Coarticulation in music-related gestures

Rolf Inge Godøy, Alexander Refsum Jensenius, and Kristian Nymoen
Department of Musicology, University of Oslo
e-mail: r.i.godoy@imv.uio.no, a.r.jensenius@imv.uio.no, krisny@ifi.uio.no

Sensing Music-Related Actions

http://www.hf.uio.no/imv/forskning/forskningsprosjekter/musicalactions/

fourMs

Music, Mind, Motion, Machines

www.fourms.uio.no

UNIVERSITY OF OSLO

Coarticulation

- Coarticulation = Fusion of micro-level actions and sounds into meso-level, holistically experienced chunks of actions and sounds, entailing a contextual smearing of the micro-level elements
- One advantage of coarticulation: Can account for the holistic perception, cognition, as well as motor control (anticipatory) of sound-action chunks
- The beauty of coarticulation: May work both forwards and backward in time, i.e. future events are colored by past events and past events are colored by future events

Coarticulation in various domains:

- Everyday tasks, e.g. reaching and lifting
- Animation
- Facial movements
- Fingerspelling
- Handwriting
- Music, but not well studied here
- Much studied in speech (see Hardcastle and Hewlett 1999 for an overview):

Coarticulation in speech

(from: http://person.sol.lu.se/SidneyWood/coart/coartint/coartina.html)

Coarticulating "happy"

For example, suppose you say the word happy:

Before you say anything, you will have moved your tongue into position for a

Then, while you are saying h, it will sound a bit like a

While you are saying a, you will also be closing your lips for pp

While your lips are together for pp, you will be moving your tongue to where you need it for y

Finally, while you are saying y, you will be opening your lips after pp

The whole word will usually be uttered in less than half a second

Principles of coarticulation:

- Otherwise singular events embedded in a context
- Past events influence present events, i.e. position and shape of effectors are determined by recent action
- Future events influence present events, i.e. position and shape of effectors are determined by preparation for future actions (anticipatory movements)
- Seems to be a biomechanical necessity
- Seems to be a motor control necessity, i.e. anticipation in motor control

Principles of coarticulation:

- coarticulation can be seen as an advantageous element: "…it is a blessing for us as behaving organisms. Think about a typist who could move only one finger at a time. Lacking the capacity for finger coarticulation, the person's typing speed would be very slow. Simultaneous movements of the fingers allow for rapid responding, just as concurrent movements of the tongue, lips and velum allow for rapid speech. Coarticulation is an effective method for increasing response speed given that individual effectors (body parts used for movement) may move relatively slowly." (Rosenbaum 1991, 15)

Principles of coarticulation:

- Basically: Body movement tends to be continuous, and also results of actions tend to be continuous (however sometimes very briefly)
- Can in some cases also be understood as a mass-spring phenomenon, i.e. as overlapping resonating events
- Has consequences for perception
- Contextual smearing in sound
- Contextual smearing in movement
- High-speed video gives and intuitive impression:

Some studies of coarticulation in sound production:

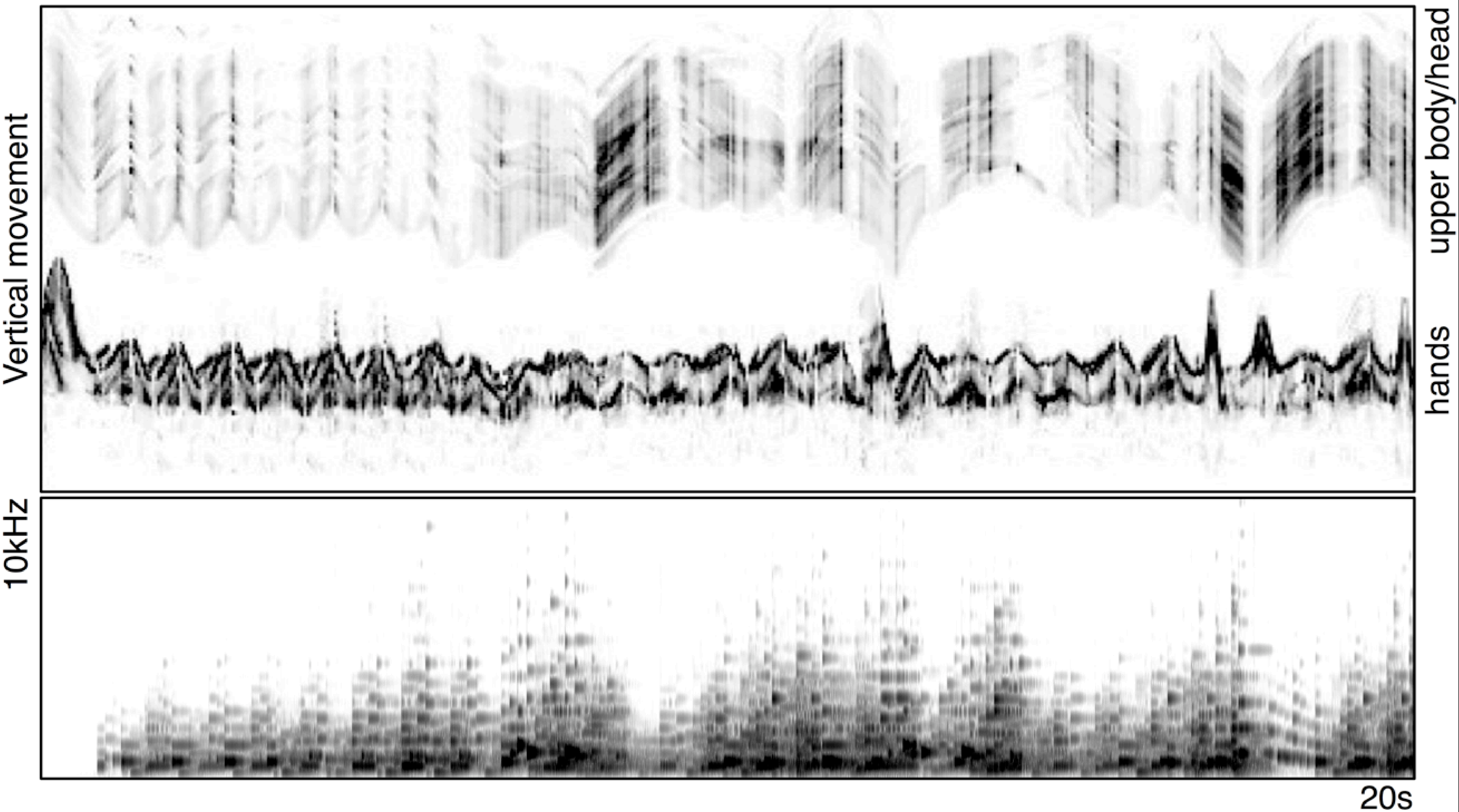
- In piano playing: fingers move to optimal position before hitting key (Engel, Flanders, and Soechting 1997)
- In string playing: left hand fingers in place in position well before playing of tones (Wiesendanger, Baader and Kazennikov 2006) and contextual smearing of bowing movements (Rasamimanana and Bevilacqua 2008)
- In drumming: In some cases, a drummer may start to prepare an accented stoke several strokes in advance (Dahl 2004)

