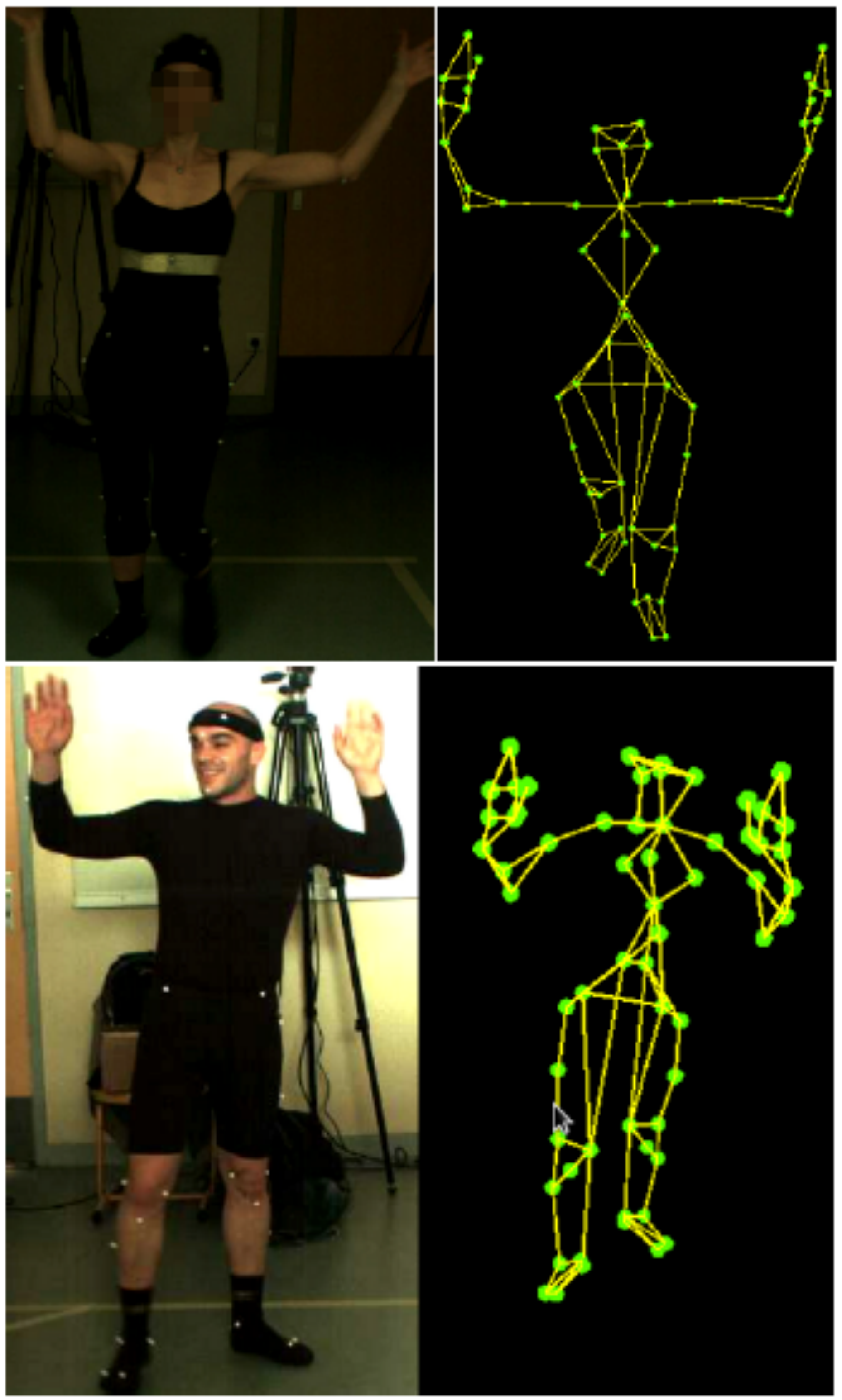
INF 3002 - Movement and Artificial Intelligence

Computable Descriptors of Human Motion

Sylvie Gibet

Introduction



- **Large amount of motion capture data, but…**
- **Objectives of recorded mocap databases:**
    - Analysis, classification, motion indexing and retrieving
    - Synthesis (replay, edition, re-composition)
- Meany application fields: sport, biomechanics, handicap, computer animation, video games, HCI, etc.
- Very few available motion databases
- Very few annotated movements with their meaning, emotion, etc.

Introduction: computable measures of movements

- **Review of computable measures of movements**
  - Extracted from full-body motion capture, or specific to body parts for specific applications (musical gestures, sign languages, etc.)
  - Generalizing, re-formalizing used descriptors to give a unified view (2D, 3D, different levels of precision, frequency rate, noise, etc.)
  - Distinction between
    - Low-level descriptors: computed from raw data
    - High-level descriptors: manually or automatically interpreted (human observations, annotations, automatic computations)

Outline of the lecture

Movement representation

Low-level motion descriptors

- Kinematic descriptors
- Geometry descriptors
- Energy descriptors

High-level motion descriptors

- Categories based on Laban Movement Analysis theory

Examples

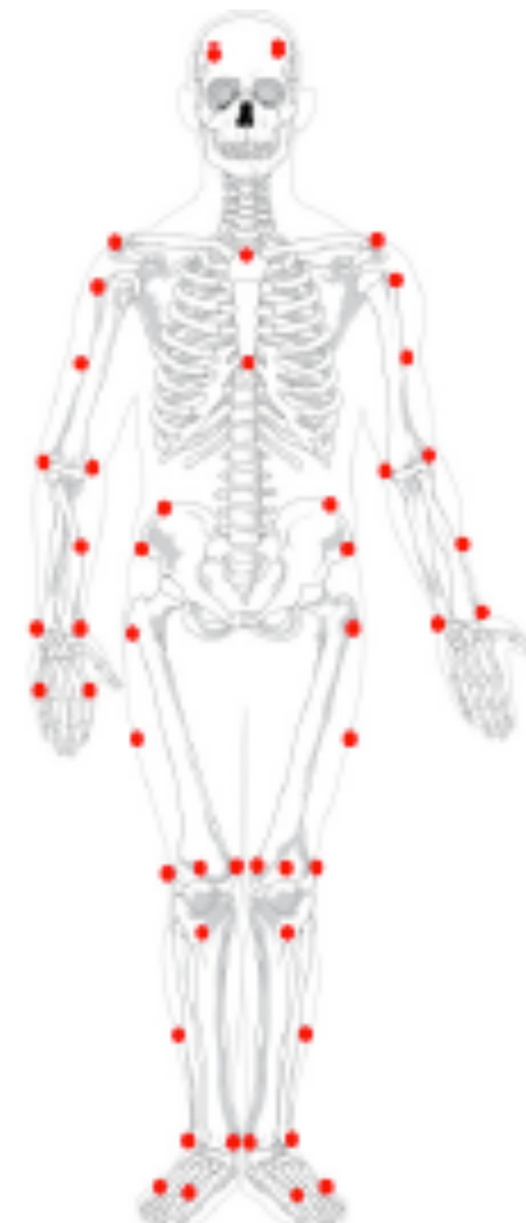
Motion representation

Based on position representation of the joints

- **Set of *m* joint positions: one pose at time *ti***
  - absolute positions relatively to the root or the world coordinate system

    $$\mathbf{x}(t_i) = \{\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^m\}(t_i)$$

    x= (x, y, z)
  - relative positions (position of the fingers' joints relatively to the wrist joint), …

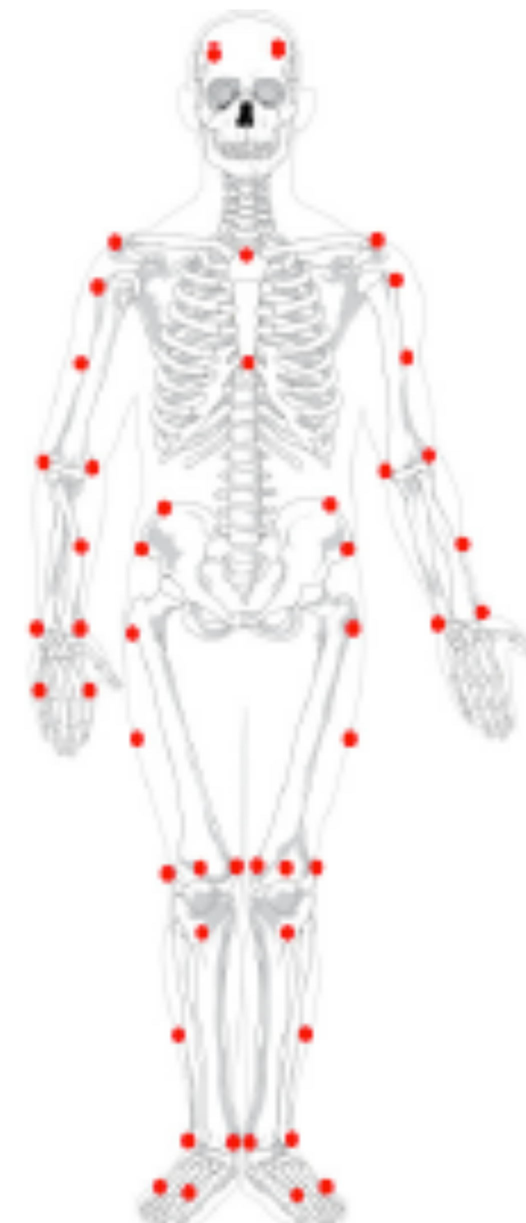


Motion representation

Based on angular representation of the joints

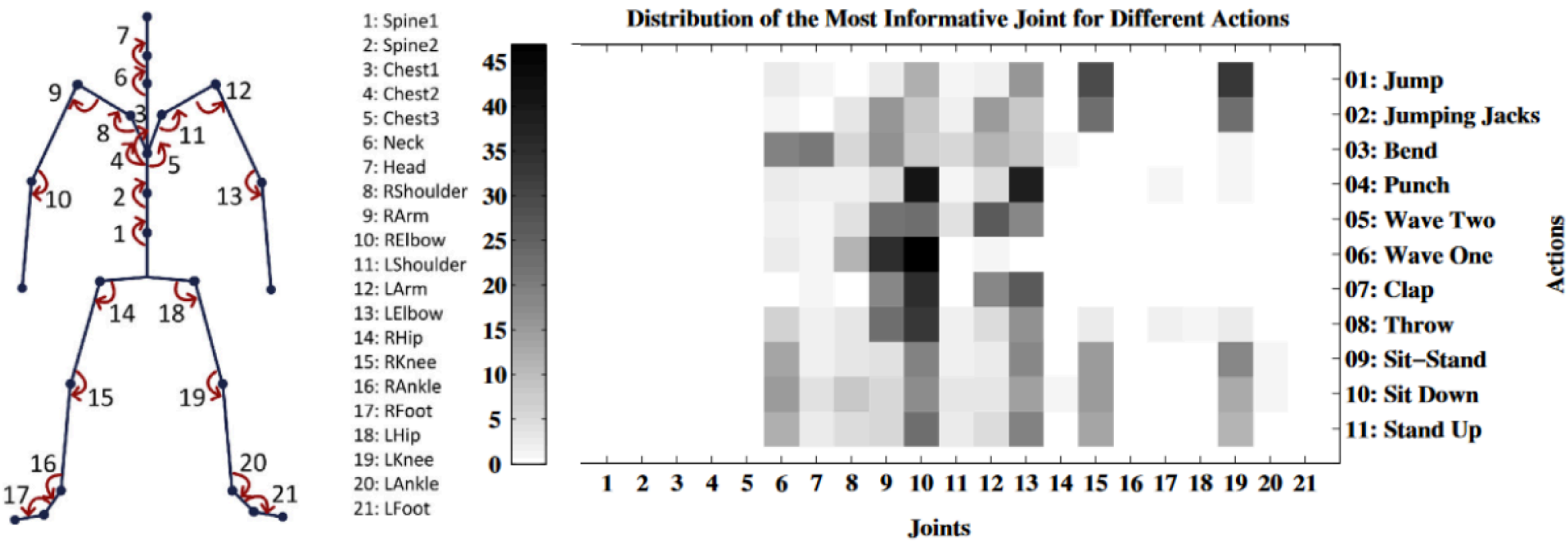
- **Set of *m* joint angles**
  - relative or absolute rotations relative to the parent or root joint
  - pose encoded by a 3x*m* dimensional vector (Euler representation) or a 4x*m* dimensional vector (quaternion representation)

$$\mathbf{x}(t_i) = \{\mathbf{q}^1, \mathbf{q}^2, ..., \mathbf{q}^m\}(t_i)$$



Motion representation

Based on angular representation of the joints



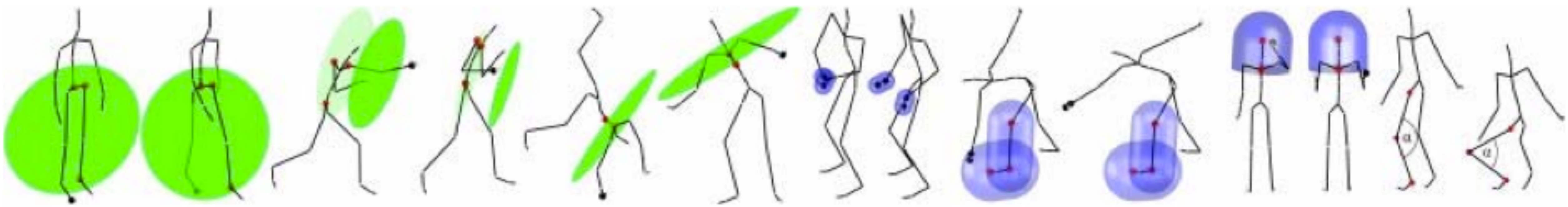
(a) (b)

- 21 raw angles (MHAD database); distribution of the most informative joints for different classes of actions; each input corresponds to the % of time where a joint is considered to be the most informative for a given action (dark: high %)

