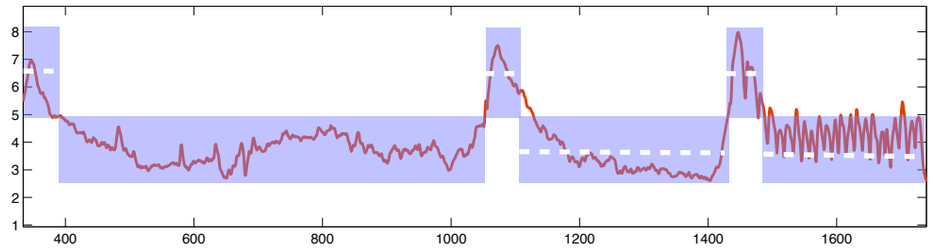
Towards Segmental Models

Goal: classification / segmentation of sounds and gestures based on their temporal evolutions

Approach: segmental HMM models

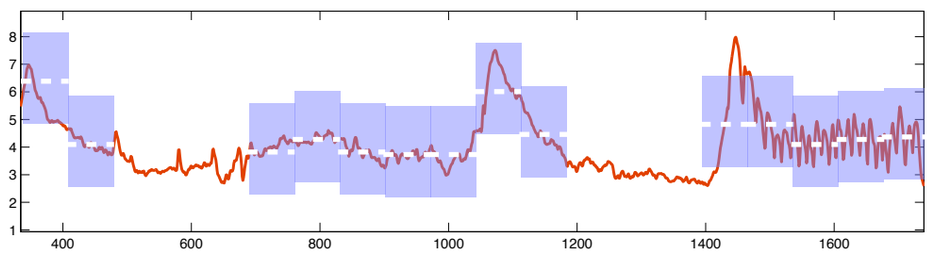
"classical" HMM

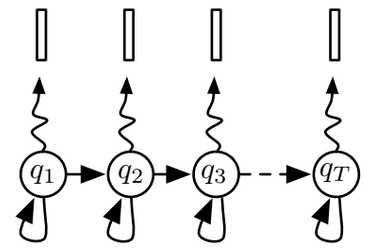
steps



Gesture Follower

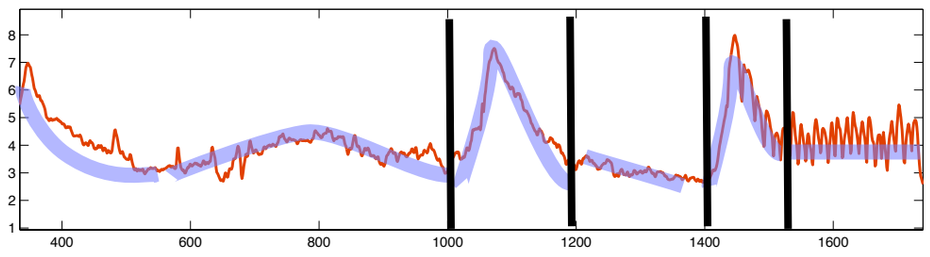
steps

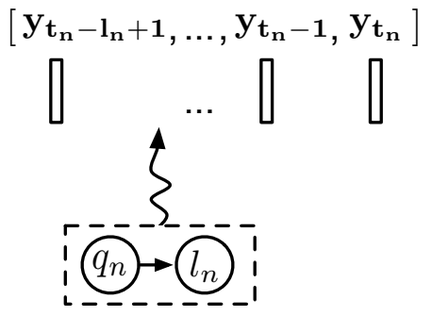