Coarticulation in piano performance:

- Consider the hand movements of François René Duchable in his performance of the opening of the third movement of Beethoven’s *Tempest* Sonata:

And notice how the hand movements are in relation to the notated rhythm:

A motiongram and spectrogram of the same passage:

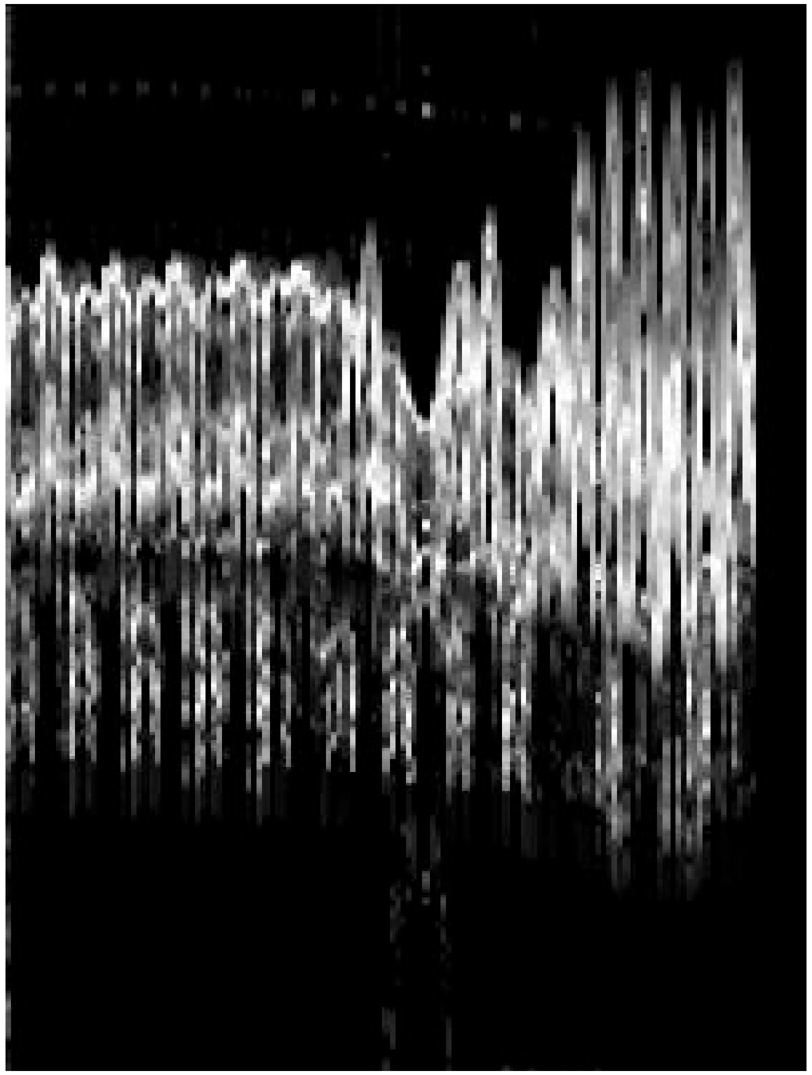


Coarticulation in sound perception

- Coarticulation has perceptual effects of creating cohesion
- Some examples, both artificial and natural
- Simulation with diphone model (bell-alphorn):
- Simulation with source-filter (trumpet + filter):
- Actual performance ("Winter"):
- Coarticulation as a bonus in physical models (multiple excitations, mass-spring, and dissipation resulting in smearing of sound):

Coarticulation in sound perception

- By the way: Lack of coarticulation one of the reasons for why sampler instruments sound “unnatural”
- Coarticulation in various other kinds of "fused" musical objects such as in Schaeffer’s *l'objet composite* (the two components separately, then fused):
- Coarticulation obvious in ornaments of various kinds of music (Norwegian fiddler):
- And as can be seen in the piano performance:



Chunking by coarticulation

- Coarticulation as periodic movement in various textural fragments, including cyclical patterns (cf. Large 2000, Waadeland 2000, etc.)
- Tutti texture chunk:
- Concatenated tutti texture chunks where detail variation can still be included in the same coarticulated movement trajectory:
- In other words: meter could be understood in the light of coarticulation

Our initial setup for recording coarticulation data:

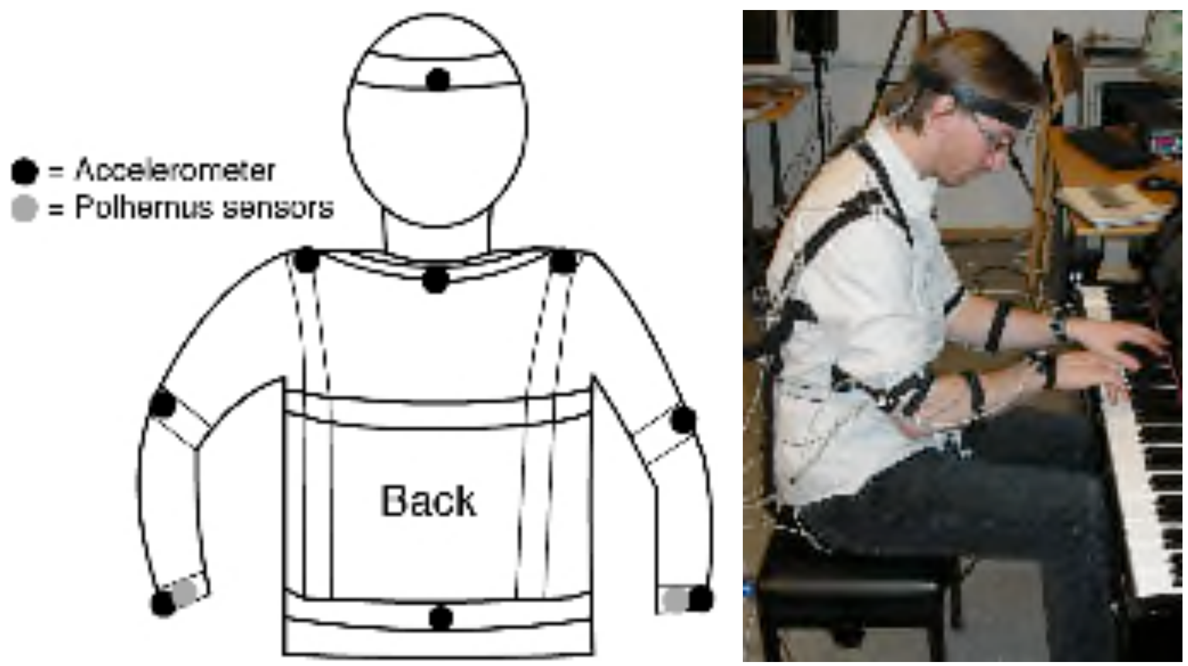


Figure 1. Positions of the accelerometers and polhemus sensors