Motion representation

Based on geometric features (Müller et al., 2005)

- **Applied to motion retrieval**
    - boolean function
    - describe relation between body parts



Motion representation

Based on a set of *distances (Naour et al, 2012, 2013)*

- The skeleton is represented by a graph

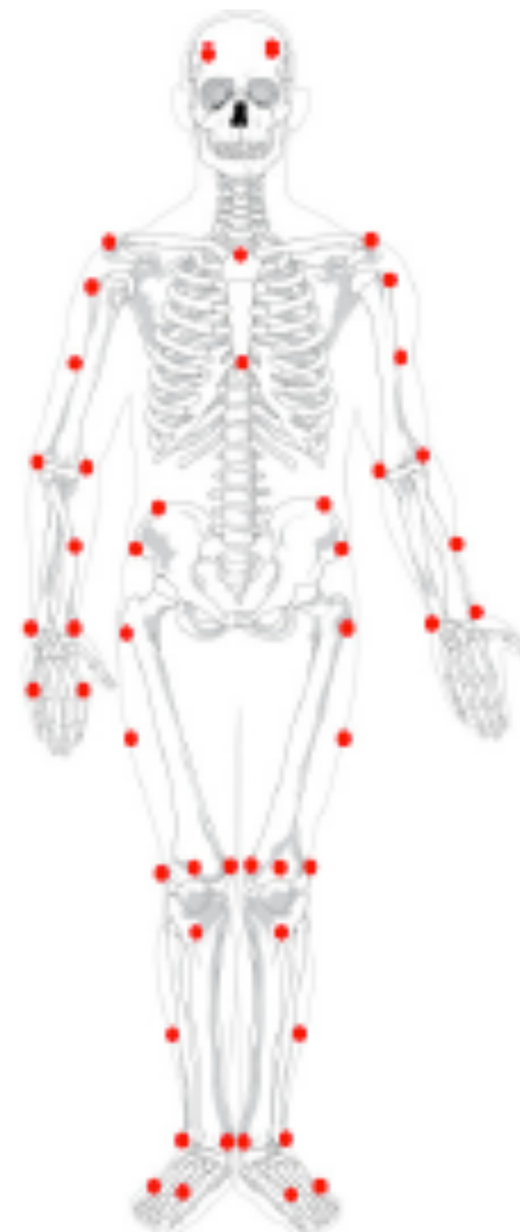
G = (V, E, d) where

V : set of points (centers of joints)

E : arcs between points

d: set of distances (VxV -> R+)

$$\forall (\boldsymbol{x}_i, \boldsymbol{x}_j) \in E,\ \|\boldsymbol{x}_i - \boldsymbol{x}_j\| = d_{ij},$$



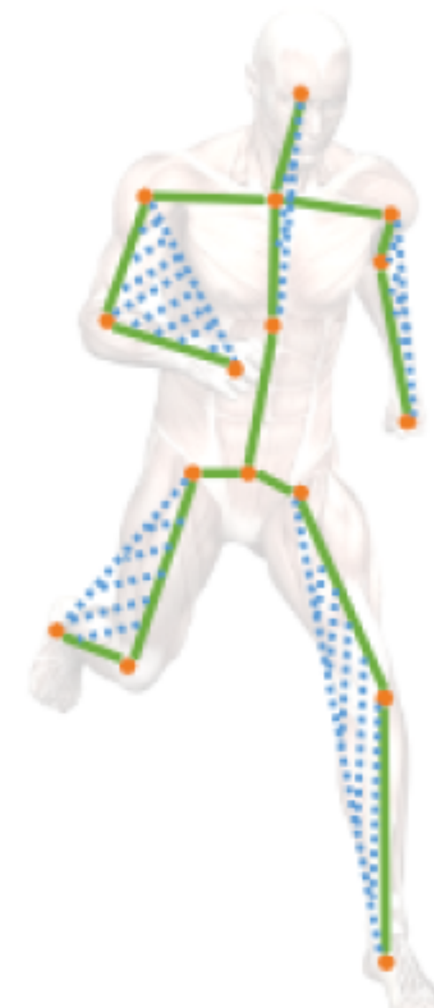
Applications: motion retrival, inverse kinematics,

motion editing, motion reconstruction, motion recognition

Motion representation

Based on a set geometrical relations between joints *(Zhang et al, 2018)*

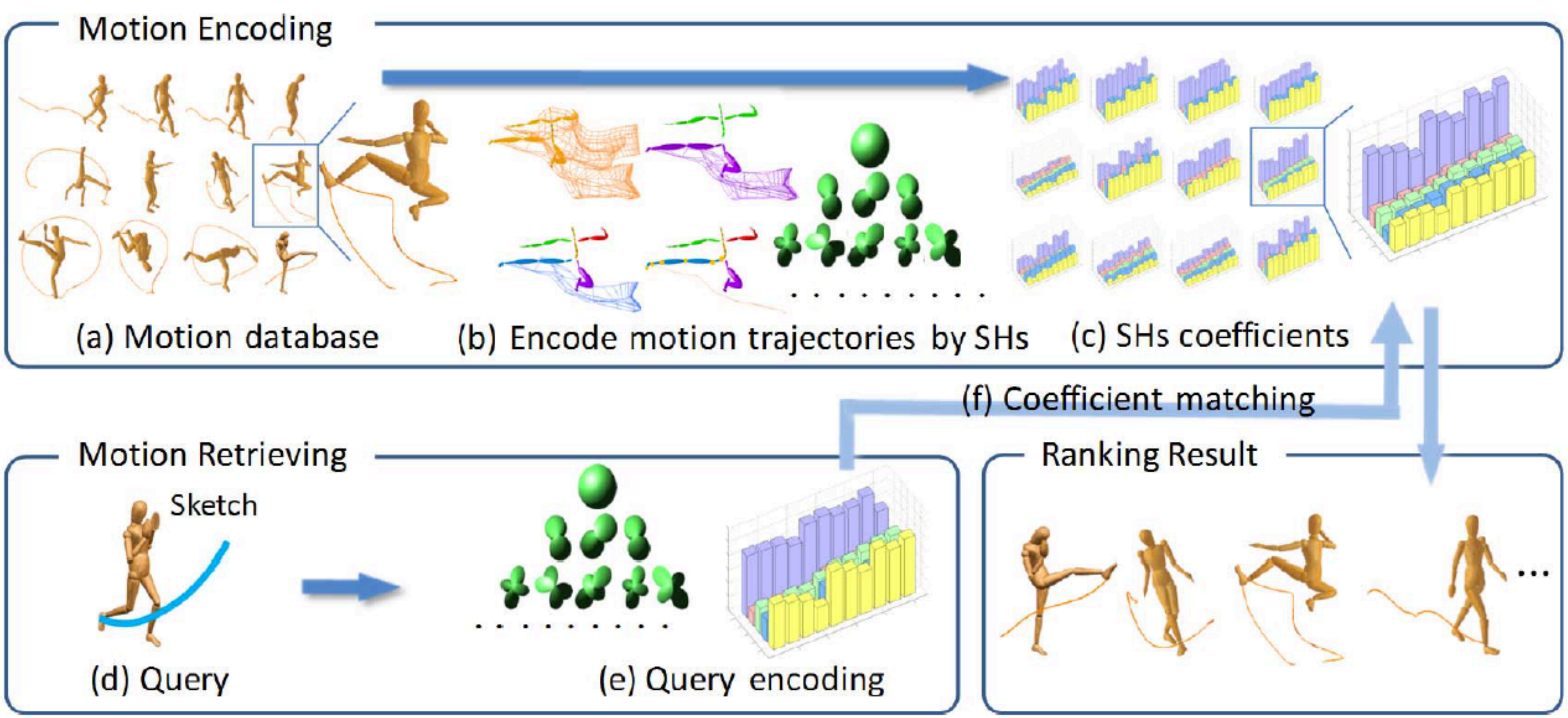
- *Relations between the joint planes defined by at least 3 articulations*
- *Only select those planes that are relevant for specific actions*



Motion representation

Based on spherical harmonics (Chao et al., 2012)

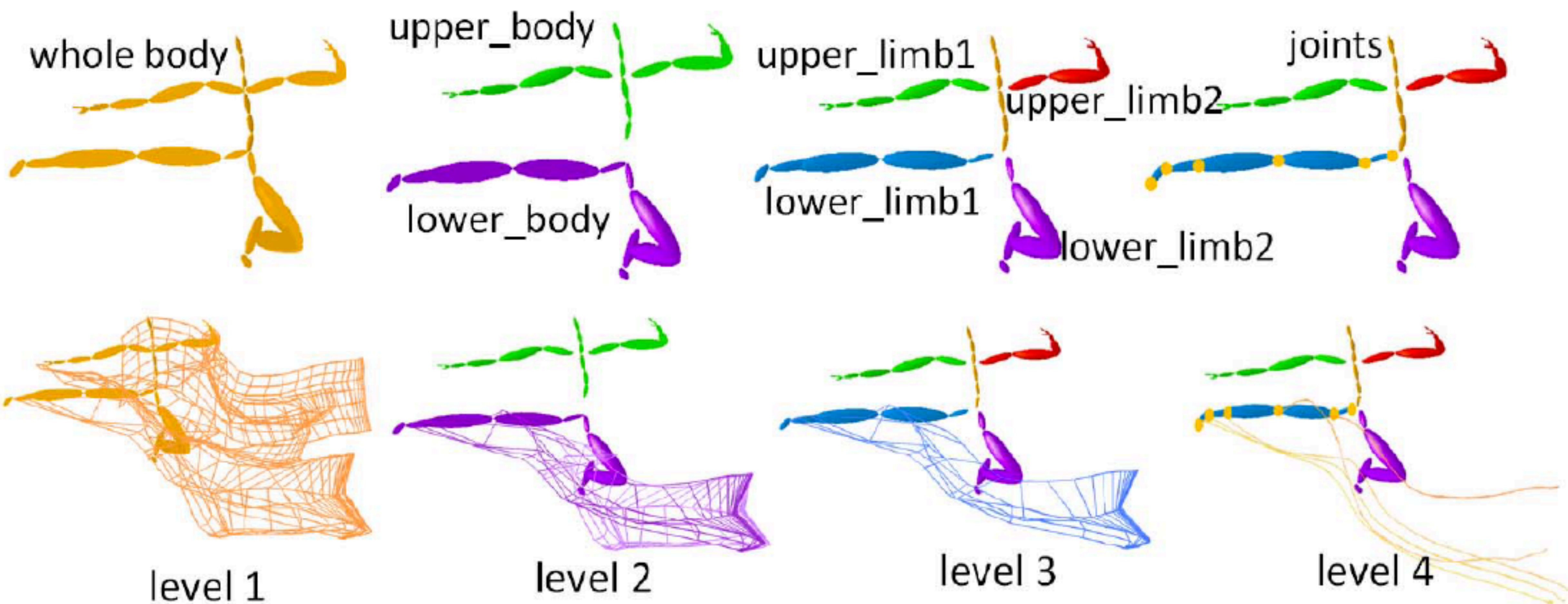
- **Applied to motion retrieval**
    - encoding of mocap data
    - sketching interface for describing the query



Motion representation

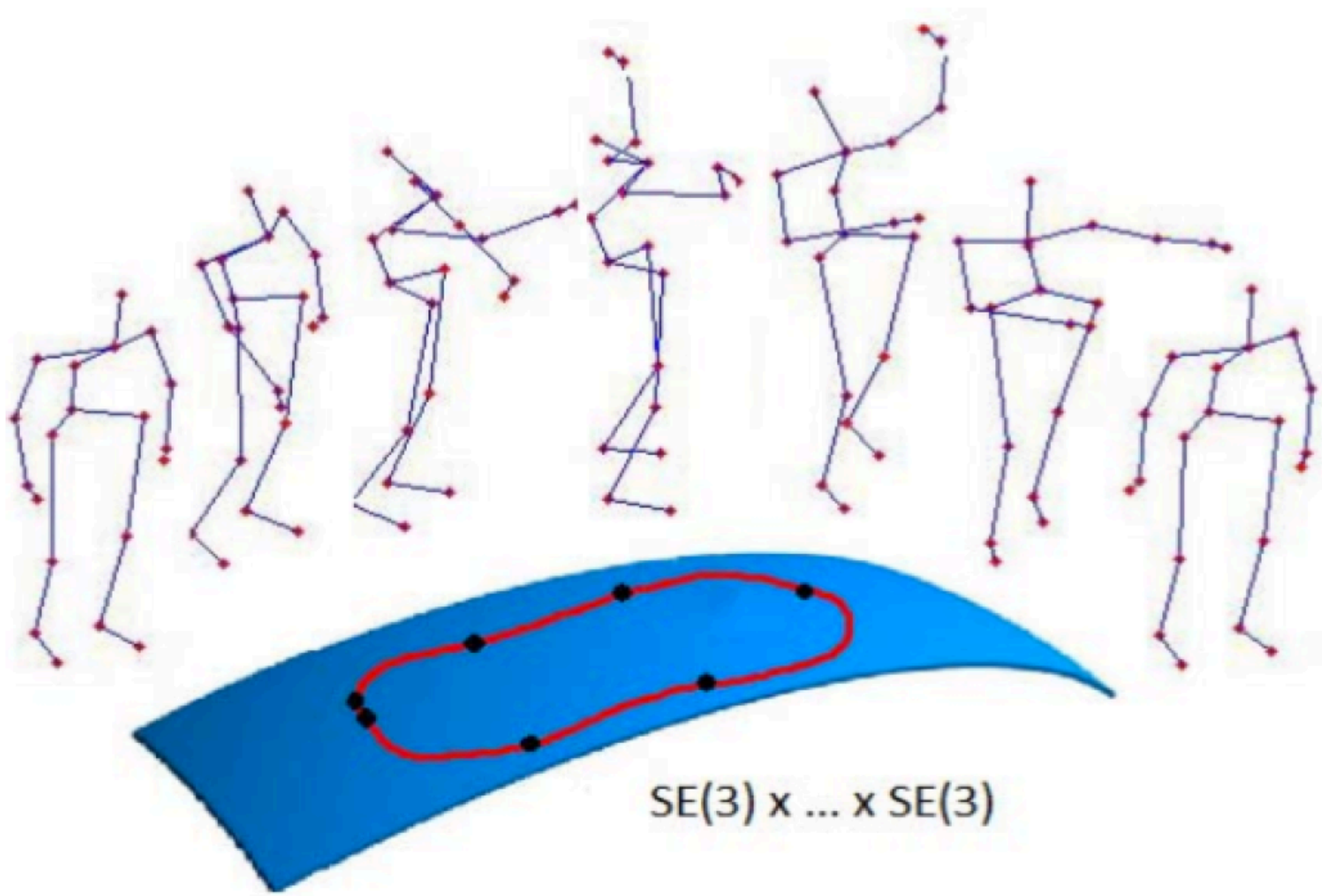
Based on spherical harmonics (Chao et al., 2012)