segmental HMM

trajectories



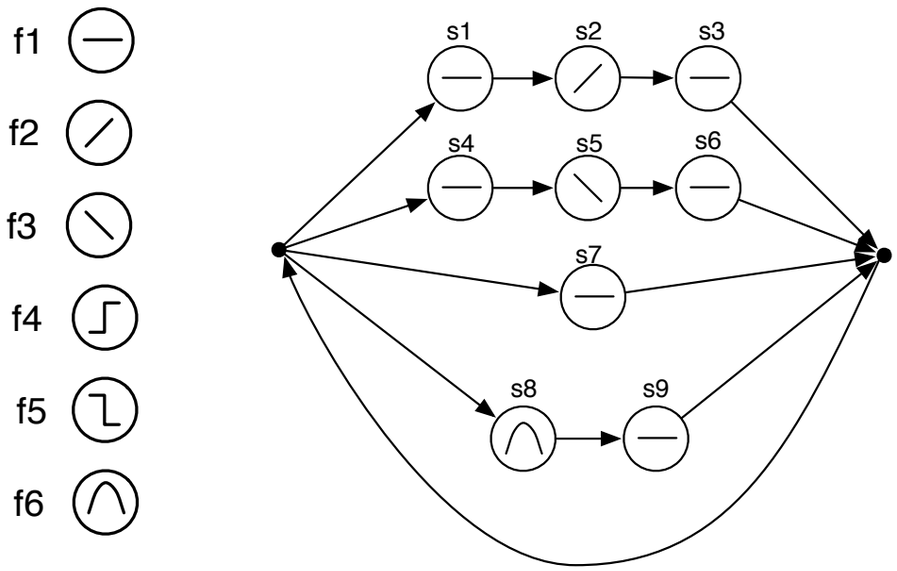


Sound and gesture morphologies

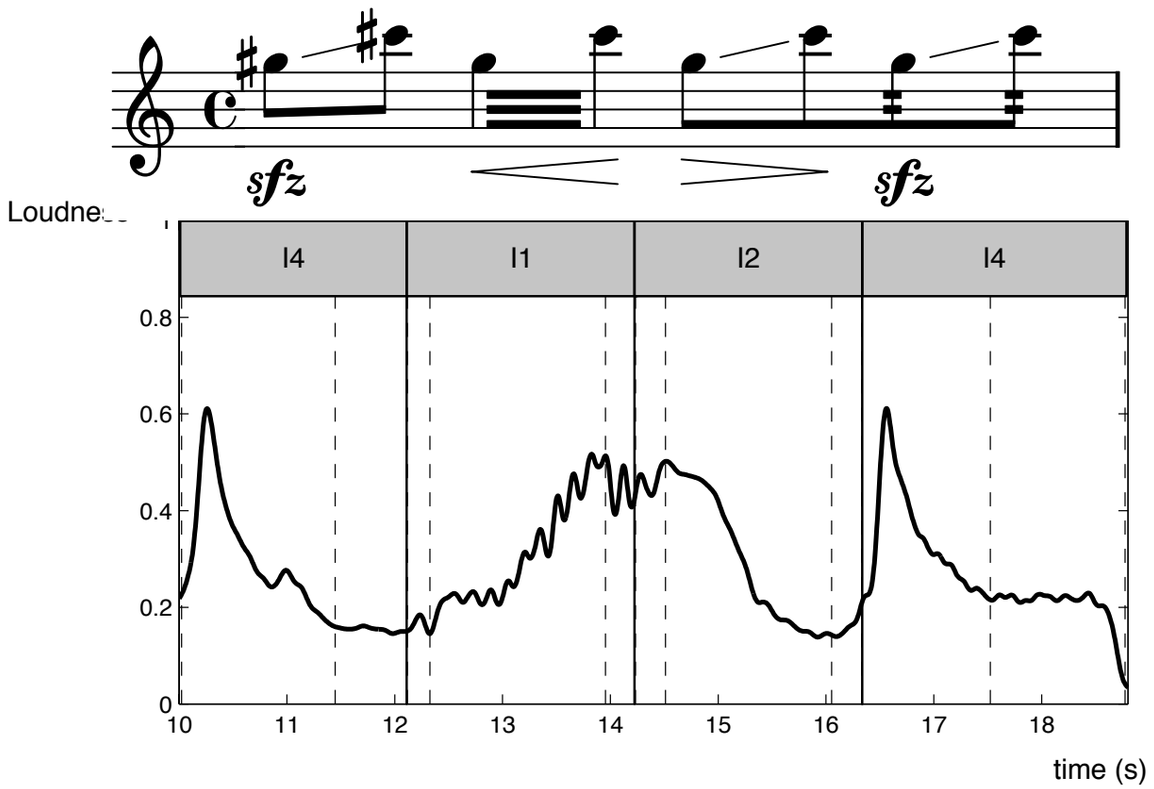
classification/segmentation on a violin database (pitch/loudness profiles)

PhD Julien Bloit & Projet Interlude

Modelling by primitive assembling:



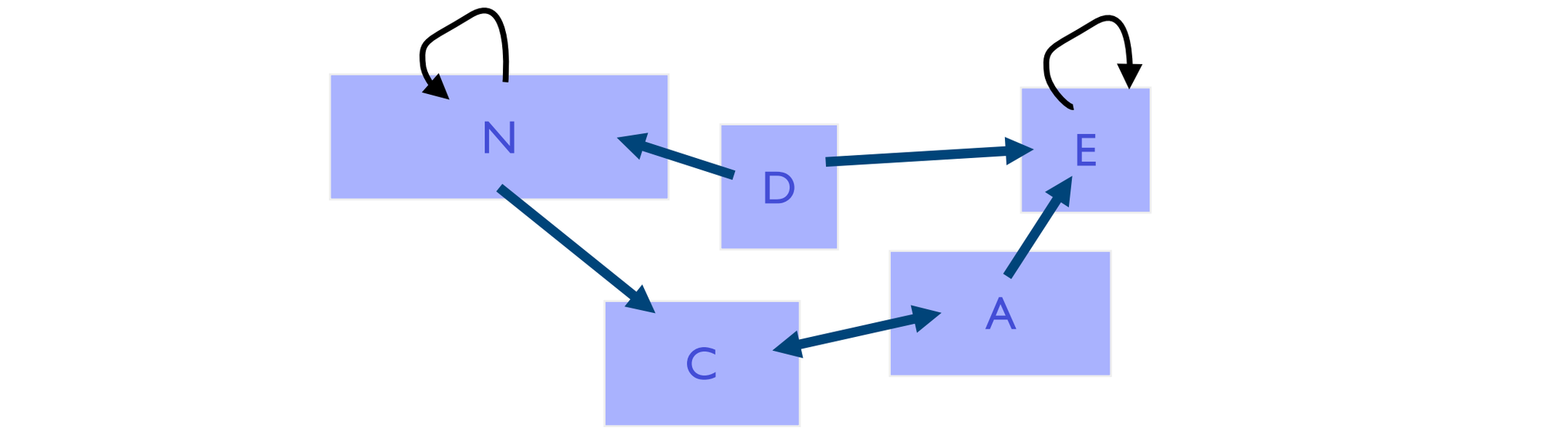
Segmentation on a continuous stream:

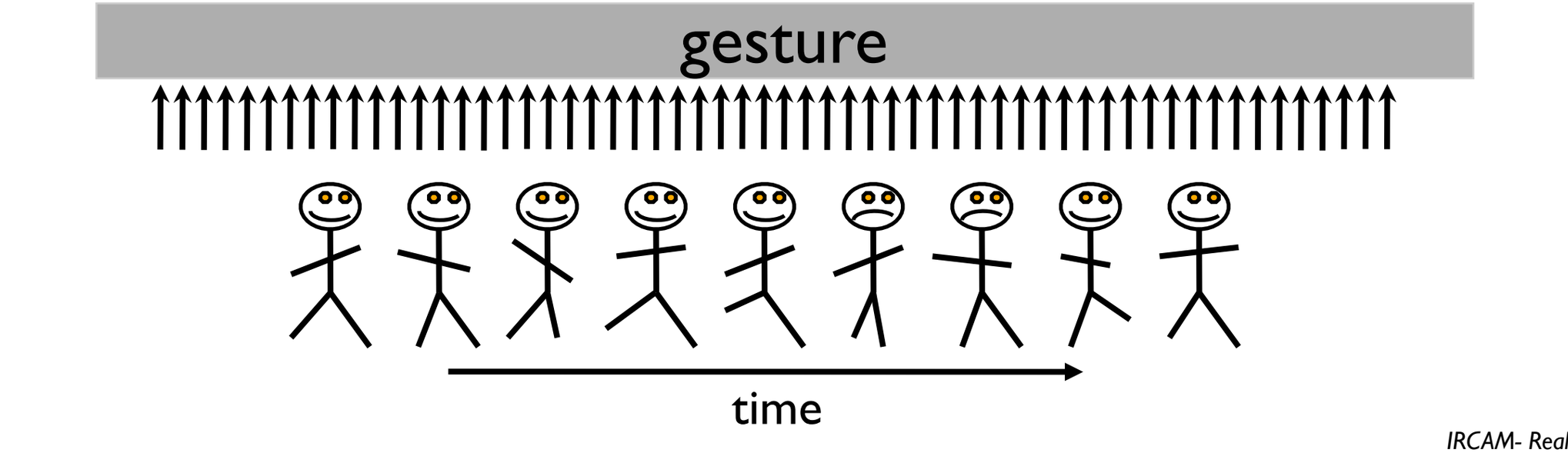


[1] J. Bloit, N. Rasamimanana, and F. Bevilacqua. Modeling and segmentation of audio descriptor profiles with segmental models. Pattern Recognition Letters, 2009.
[2] J. Bloit, N. Rasamimanana, and F. Bevilacqua. Towards morphological sound description using segmental models. In DAFX, Como, Italy, 2009.

Hierarchical / Two-level Modeling

1. Temporal Segments Temporal
2. Sequence of Segments





Credits and Acknowledgements

- Real Time Musical Interaction team:

  Frédéric Bevilacqua, Tommaso Bianco, Julien Bloit, Riccardo Borghesi, Baptiste Caramiaux, Arshia Cont, Arnaud Dessein, Sarah Fdili Alaoui, Emmanuel Fléty, Vassilios-Fivos Maniatakos, Norbert Schnell, Diemo Schwarz, Fabrice Guédy, Alain Bonardi, Nicolas Rasamimanana, Bruno Zamborlin

- Current Support of the projects:
  - ANR projects: Interlude, Topophonie (France).
  - EU-ICT Project SAME
- Thanks to

  Atelier les Feuillantines and students, Remy Muller, Jean-Philippe Lambert, Alice Daquet, Anthony Sypniewski, Donald Glowinski, Bertha Bermudez and EG|PC, Myriam Gourfink, Richard Siegal, Hillary Goidell, Florent Berenger, Florence Baschet