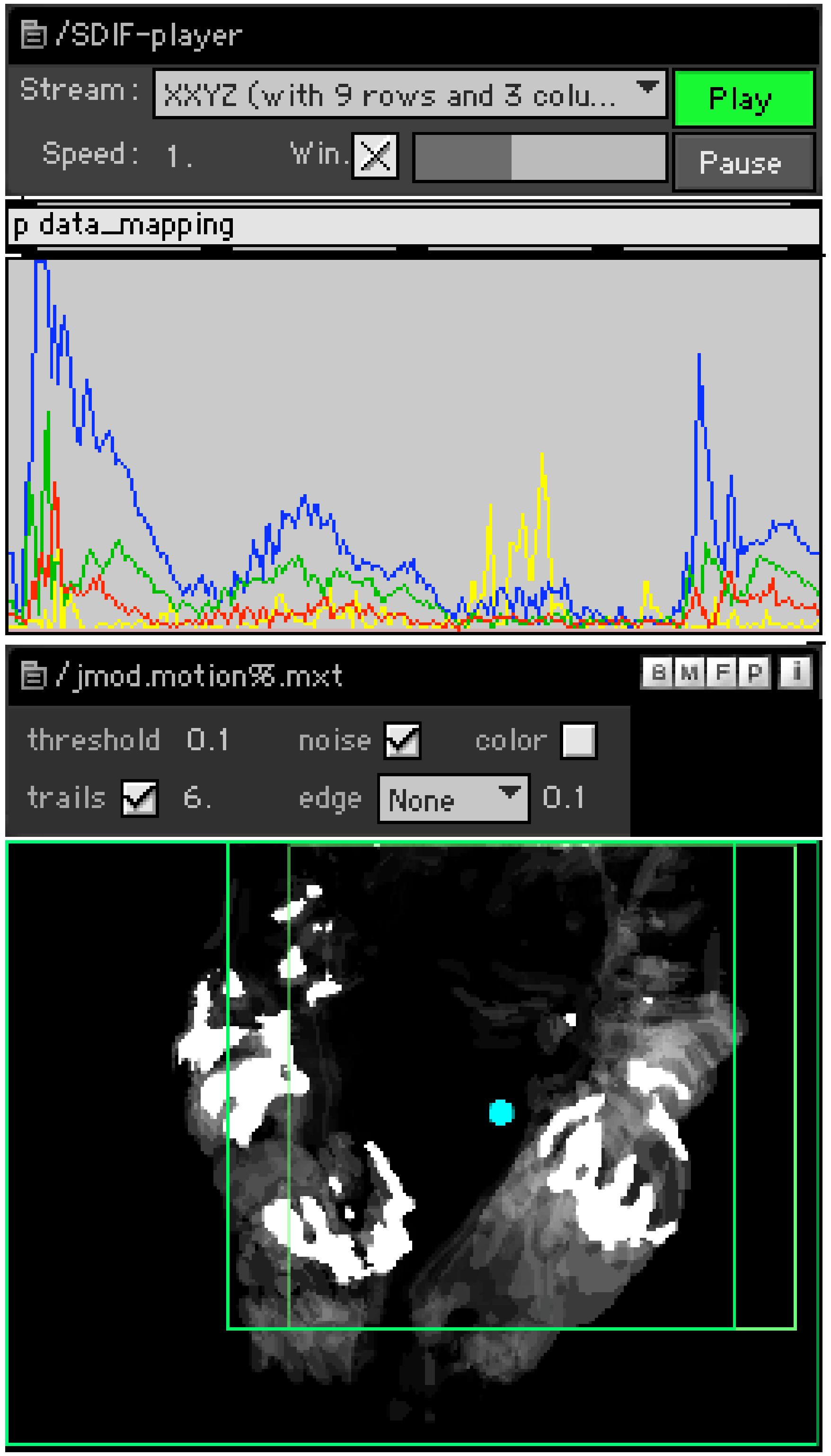
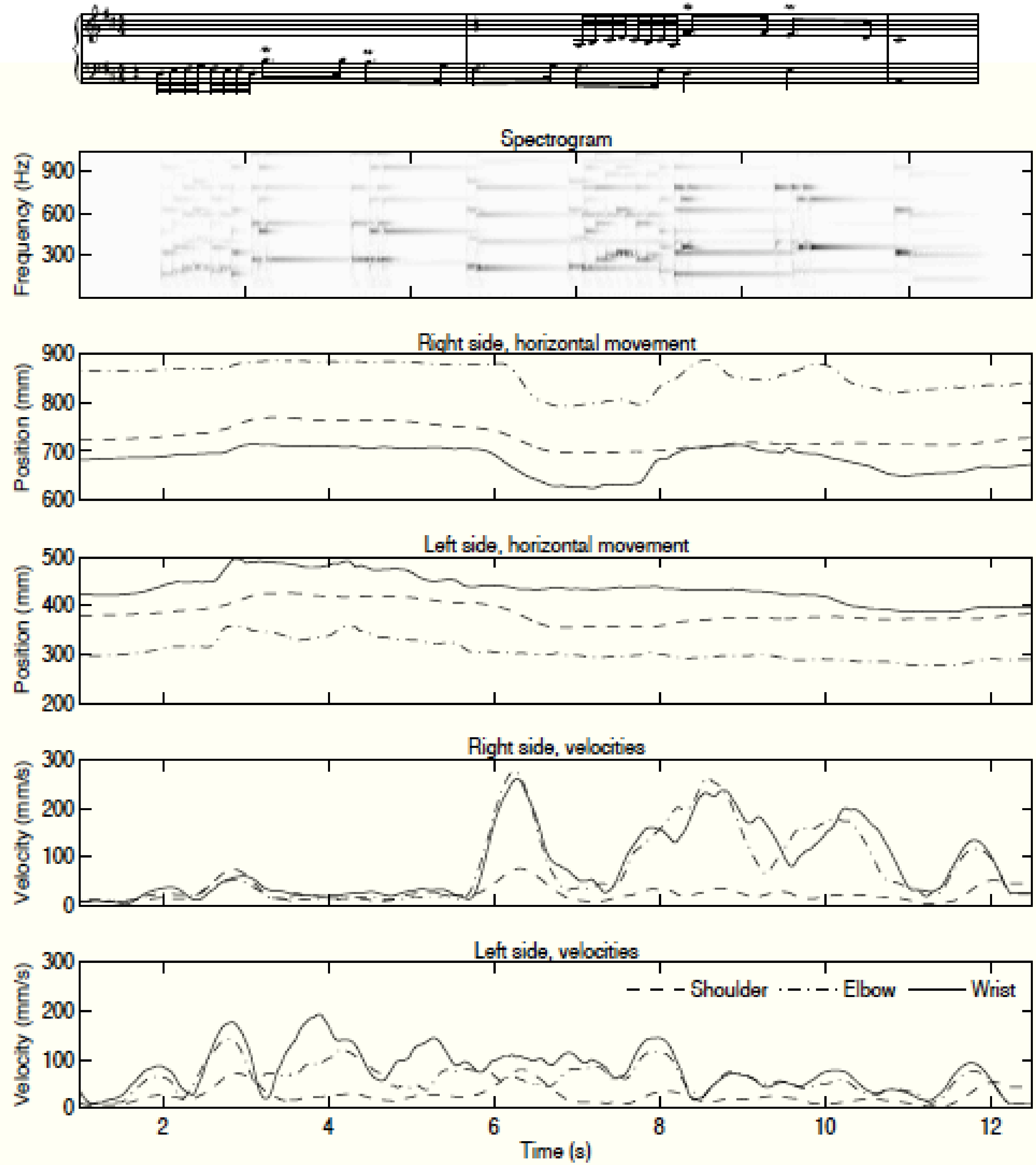