- Novel motion encoding method for retrieving motion in a coarse-to-fine manner



Motion representation

- **Statistical covariance between absolute values of joint positions (Hussein et al., 2013)**
  - 3D trajectories represented as a curve in the Lie group after changing the representation space



Low-level motion descriptors

- Kinematic descriptors
- Geometry descriptors
- Energy descriptors

Kinematic descriptors

- **For one joint *k* at time *ti***
  - **Velocity:** instantaneous velocity vector for joint k (derivative of the position)

$$\mathbf{v}^k(t_i) = \frac{\mathbf{x}^k(t_{i+1}) - \mathbf{x}^k(t_{i-1})}{2\delta t}$$

with the velocity amplitude (scalar value)

$$v^k(t_i) = \sqrt{\mathbf{v}_x^k(t_i)^2 + \mathbf{v}_y^k(t_i)^2 + \mathbf{v}_z^k(t_i)^2}$$

Kinematic descriptors

- **For one joint *k* at time *ti***
  - **Acceleration:** instantaneous acceleration vector (derivative of the velocity)

$$\mathbf{a}^k(t_i) = \frac{\mathbf{x}^k(t_{i+1}) - 2.\mathbf{x}^k(t_i) + \mathbf{x}^k(t_{i-1})}{\delta t^2}$$

  - with the acceleration amplitude (scalar value)

$$a^k(t_i) = \sqrt{\mathbf{a}_x^k(t_i)^2 + \mathbf{a}_y^k(t_i)^2 + \mathbf{a}_z^k(t_i)^2}$$

Kinematic descriptors - exercice

- Pour la semaine prochaine : retrouvez les expressions pour calculer les vitesses et les accélérations (dérivations) avec :
  - Euler à l’ordre 1
  - Euler à l’ordre 2
- En déduire les schémas d’intégration utilisés en simulation de particules par exemple

Kinematic descriptors

- **For one joint *k* at time *ti***
  - **Jerk:** measure of smoothness for point-to-point reaching movements (derivative of the acceleration):

$$\mathbf{j}^k(t_i) = \frac{\mathbf{x}^k(t_{i+2}) - 2.\mathbf{x}^k(t_{i+1}) + 2.\mathbf{x}^k(t_{i-1}) - \mathbf{x}^k(t_{i-2})}{2.\delta t^3}$$

    - with the jerk amplitude (scalar value)

$$j^k(t_i) = \sqrt{\mathbf{j}_x^k(t_i)^2 + \mathbf{j}_y^k(t_i)^2 + \mathbf{j}_z^k(t_i)^2}$$

    - and the dimensionless value

$$\frac{(t_{start} - t_{end})^5}{{H_{max}}^2} . \sum_{t_{start}}^{t_{end}} j(t_i)^2$$

Kinematic descriptors

- **Curvature (resp radius of curvature):** measures how fast a curve is changing direction

$$C^k(t_i) = \frac{\|\mathbf{a}^k(t_i) \times \mathbf{v}^k(t_i)\|}{v^k(t_i)^3}$$

$$R^k(t_i) = 1/C^k(t_i)$$

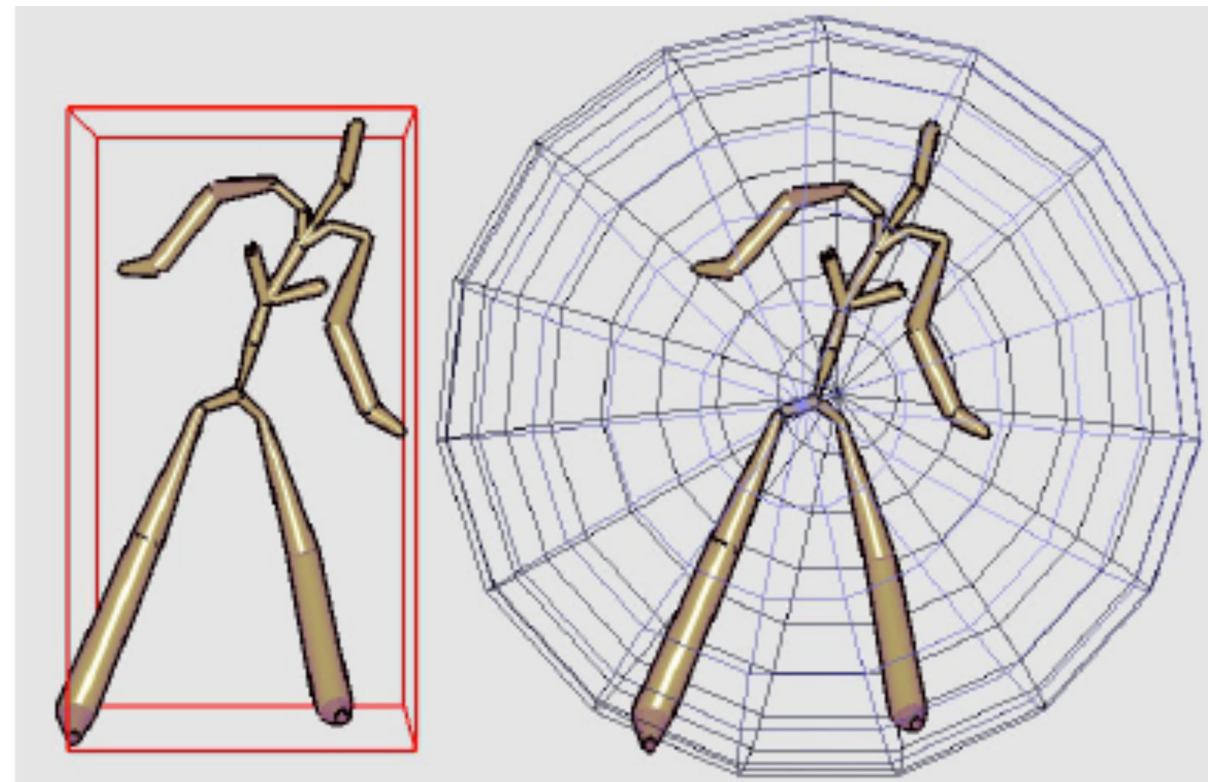
Kinematic descriptors

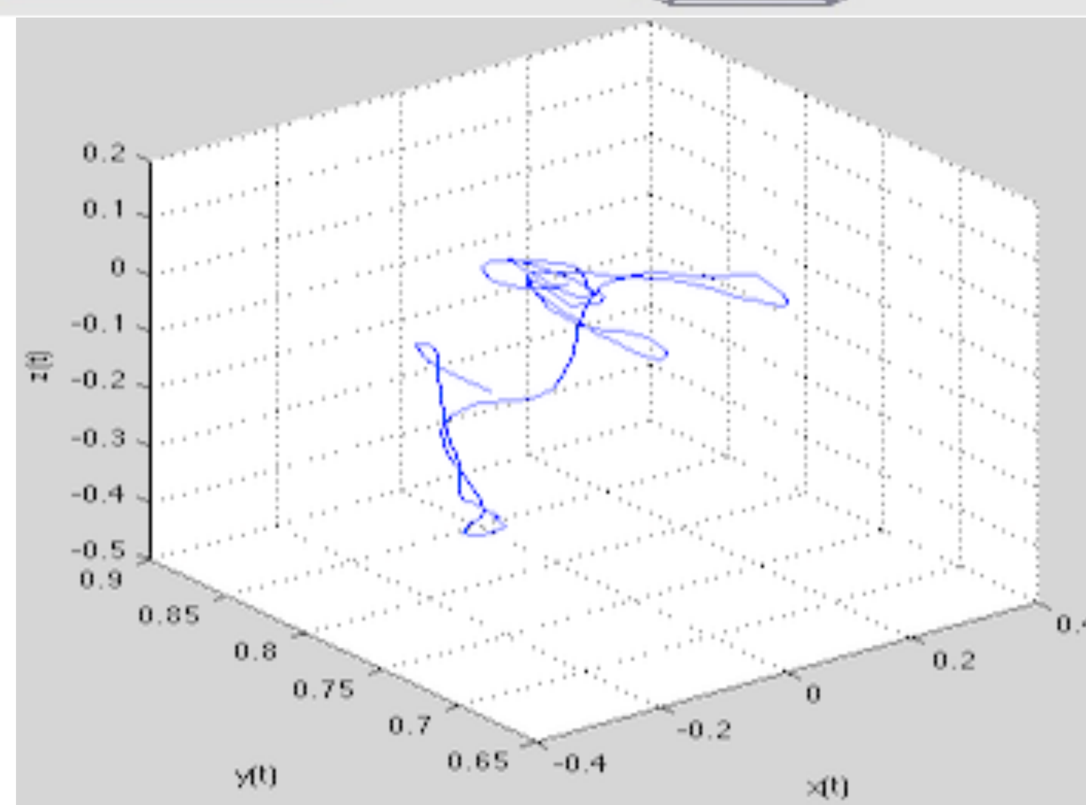
- **Curvilinear displacement:** measures the displacement along the trajectory
  - computed as the integration of the absolute value of the velocity

$$Ds = \sum_{t_b}^{t_e} |v(t)|dt$$

Geometry descriptors

- **For one joint *k* at time t_i**
  - **Bounding shapes:** describes the 3D shape of the body (box, sphere, ellipsoid)
  - **Distances** between specific joints (two hands, two feet, etc.)





Energy descriptors

- **For one joint *k* at time *ti***
    - **Center of mass:** computes the weighted position of the center of mass of the body for a set of joints

$$\mathbf{com}(t_i) = \frac{\sum_{k \in \mathbb{K}} w_k . \mathbf{x}^k(t_i)}{\sum_{k \in \mathbb{K}} w_k}$$

    - with w_k: weighting of the joint k (may depend on the masses of the different segments)

Energy descriptors

- **For one joint *k* at time *t_i***
  - **Quantity of Motion :** weighted average of the speeds of the joints' body, with K a set of joints
  - w_k: mass of the segment k

$$qom(t_i) = \frac{\sum_{k \in \mathbb{K}} w_k . v^k(t_i)}{\sum_{k \in \mathbb{K}} w_k}$$

Energy descriptors

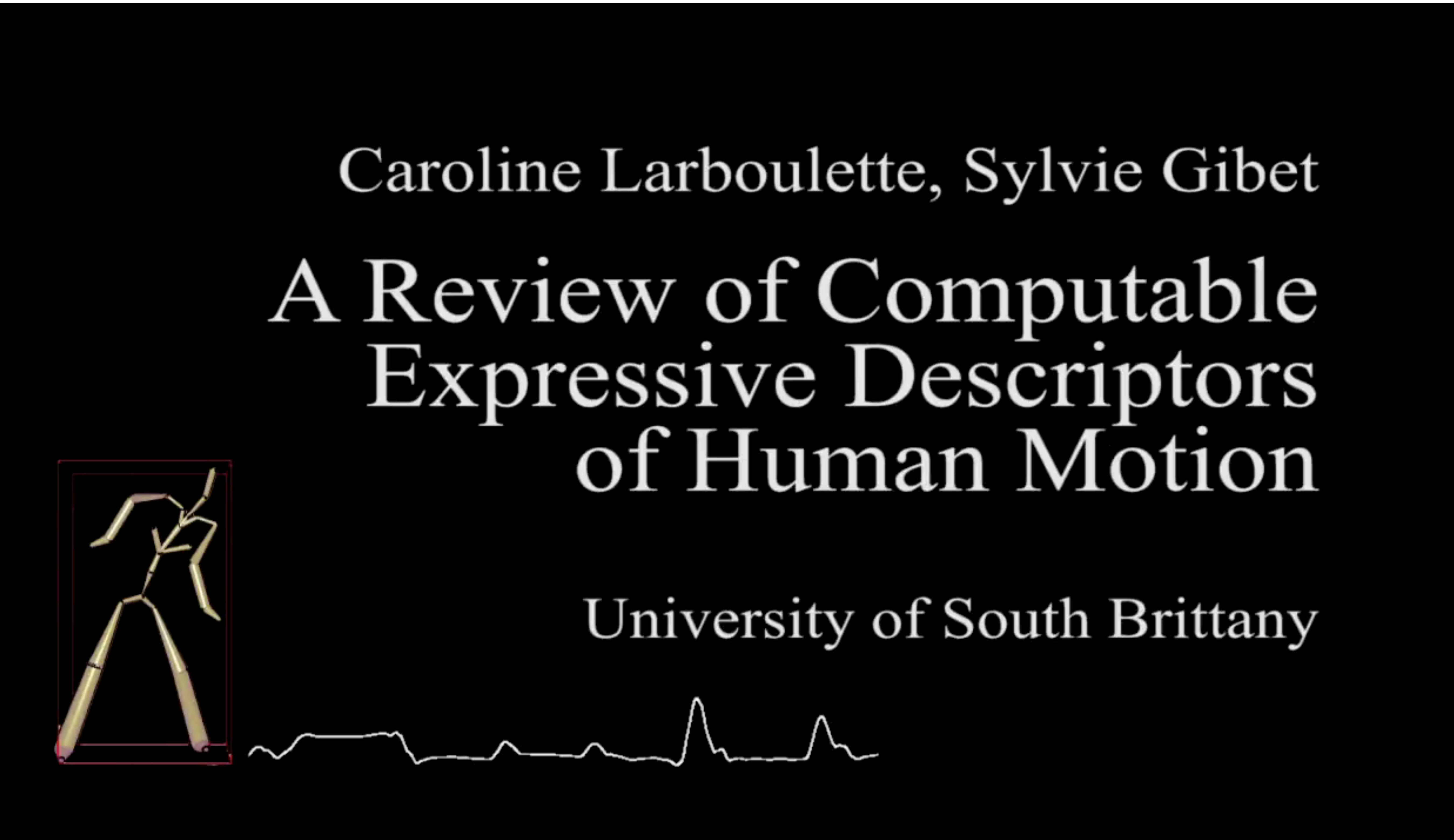
- **For one joint *k* at time *ti***
  - **Kinetic energy :**

$$\frac{1}{2} . \sum_{k \in \mathbb{K}} m_k . v^k(t_i)^2$$

with m_k the mass of the segment k

  - **Potential energy**

$$\sum_{k \in \mathbb{K}} m_k . h . g$$


Caroline Larboulette, Sylvie Gibet
A Review of Computable Expressive Descriptors of Human Motion
University of South Brittany

High-level motion descriptors

High-level motion descriptors

- **Why high-level descriptors: interpretated by movement experts or perceptual evaluations**
  - Semantic labels
  - Discrete (quantified) variables

- **How are they computed**
  - Use low-level features
  - Expressed as continuous features, but for full-body motion

High-level motion descriptors

- **Various levels of spatial discretization**
  - One joint k
  - A set of joints
  - Full-body postures

- **Various levels of temporal discretization**
  - One frame
  - A set of key-frames at t_i
  - A movement chunk (action, sequence, sign, etc.)

High-level motion descriptors: examples

- **Geometric descriptors**
    - See Müller et al
- **Adhoc descriptors**
    - **Periodicity**: computed as the mean of auto-correlation (Alaoui et al.)
    - **Repetition**: for repeated strokes (Pelachaud et al.)
    - **Movement phase**: computes the product of the speed with the curvature
- **Laban Movement Analysis (LMA) descriptors**
    - see at the end of the course

Discussion

- **Challenges**
  - **Find a consistent and reduced set of descriptors** that best characterizes motion
    - Most of current work choose descriptors in an ad hoc manner, which may not fully preserve the expressive qualities of motion
    - We will favor automatic approaches where there is no *a priori* assumption and descriptors will emerge automatically

Discussion: challenges

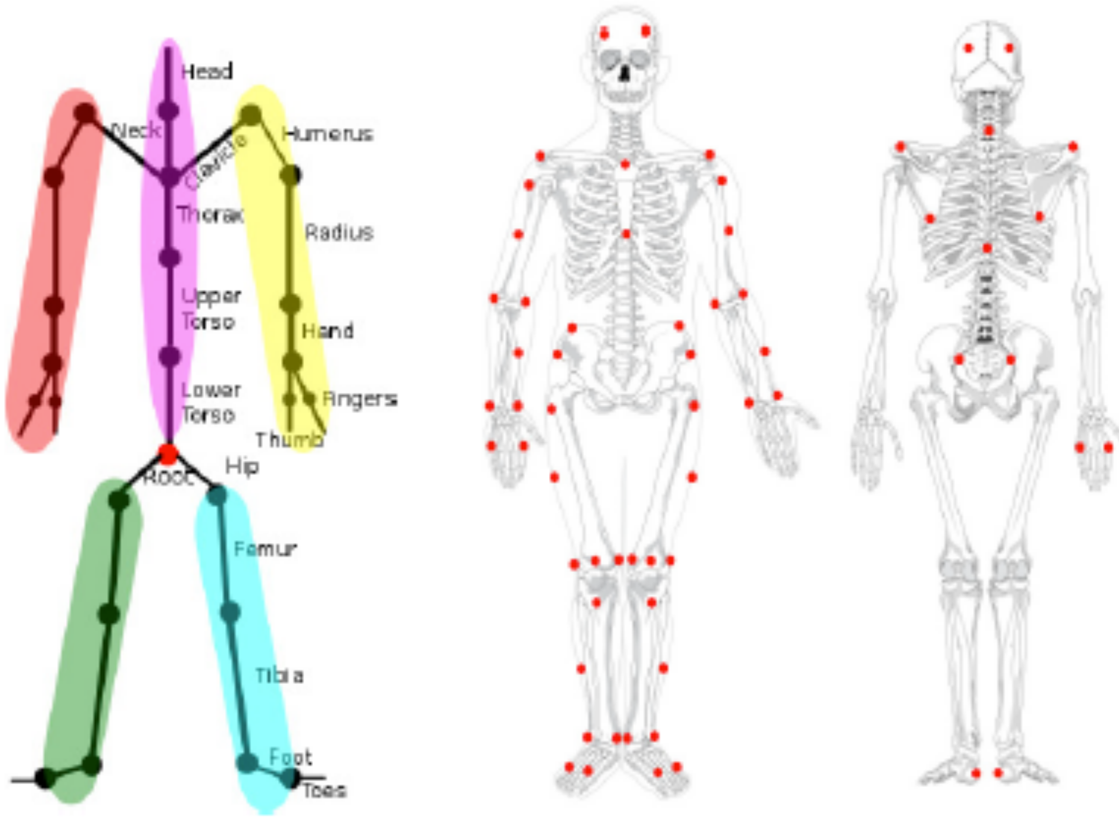
- **Find a consistent and reduced set of descriptors**
  - **1. Level of quantization**
  - **2. Spatial reduction**
  - **3. Temporal segmentation**

Discussion: challenges

- **1. Level of quantization**
  - Ranges from boolean to continuous values
    - ✦Determine an optimized set of values for each descriptor: depends on the range of motion of the movement, and on the choice of the database
    - ✦Evaluate the sensitivity of the results

Discussion: challenges

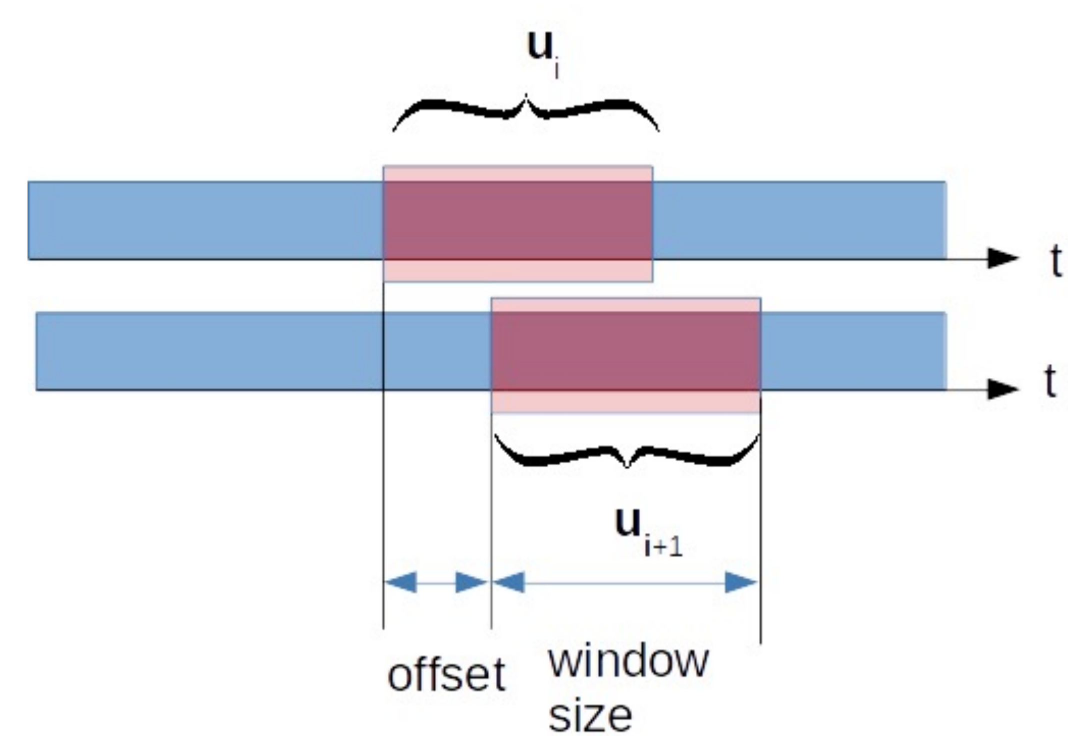
- **2. Spatial reduction of the skeleton**
- Define on which body part the measures should be estimated
  - ✦Find the optimal splitting of the body
  - ✦Find the optimal weights associated to the body parts



•

Discussion: challenges

- **3. Temporal segmentation**
  - Key-frame extraction relatively to the actions [Zhang et al., 2013]
  - Use of a sliding window: offset and size of the window [Carreno et al. 2015]



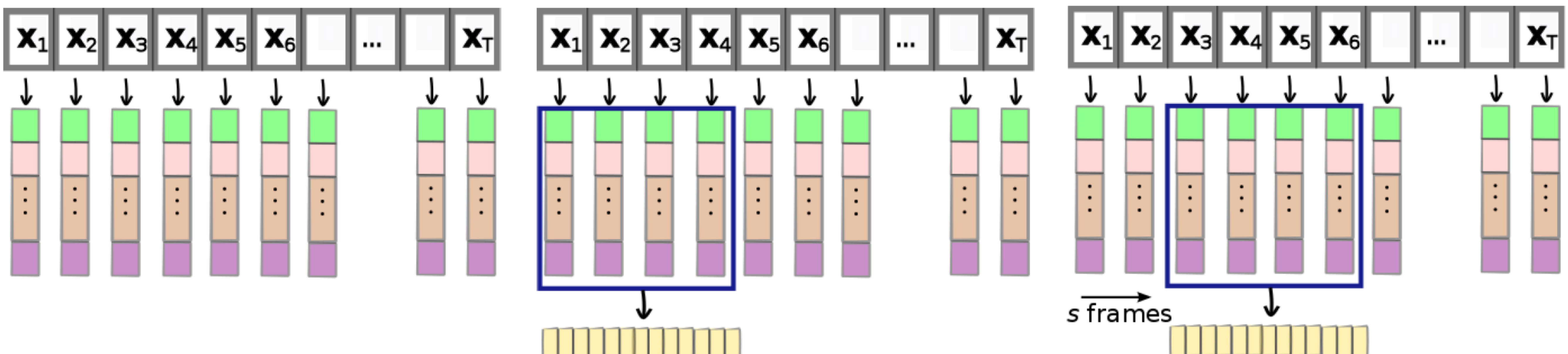
Discussion: challenges

Example of descriptors for body motion (Pamela Carreno PhD, 2016)

- A motion $\mathbf{M} = \{\mathbf{X}_1, ..., \mathbf{X}_T\}$, with $\mathbf{X}_i \in \mathbb{R}^{3\times27}$ and emotion label c.
- Kinematic features for each joint: velocity ($\dot{\mathbf{X}}_i$), acceleration ($\ddot{\mathbf{X}}_i$), jerk ($\dddot{\mathbf{X}}_i$), and curvature (κ_i).
- Features are summarized by their μ and σ (1 joint = 8 features by vector).

Motion descriptors (Carreno et al., ACII 2015)

A feature vector for each *motion chunk* (sampling of L contiguous frames from $\mathbf{M}$).



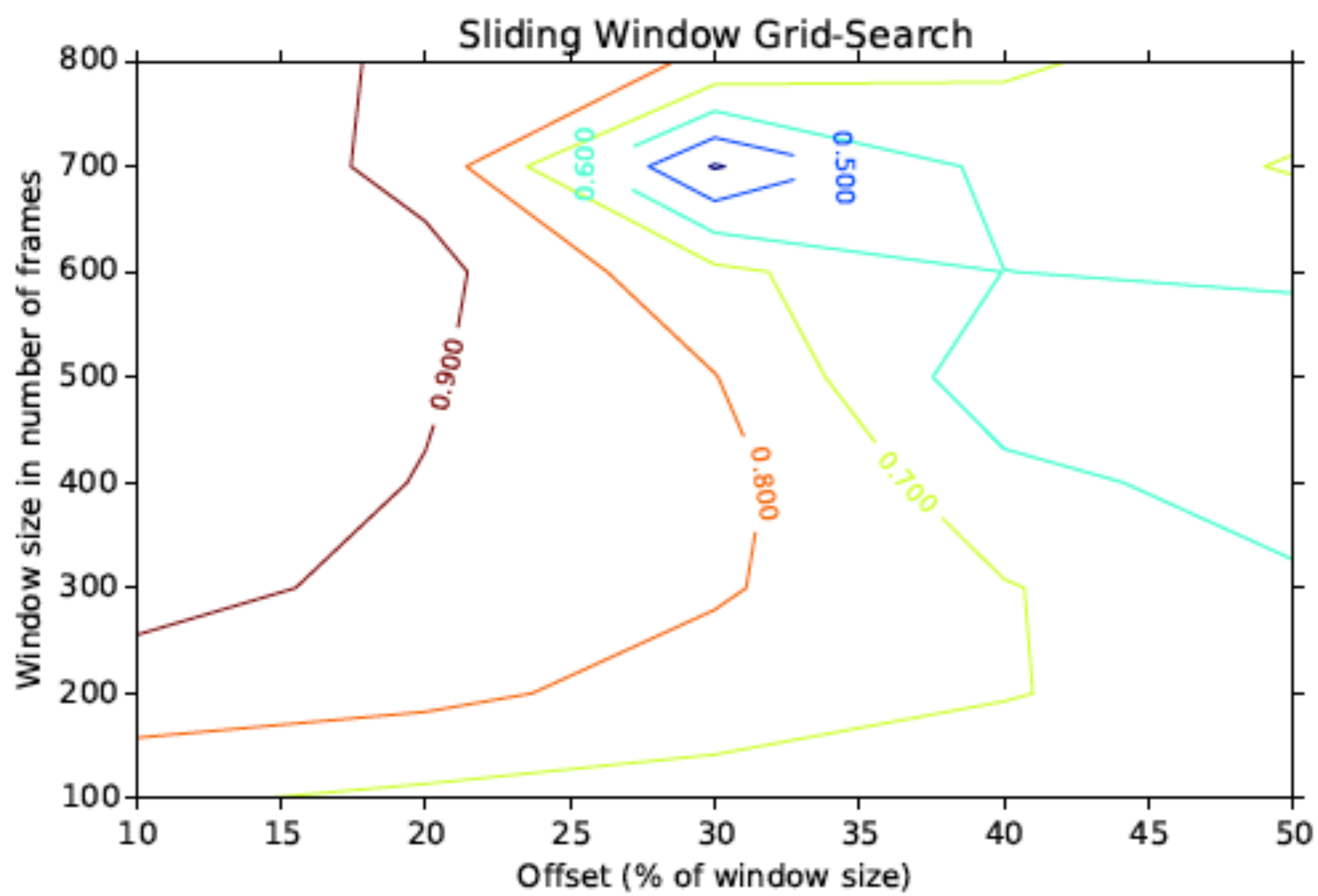
Compute features for all frames in $\mathbf{M}$