Figure 3. Plot of accelerometer values from the left wrist, elbow and shoulder, and EMG from the left lower arm. This is an example of a rather inaccurate multislider-plot in Max/MSP, here used to demonstrate temporal differences between the shoulder, elbow and wrist. For more accurate displays the recorded data should be imported to a more advanced analysis program.

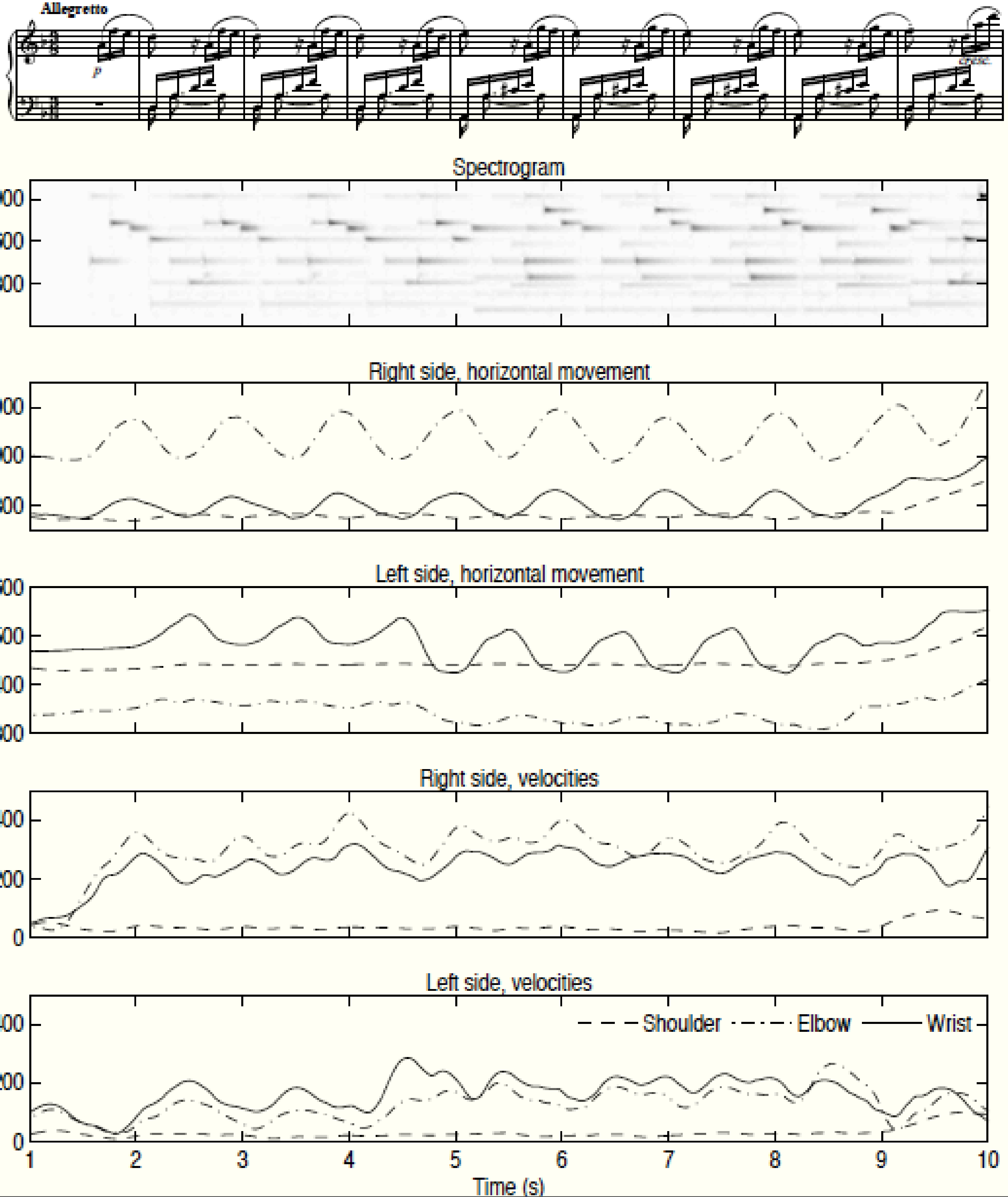
Simplified setup:

Comments on collecting sound-action data:

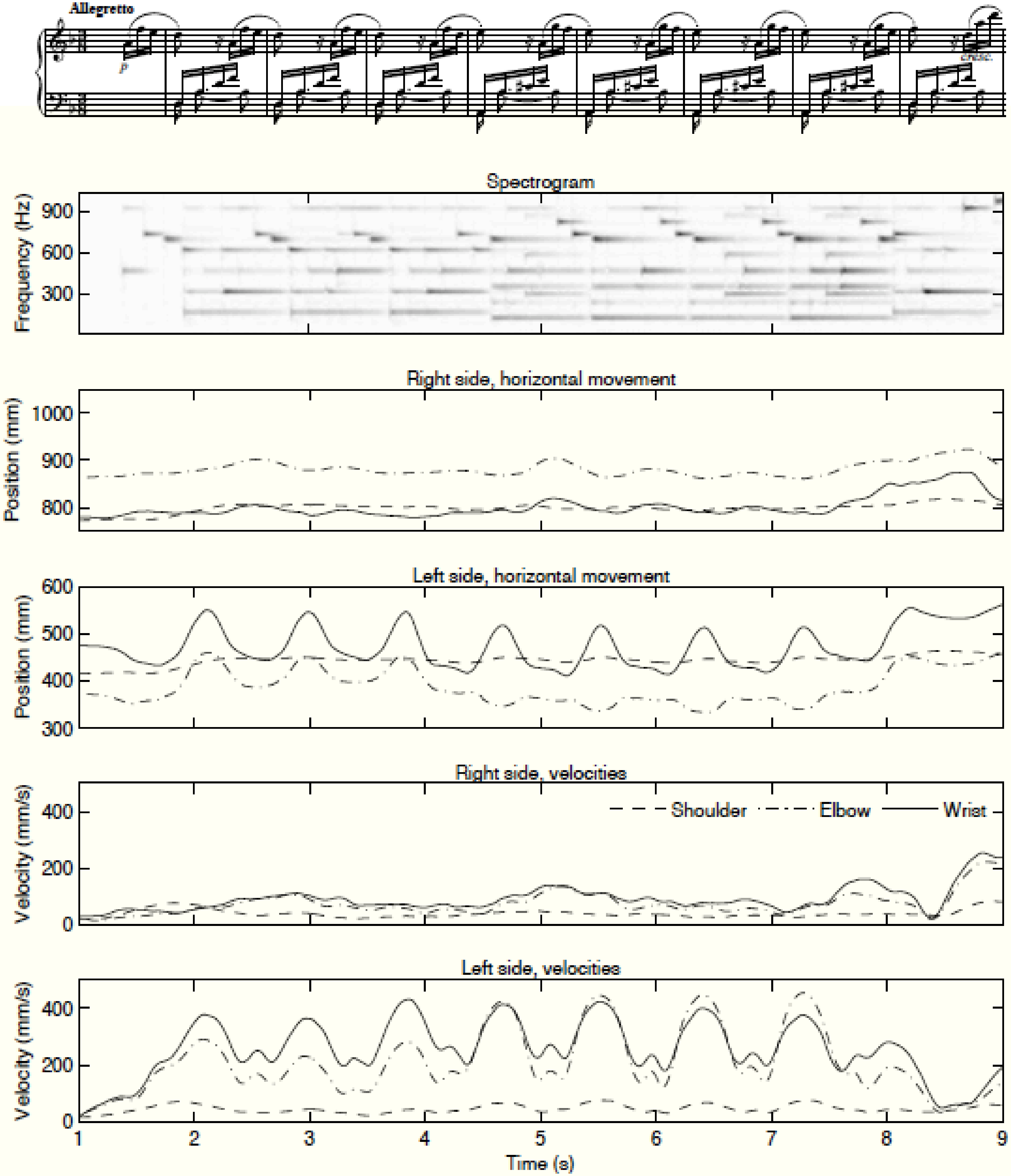
- Develop schemes for coding and representations of motion capture data and sound
- Challenges of synchrony and getting all components of the system to work smoothly together
- Challenges of representing multidimensional movement data
- Ambition: GDIF (Gesture Description Interchange Format) with “multi-track” style representation and playback possibilities
- Here some examples of our present work in progress with infrared motion capture data of piano performance:



Spectrogram
Frequency (Hz)
900
600
300
Right side, horizontal movement
Position (mm)
900
800
700
600
Left side, horizontal movement
Position (mm)
500
400
300
200
Right side, velocities
Velocity (mm/s)
300
200
100
0
Left side, velocities
Velocity (mm/s)
Shoulder
Elbow
Wrist
2
4
6
8
10
12
Time (s)



Allegretto
p
cresc.
Spectrogram
Frequency (Hz)
900
600
300
Right side, horizontal movement
Position (mm)
1000
900
800
Left side, horizontal movement
Position (mm)
600
500
400
300
Right side, velocities
Velocity (mm/s)
400
200
0
Left side, velocities
Velocity (mm/s)
400
200
0
Shoulder
Elbow
Wrist
1
2
3
4
5
6
7
8
9
10
Time (s)



Allegretto
p
cresc.
Spectrogram
Frequency (Hz)
900
600
300
Right side, horizontal movement
Position (mm)
1000
900
800
Left side, horizontal movement
Position (mm)
600
500
400
300
Right side, velocities
Velocity (mm/s)
400
200
0
Shoulder
Elbow
Wrist
Left side, velocities
Velocity (mm/s)
400
200
0
1
2
3
4
5
6
7
8
9
Time (s)

Goal-postures in coarticulation:

- So far: We believe there are indications of coarticulation in sound-action chunks, both in trajectory and velocity data
- And: We believe that coarticulation concerns both the sound and the sound-producing action, hence both perception and production
- But we also believe these sound-action chunks are centered on certain salient points in the music such as downbeats, other accents, and melodic peaks
- These salient points we consider as *goal-postures* both for the sound-producing movement and for the perceived sound

Goal-postures in coarticulation:

- Our intuitive sensations of such goal-postures are supported by the more general principles of goal-directed actions and goal-directed imitation
- In our analysis of music-related actions, we try to look for such goal-postures
- Trajectories to and from these goal-points we call *prefix* and *suffix*, and these trajectories are coarticulated in relation to the goal-postures
- Goal-posture chunks may either stand alone or overlap (see Godøy 2008 for details):