Compute features norm and summarize by μ and σ across i-*th motion chunk*

Move s frames forward and repeat step 2

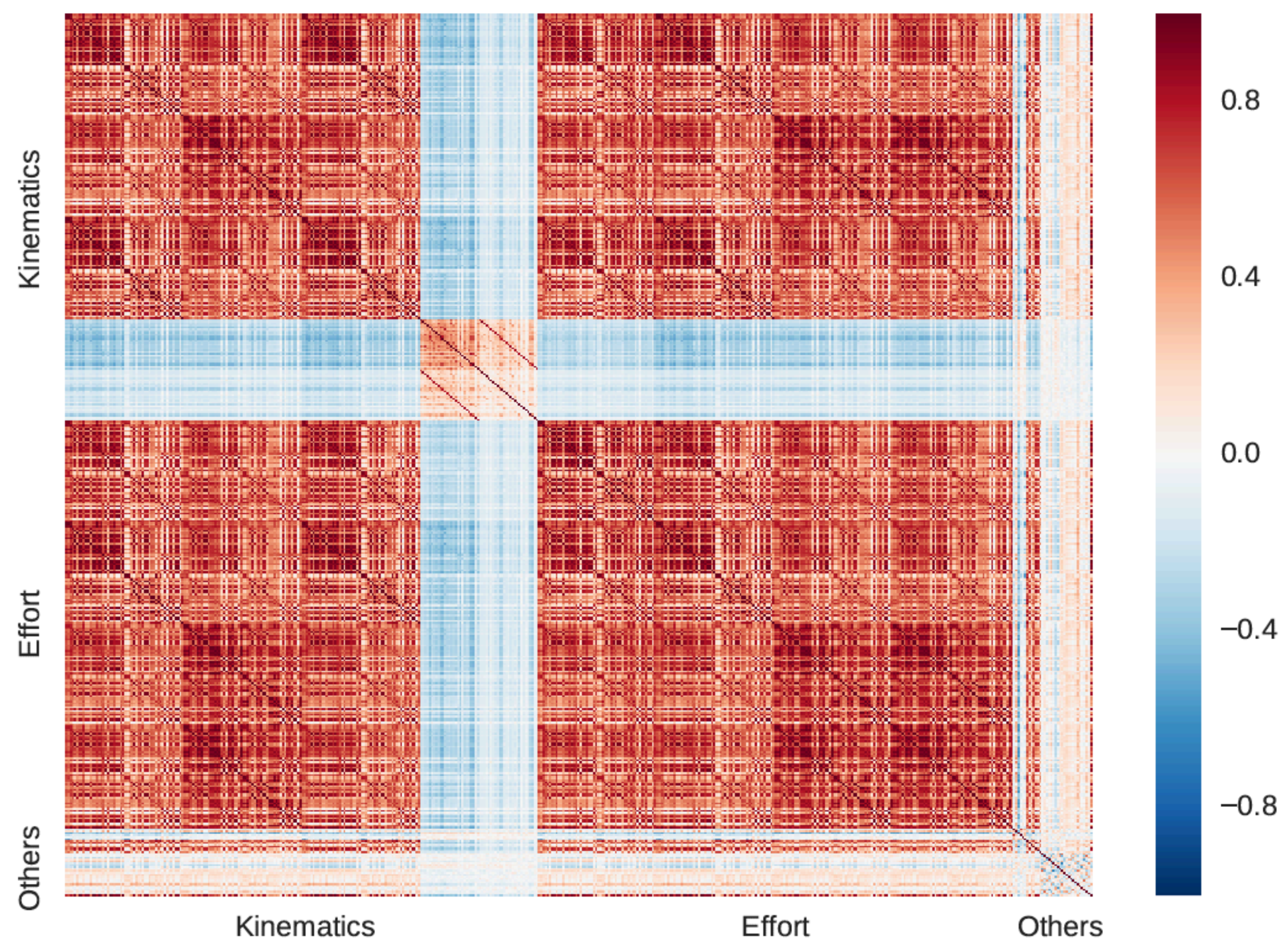
Discussion: challenges

Evaluation [Carreno et al., ACII 2015]: segmentation into sub-actions - influence of the offset and the size of the window on the classification rates - affective classes (6+1 emotions: angry, happy, sad, surprise, stressed, relaxed)



Discussion: challenges

- **How to evaluate the set of descriptors**



- Correlation matrix (Carreno et al, 2017)

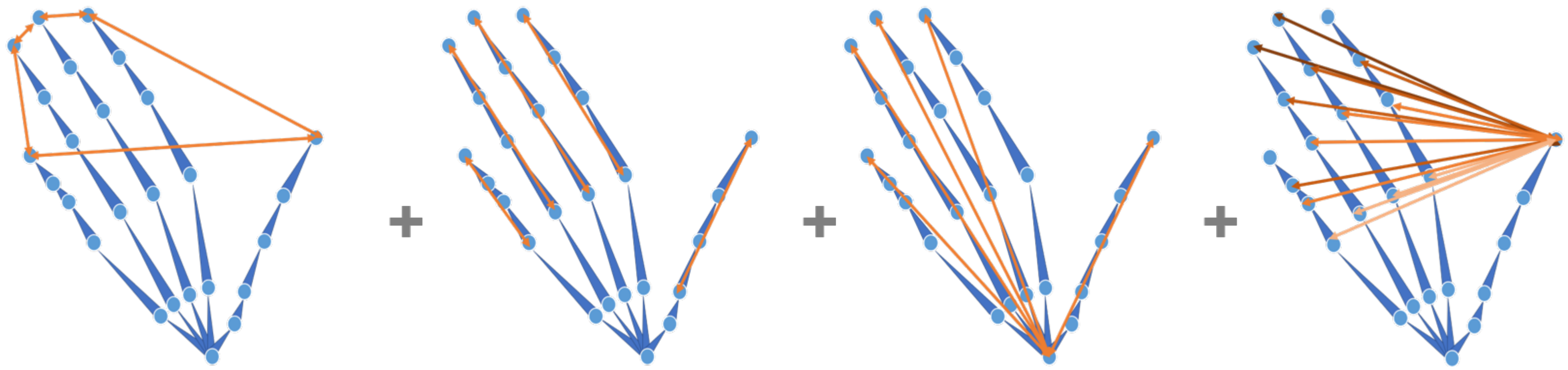
Discussion: challenges

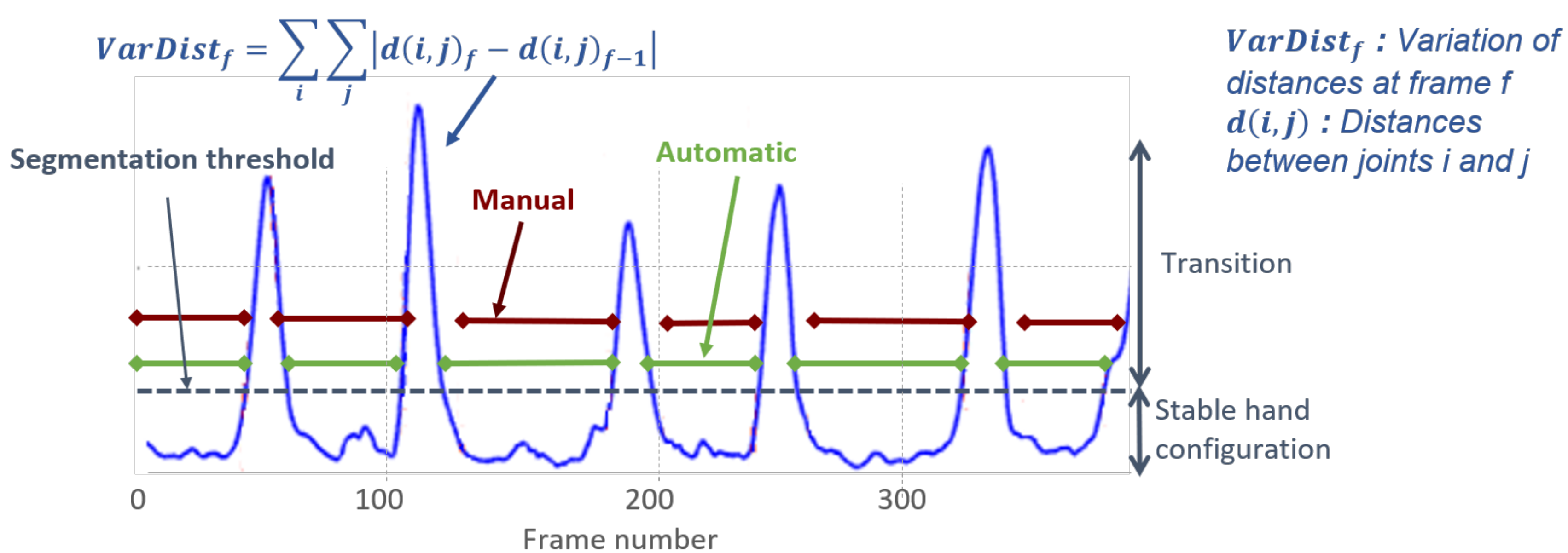
- **How to evaluate the set of descriptors**
  - Selecting methods to extract the more relevant descriptors (random forests)
  - CNN networks
  - Perceptual studies

Examples

Hand descriptors (used for handshape recognition in sign languages)

Which descriptors



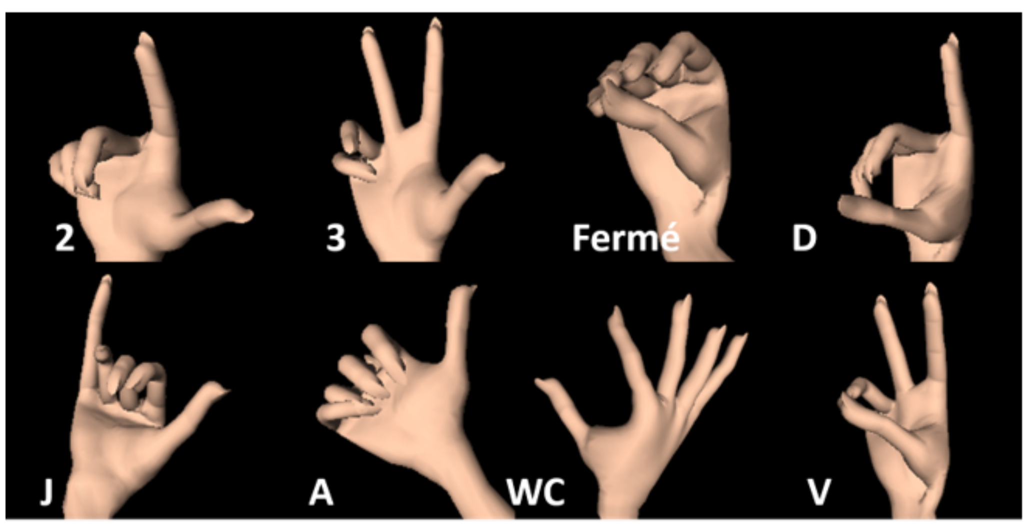


Automatic segmentation

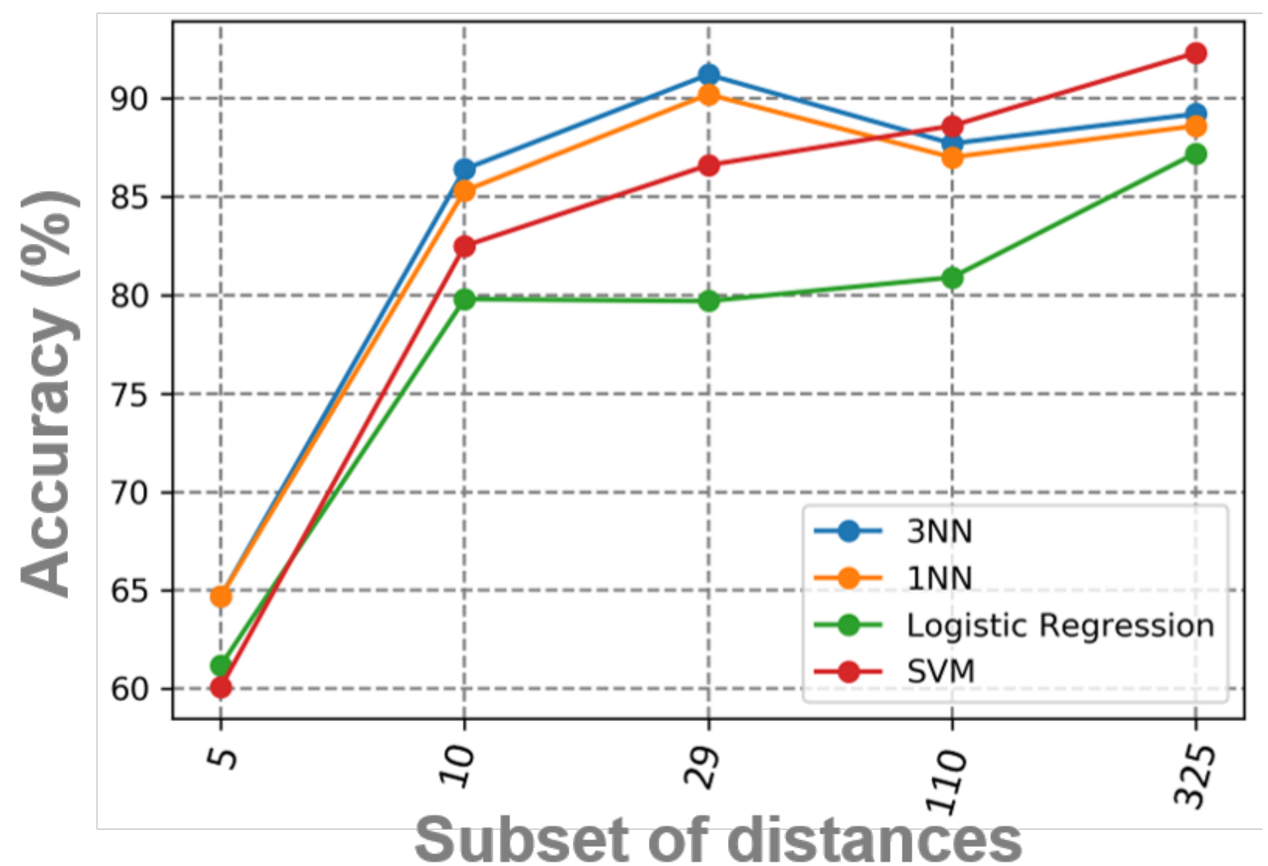
Automatic Annotation of Hand Configurations

- Recognition through machine learning (K-NN, LR, RF): accuracy results

32 hand configurations



...



EXPRESSION
Expressiveness in Human Centered Data/Media

Automatic Annotation of Hand Configurations

Accuracy = 86.2%
Precision = 84.7 %
Recall = 91.3 %
F-score = 87.9 %

"The entrance fee is 2.5 euros"

Accuracy = 81.4%
Precision = 75.2 %
Recall = 91.3 %
F-score = 82.4 %

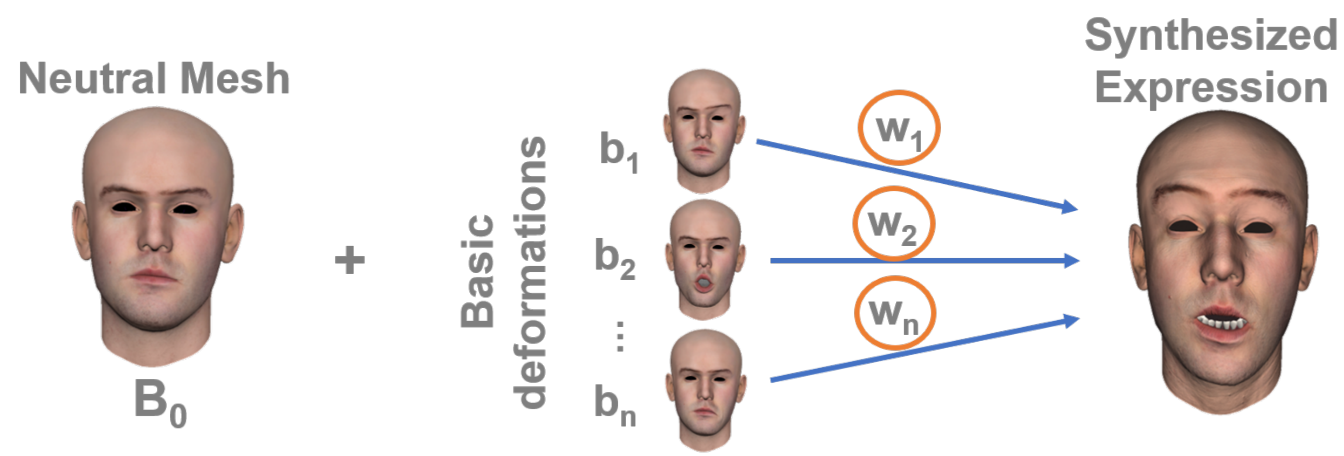
"The child is named Antoine. He lives with his parents"

3NN classification on 29 distances

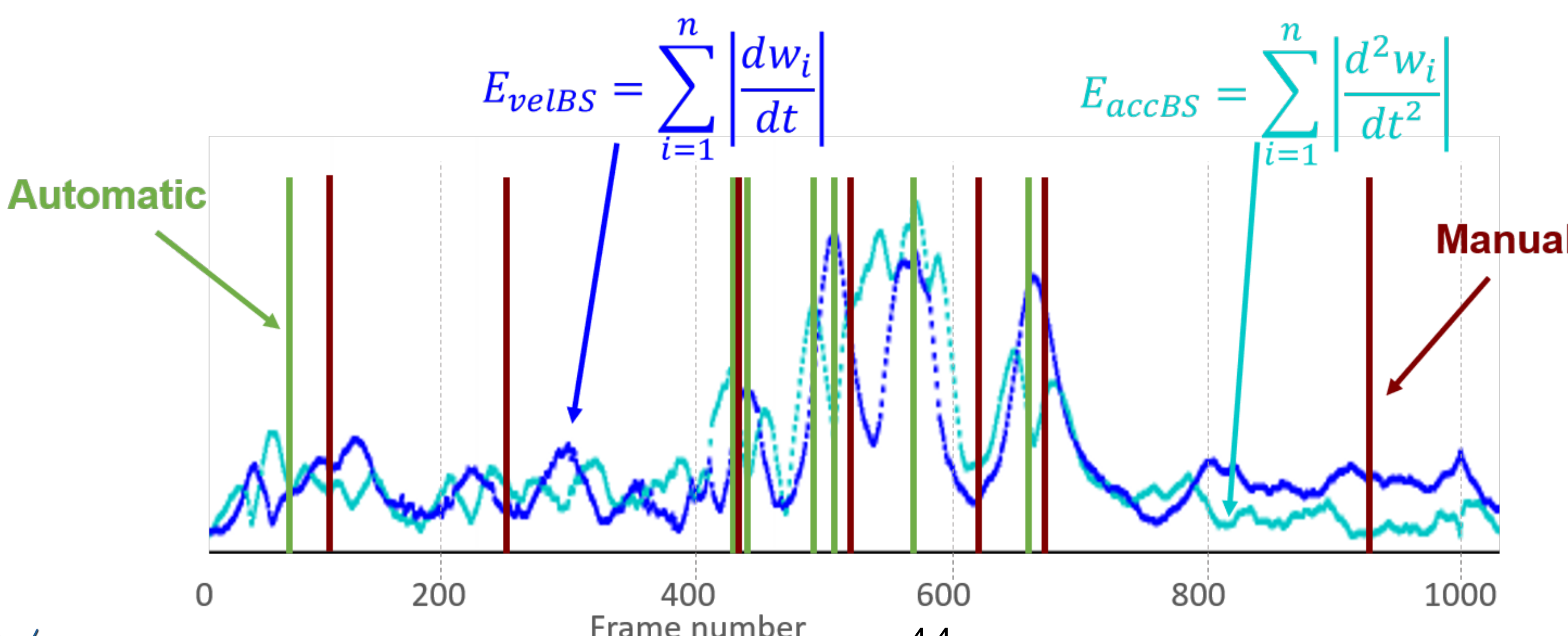
Lucie Naert, Clément Reverdy, Caroline Larboulette, Sylvie Gibet. Per Channel Automatic Annotation of Sign Language Motion Capture Data. Workshop on Representation and Processing of Sign Languages: Involving the Language Community, Miyazaki, Japan, 2018.

Facial descriptors

- Which descriptors?



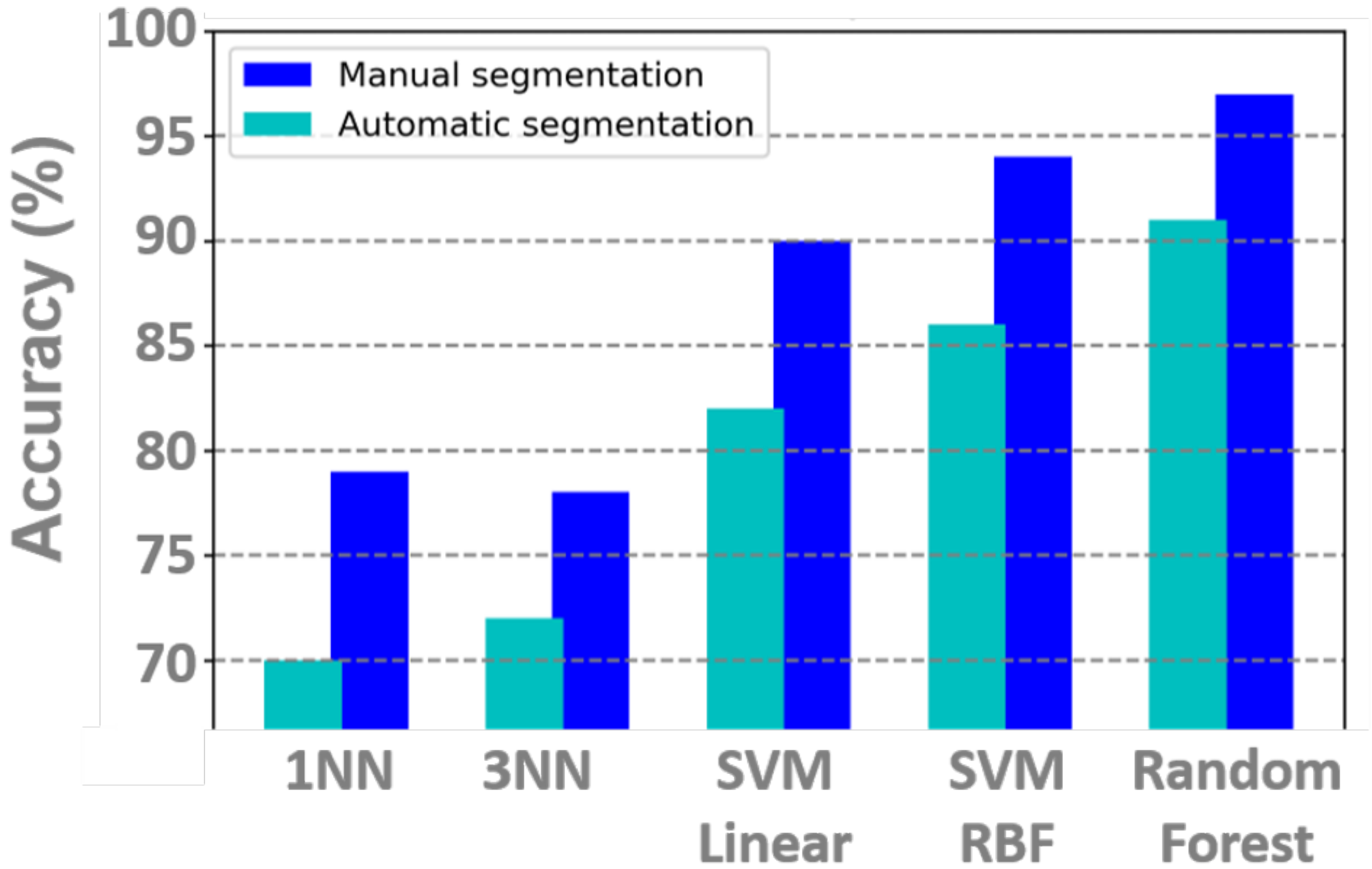
- Automatic segmentation: variations of distances



EXPRESSION
Expressiveness in Human Centered Data/Media

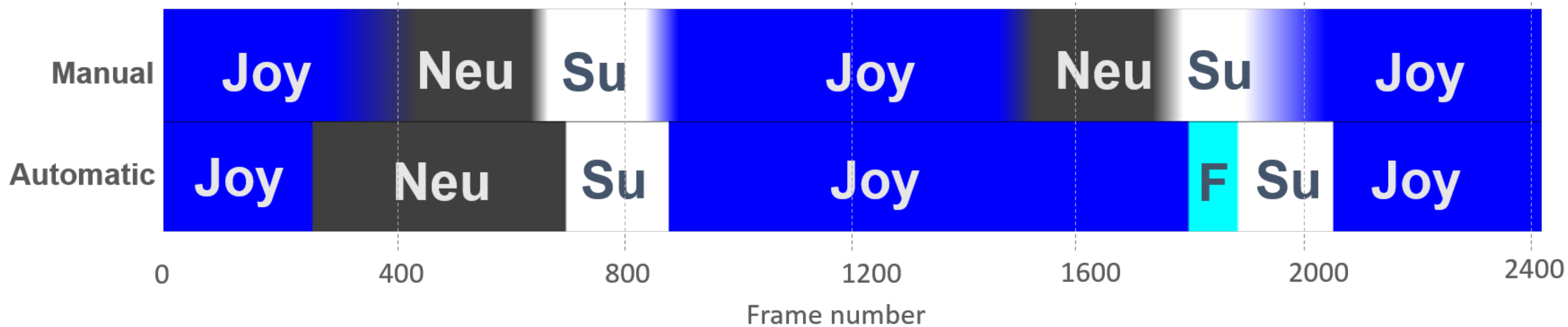
Facial descriptors

- Recognition through machine learning: accuracy results



Facial descriptors

- Annotation results



Expressive motion and descriptors

Expressive motion

- Combination of meaning, style, expressiveness
  - Meaning: linguistic elements or significant actions (narrative scenarios, signs or sentences in SL, etc.)
  - Style: identity (morphology, gender, personality, etc.)
  - Expressiveness: nuances superimposed on the motion, guided by the emotional context or with a willful intent (improvisation theater, gestures / postures to show the emotion)
- Non-separable components
  - difficult to specifically characterize their influence on the whole performance