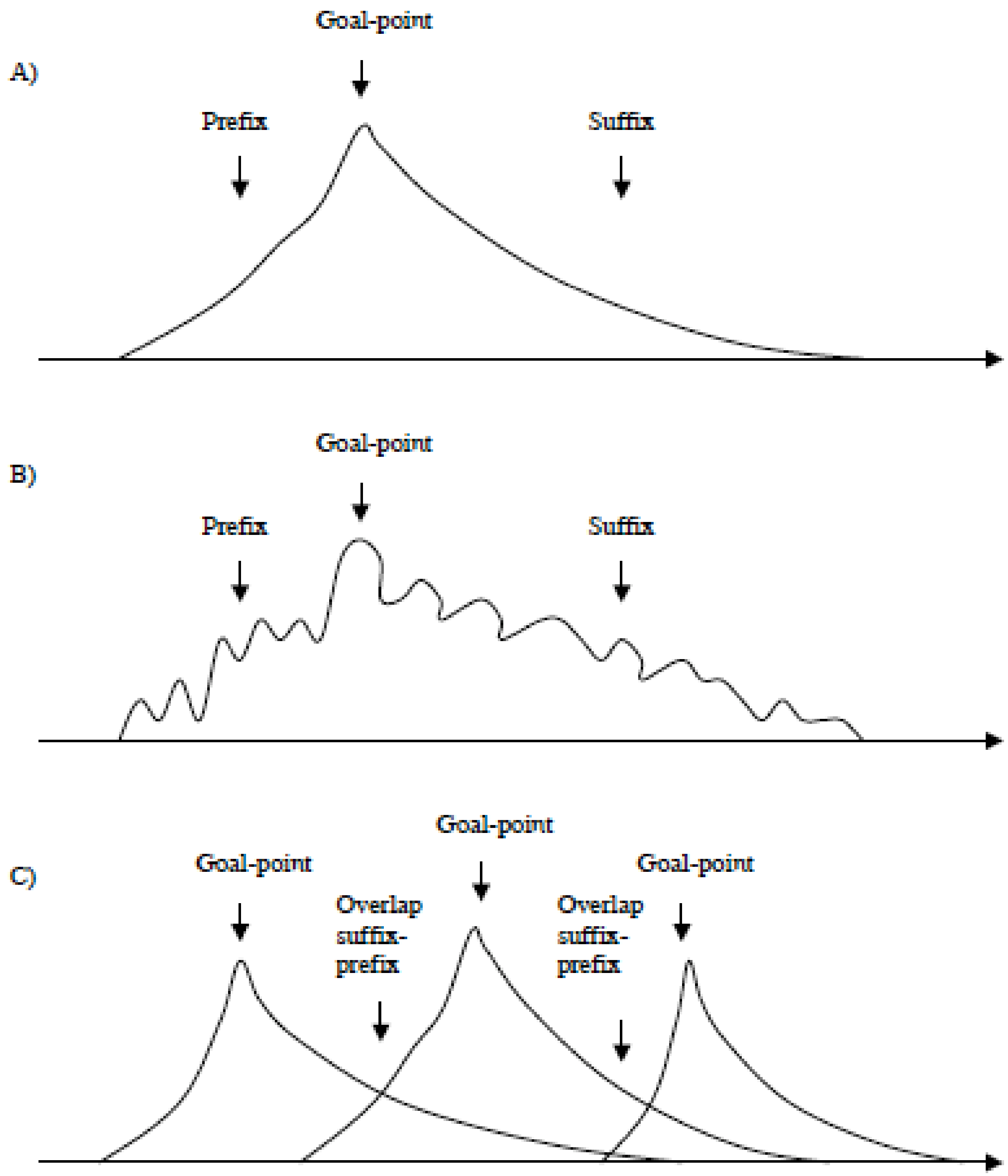
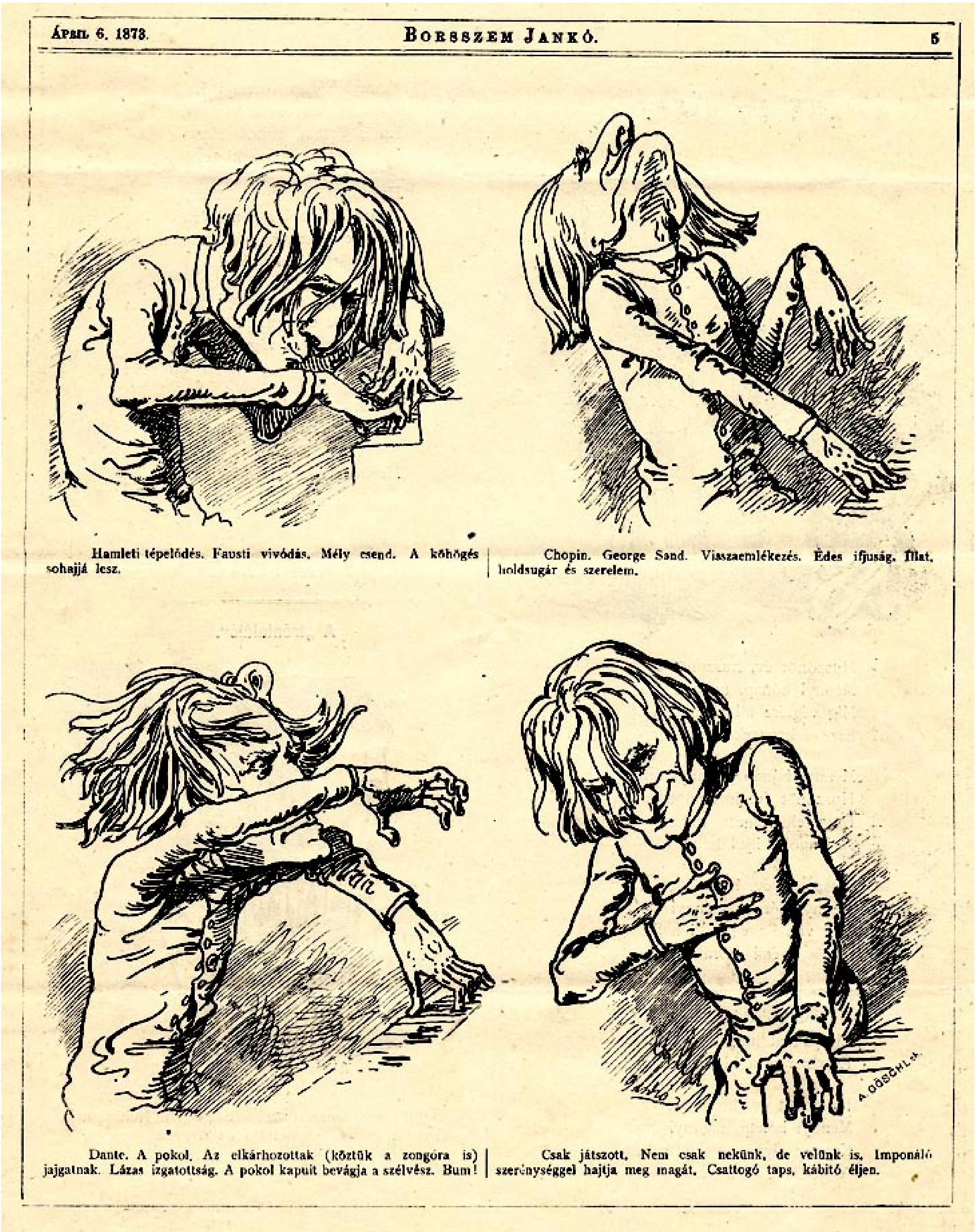


Figure 2. Schematic depiction of chunking by goal-points, A) with a singular chunk with just a prefix straight to the goal-point followed by a suffix from the goal-point, and B) with a prefix and a suffix containing several coarticulated sub-actions, and C), several goal-point chunks with overlapping prefixes and suffixes creating a sensation of continuity.

Goal-postures well-known from caricatures:



Ápril 6. 1873. BORSSZEM JANKÓ. 5

Hamleti tépelődés. Fausti vivódás. Mély csend. A köhögés sohajjá lesz.

Chopin. George Sand. Visszaemlékezés. Édes ifjuság. Illat, holdsugár és szerelem.

Dante. A pokol. Az elkárhozottak (köztük a zongora is) jajgatnak. Lázas izgatottság. A pokol kapuit bevágja a szélvész. Bum!

Csak játszott. Nem csak nekünk, de velünk is. Imponáló szerénységgel hajtja meg magát. Csattogó taps, kábító éljen.

Goal-postures well-known from cartoons:

Goal-postures in coarticulation:

- Actually, a similar idea of coarticulation centered around goal-postures has been presented in linguistics
- One advantage of this model is that it can accommodate continuous sound-actions since suffixes of past goal-postures may overlap with prefixes of new goal-postures, resulting in continuous motion
- This may also address the problem of chunking that perceivers experience but are hard to pinpoint in motion data