Expressive motion

- **No consensus on which descriptors characterize best motion, quantitatively and qualitatively**
    - Early work: inspired from linguistic or psychological studies [Wallbott 1998], [Gallaher 1992]
    - Notation systems provide efficient means to describe motion properties, both semantically, structurally, and in details
        - **Verbal labels** (Kendon 1980, McNeill 1992): define semantic categories for co-verbal gestures
        - **Structural notations**: Laban Movement Analysis (LMA) expressed for dance and everyday life movements
        - **Linguistic notations**: phonetics, phonological elements, syntactic structure, prosody for sign languages

Expressive motion

- **Partial body measures (hand-arm movements,…)**
  - **Based on LMA theory**
    - EyesWeb system [Camurri et al., 2004]: computer vision
    - Laban Effort and Shape analysis of affective hand-arm movements [Samadani et al., 2013]: continous features in time
  - **Based on psychological studies (Wallbott, Gallaher)**
    - [Pelachaud et al., 2009], [Glowinski et al., 2011] (extending EyesWeb system)
  - **Based on spatial and temporal movement qualities**
    - [Alaoui et al., 2017]

Laban descriptors

High-level motion descriptors

- **Based on LMA**
  - Theory that covers most motion characteristics: describes the structural, geometric and dynamic properties of human motion, within the body and in relation to the surrounding environment
  - Four components
    - Body descriptors
    - Space descriptors
    - Shape descriptors
    - Effort descriptors

High-level motion descriptors: Body

- **Spatial characteristics of the motion:** determines which body parts are moving, what body parts support the body, ensure the balance
  - **Action presence**: detects if the displacement of the body segment crosses a threshold

$$Action^k(t_i) = ||\mathbf{x}^k(t_i) - \mathbf{x}^k(t_{i-1})|| > \epsilon \; ? \; 1 : 0$$

  - **Center of mass displacement:**

$$com\text{-}disp(t_i) = ||\mathbf{com}(t_i) - \mathbf{com}(t_{rest})||$$

  - **Balance:** boobean value indicating the relative location of the CoM with respect to the support polygon

$$\begin{aligned} Balance(t_i) = proj(&\mathbf{com}(t_i)) \in \\ &proj(\text{B-Shape}(\mathbf{x}^k(t_i))) \; ? \; 1 : 0 \end{aligned}$$

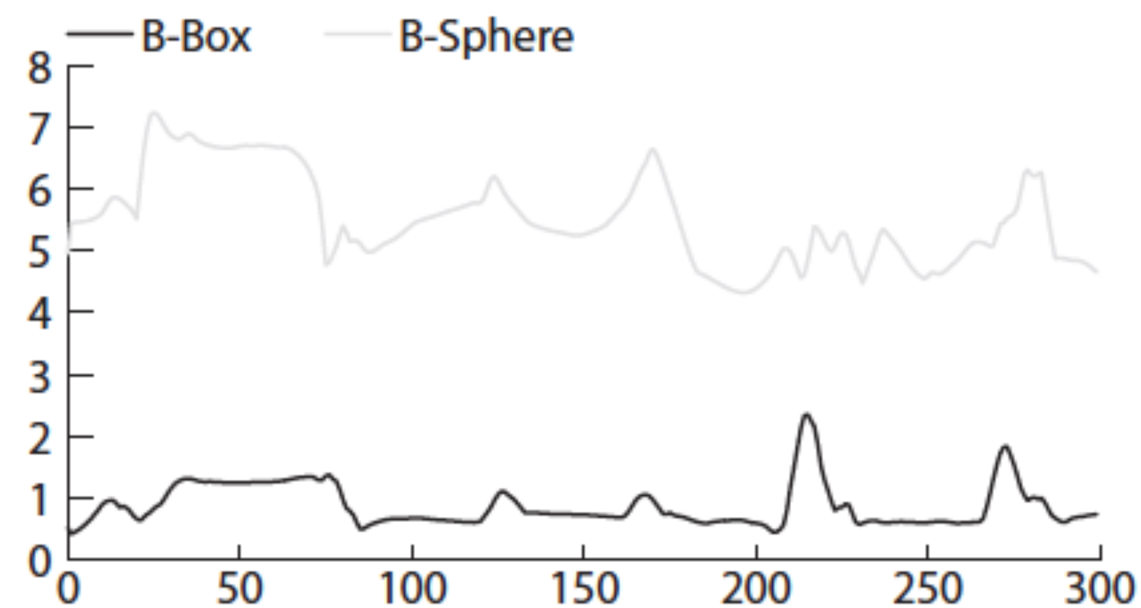
High-level motion descriptors: Space

- **Spatial characteristics of the motion:** describes the movement in relation with the 3D space surrounding the body
    - **Distance covered:** measures the total distance covered by the projection of the Root joint on the floor over a time period
    - **Area covered**: measures the area covered by the projection of the Root joint on the floor over a time period
    - **Hip height**: calculates the distance between the root joint and the ground

High-level motion descriptors: Shape

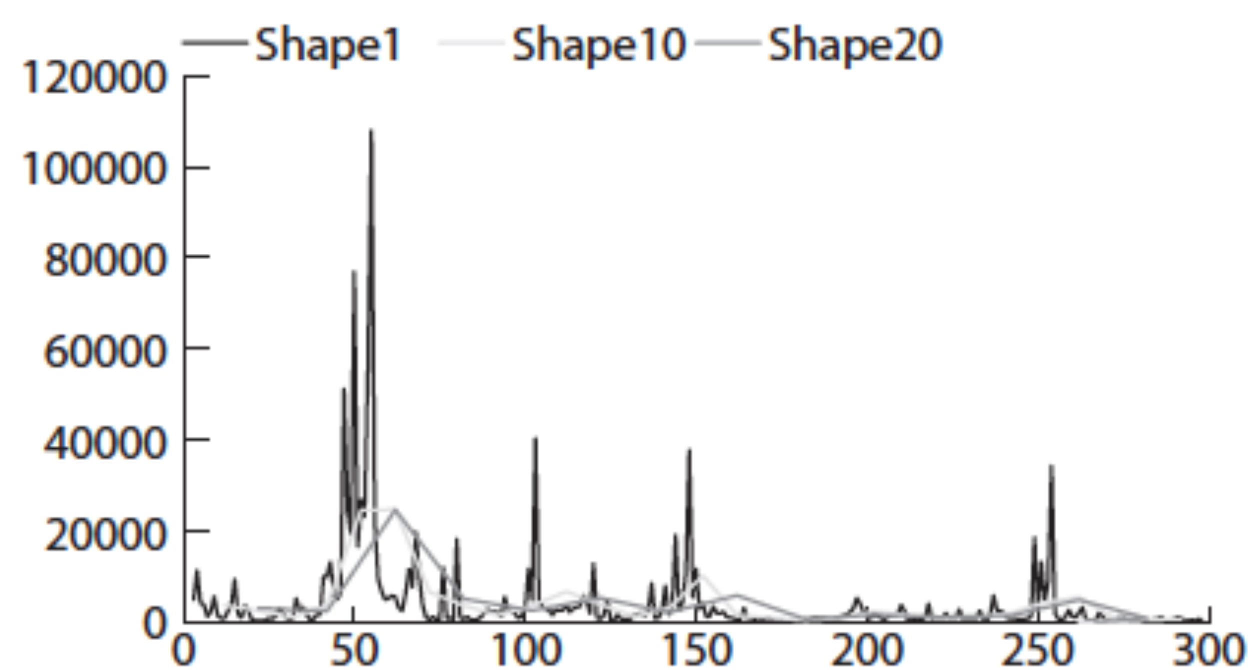
- Describes how the body changes shape during the movement
    - **Bounding volume**: computes the volume of the bounding shape

$$ShapeV^k(t_i) = Volume(\textbf{B-Shape}(\mathbf{x}^k(t_i)))$$



- **Shape directional**: determines the shape of the path along which a motion is executed

$$ShapeD^k(T) = \frac{1}{T}\sum_{i=1}^{T} C^k(t_i)$$



High-level motion descriptors: Shape

- Describes how the body changes shape during the movement

    - **Extensiveness:** maximum distance between the center of mass and the set of end-extremities (hands, feet, shoulders, head)

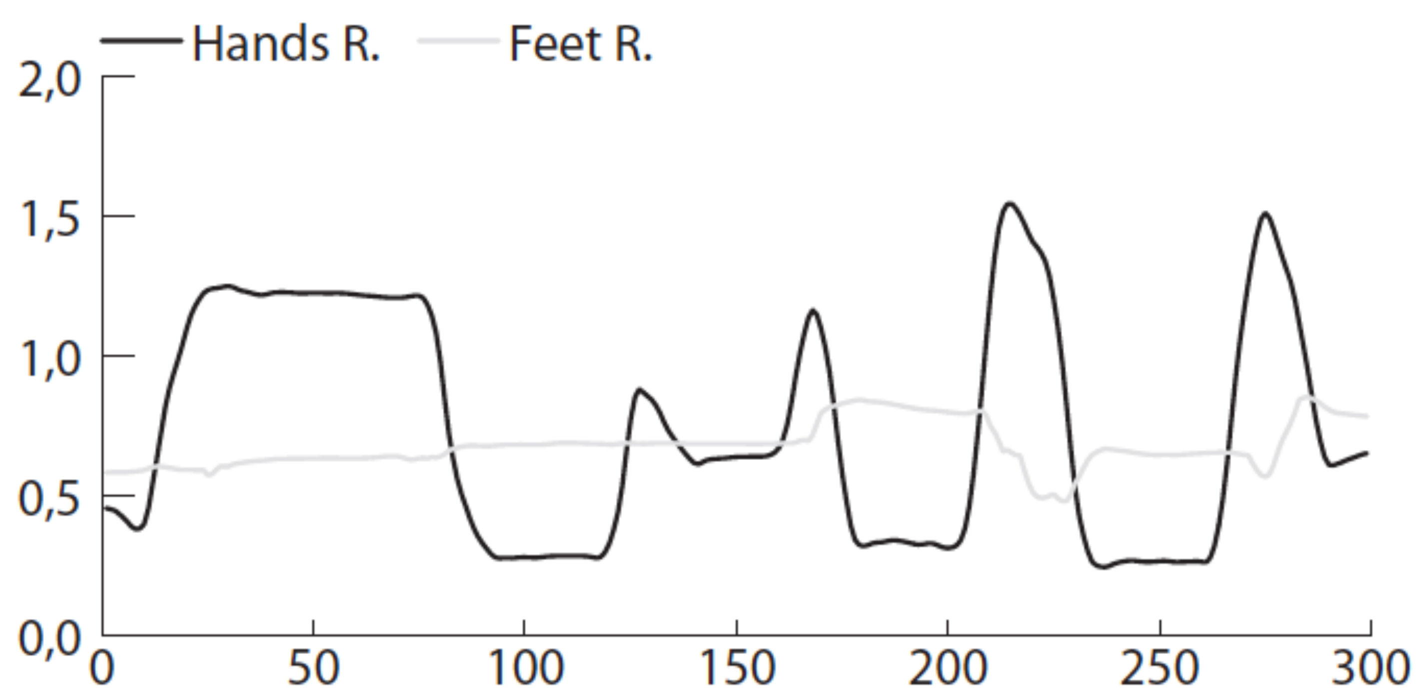
$$Extensiveness_m(t_i) = \max_k \alpha^k . \|\mathbf{com}(t_i) - \mathbf{x}^k(t_i)\|$$

    - or weighted sum of these distances

$$Extensiveness_w(t_i) = \sum_{k \in \mathbb{K}} \alpha^k . \|\mathbf{com}(t_i) - \mathbf{x}^k(t_i)\|$$

High-level motion descriptors: Shape

- Describes how the body changes shape during the movement
  - **Hands and feet relationship**: computes 3D distance between hands (respectively feet)



High-level motion descriptors: Effort

- **Focuses on the quality of motion** in terms of dynamics, energy and expressiveness
    - Spatio-temporal descriptors
    - Comprises 4 sub-categories which vary continuously in intensity between opposite poles
        - **Weight Effort**
        - **Time Effort**
        - **Space Effort**
        - **Flow Effort**

High-level motion descriptors: Weight Effort

- Characterizes physical properties of motion

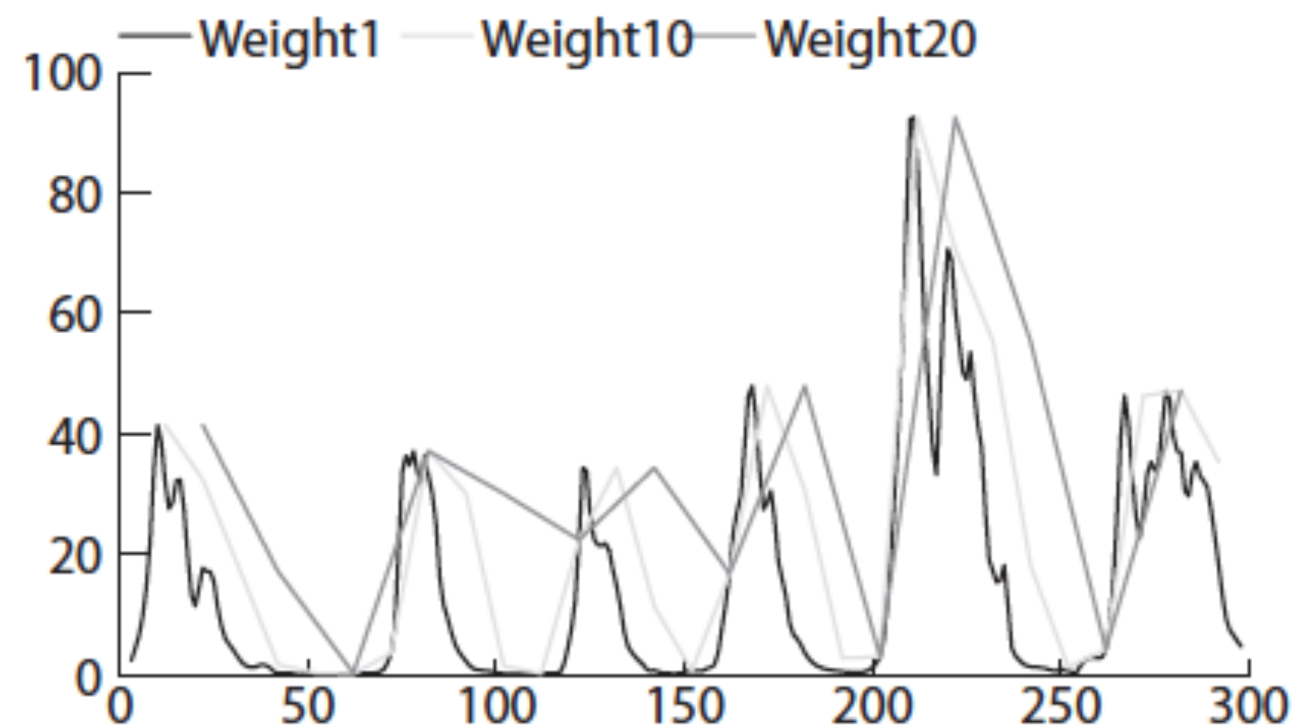
Strong (powerful, forceful)

vs.

Light (gentle, delicate, sensitive)

$$E(t_i) = \sum_{k \in \mathbb{K}} E^k(t_i) = \sum_{k \in \mathbb{K}} \alpha_k . v^k(t_i)^2$$

$$Weight(T) = \max E(t_i), i \in [1, T]$$



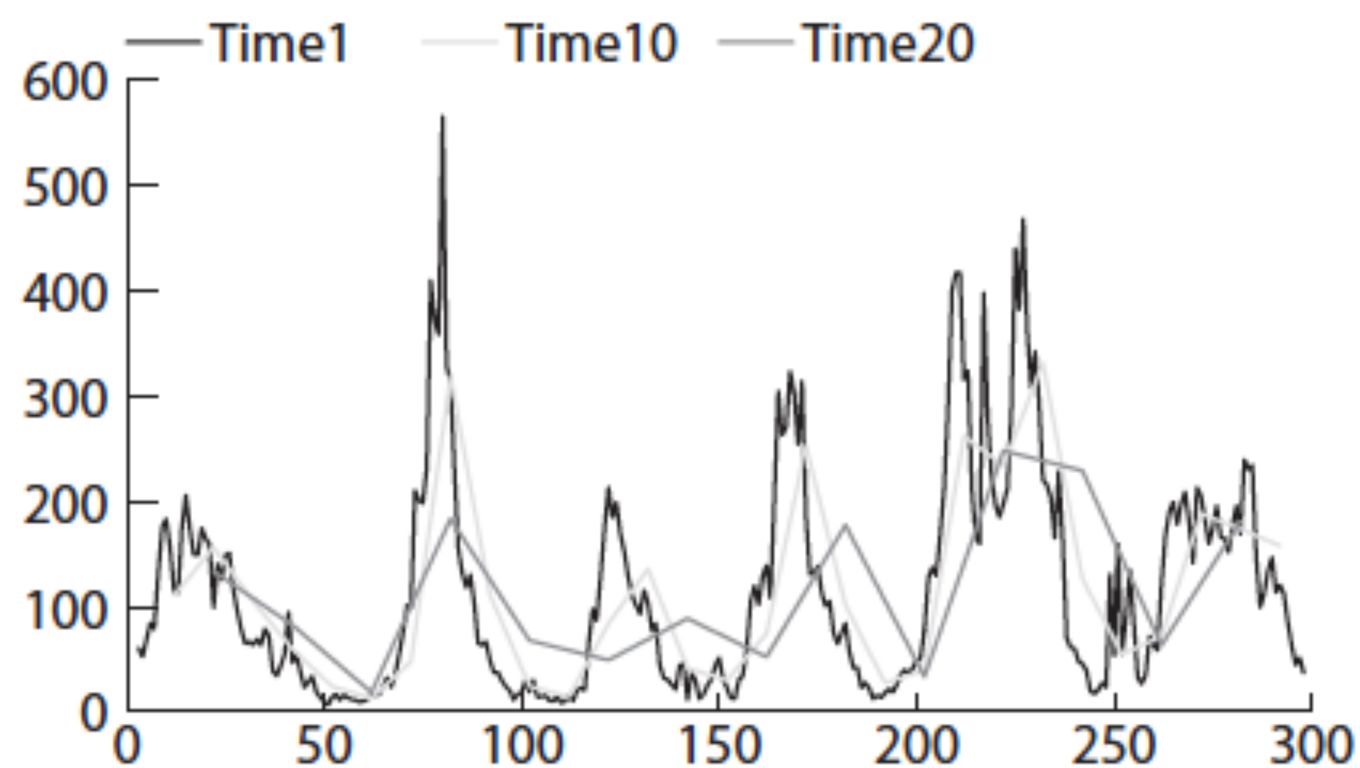
High-level motion descriptors: Time Effort

- Represents the sense of urgency

Sudden (urgent, quick)

vs.

Sustained (stretching the time, steady)



$$Time^k(T) = \frac{1}{T}\sum_{i=1}^{T} a^k(t_i)$$

$$Time(T) = \sum_{k\in\mathbb{K}} \alpha_k . Time^k(T)$$

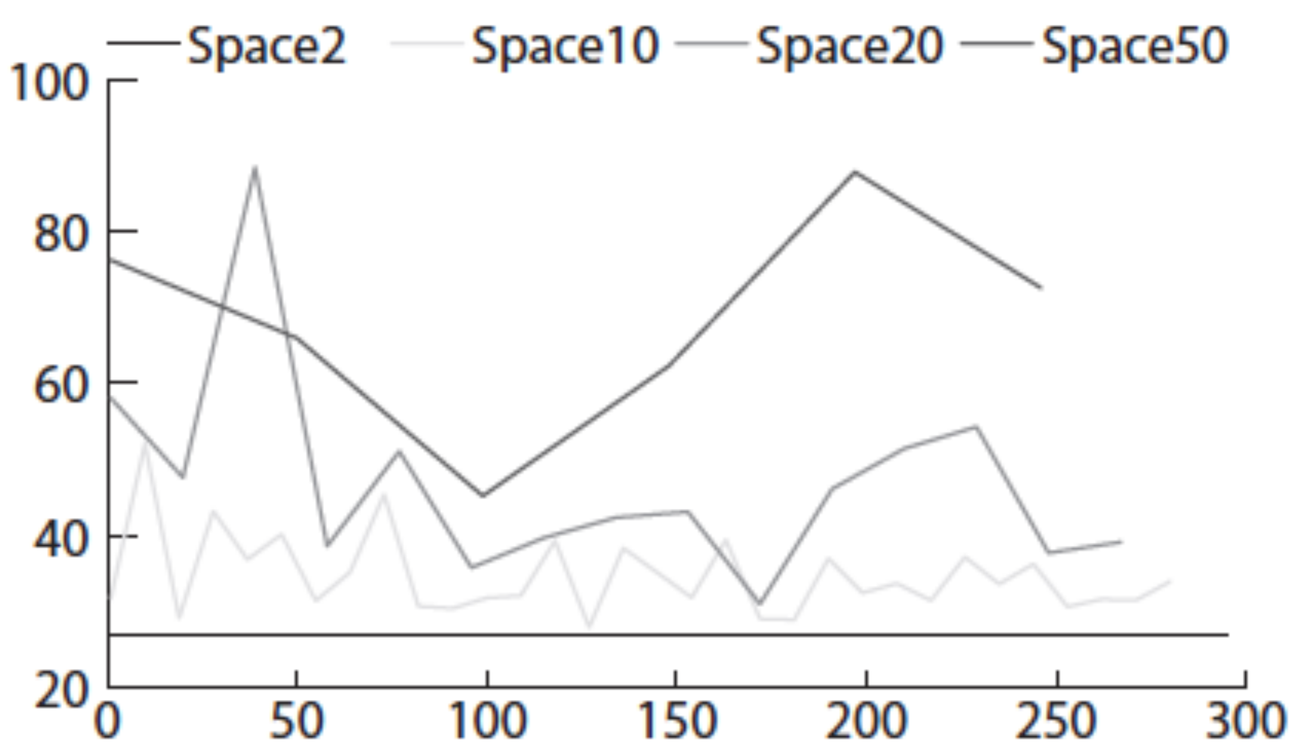
High-level motion descriptors: Space Effort

- Defines the directness of the motion, related to the attention to the surroundings

Direct (focused, toward a particular spot)

vs.

Indirect (multi-focused, flexible)



$$Space^k(T) = \frac{\sum_{i=2}^{T} \|x^k(t_i) - x^k(t_{i-1})\|}{\|x^k(T) - x^k(t_1)\|}$$

$$Space(T) = \sum_{k \in \mathbb{K}} \alpha_k . Space^k(T)$$

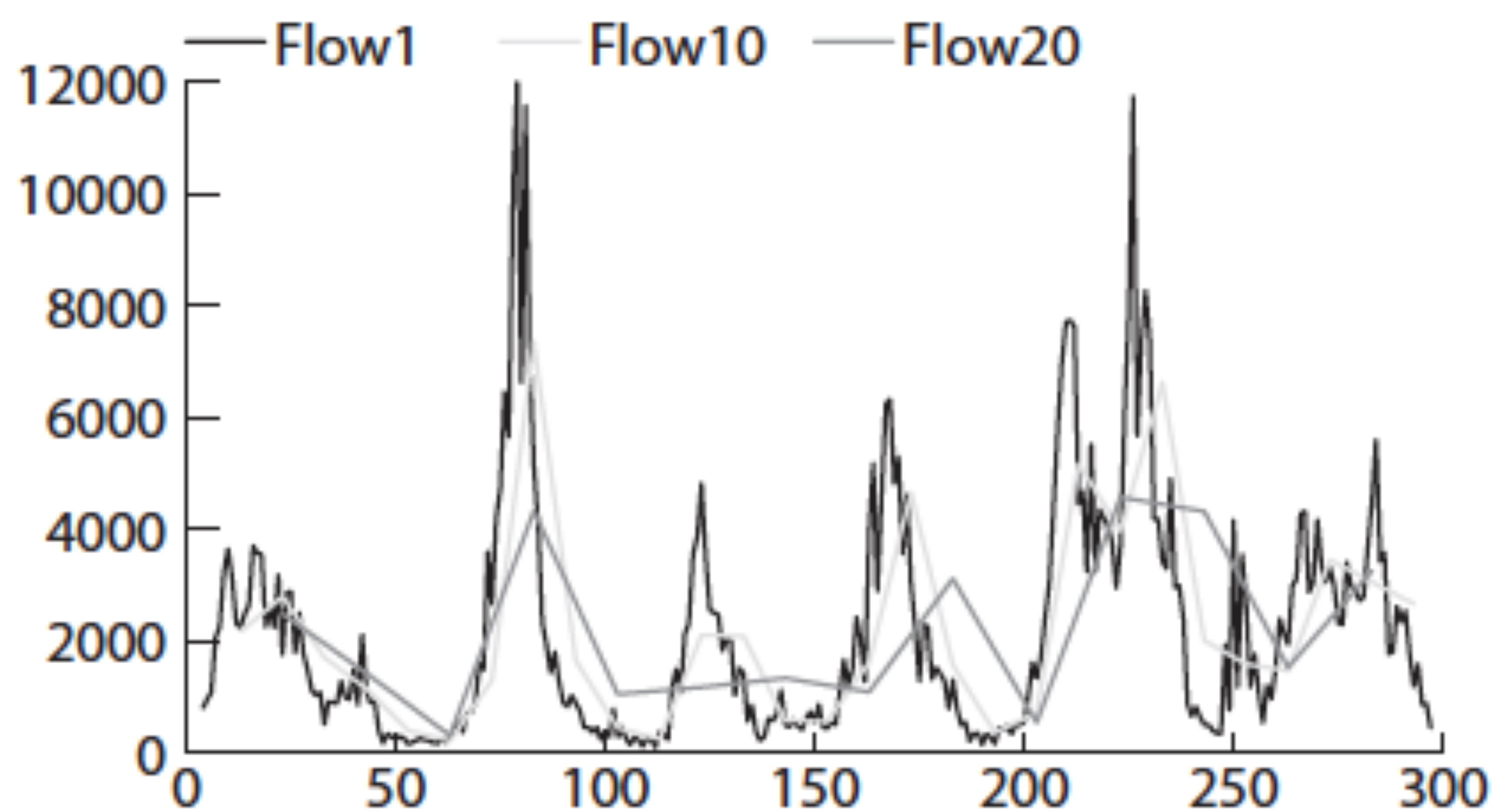
High-level motion descriptors: Flow Effort

- Defines the continuity of the motion

  Free (fluid, released)

  vs.

  Bound(controlled, careful, restrainted)



$$Flow^k(T) = \frac{1}{T}\sum_{i=1}^{T} j^k(t_i)$$

$$Flow(T) = \sum_{k\in\mathbb{K}} \alpha_k.Flow^k(T)$$