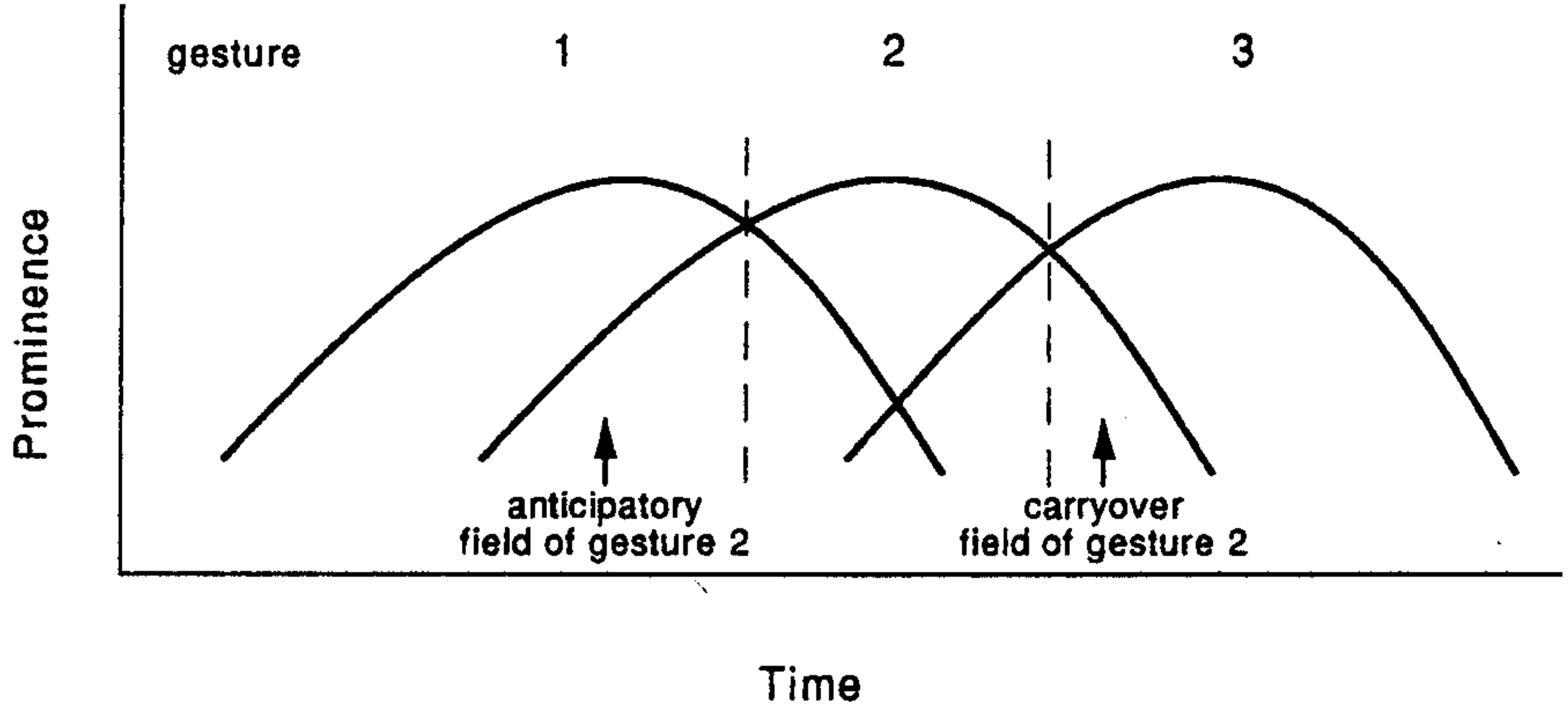


Figure 2.5 Representation of three overlapping phonetic gestures (from Fowler and Saltzman 1993). See text for details.

The gestures and their spatio-temporal organization

Figure 2.5, from Fowler and Saltzman (1993) illustrates how gestures are coproduced in speech. The activation of a gesture increases and decreases smoothly in time and so does its influence on the vocal tract shape and on the acoustic signal. In the figure, the vertical lines delimit a temporal interval (possibly corresponding to an acoustic segment) during which gesture 2 is prominent, i.e. has the maximal influence on the vocal tract shape, while the overlapping gestures 1 and 3 have a much weaker influence. Before this interval, the influence of gesture 1 predominates and the weaker influence of the following gesture gives rise to anticipatory coarticulation; analogously, after the interval of dominance of gesture 2, its influence determines the so-called carryover coarticulation. Thus both anticipatory and carryover effects are accounted for by the same principle of gesture overlap.

Challenges of understanding coarticulation:

- Advantageous to know what we are looking for when analyzing continuous movement
- The basic idea: Some moments in time are more important than others as implied with the idea of goal-postures
- Also suspicion that below a certain temporal threshold, sound and movement are perceived holistically, i.e. “in a now”
- Anticipatory motor cognition obviously “in an now”
- Hence, the distinction between static and dynamic, or between posture and movement, may become blurred

Next:

- More close-up studies of piano performance
- Drumming
- Marimba
- Cello
- Further on: Wind instruments
- Sound analysis of coarticulation effects
- Synthesis simulations of coarticulation

Main points of studying coarticulation:

- Change of paradigm, away from mostly notation-based musical analysis
- Regard all sounds as included in some action-trajectory
- Regards all features as related to actions
- Regard tones as included in chunks by coarticulation, hence as contextually smeared
- Regard chunks of musical sound as coinciding with action chunks

Applications of chunking by coarticulation:

- Rhythmic textural patterns: All kinds of rhythmical fragments, including cyclical patterns
- Timbral contours: All kinds of changes over time, various transients, fluctuations, etc.
- Melodic contours: All kinds of melodic shapes
- Expressive features, i.e. timing, articulation, accents, phrasing, “feel”/”groove”, etc.
- Music as scripts of sound-actions centered on keyframes

Requires better understanding of music-action relationships:

- Insights from neurocognitive research on perception-action links and on sensory integration
- Better understanding of anticipation and chunking in sound and in action
- A more developed conceptual apparatus for differentiating sound-actions features
- Problems with overt kinematics but covert dynamics, or:
- Discrepancy of what can be seen and/or measured and what listeners/performers feel

And also better methods for:

- Motion capture, including more unobtrusive setups and sensors
- Analysis and representation/visualization of motion capture data
- Biomechanical modeling and simulation
- Better understanding of the holistic perception of sound chunks e.g. (Grossberg and Myers 2000)
- Perceptual studies of degrees and modes of coarticulation, requiring good simulations (e.g. by diphone synthesis, time-varying filtering, physical model synthesis, etc.)

For more information, publications, and software:
www.fourms.uio.no