Communication with Emotional Body Language

COmmunication with Emotional BOdy Language (COBOL) was launched in 2006 (3 year project) by the the European Commission as project, in the 6th EU framework programme.

Tilburg University
Prof. Beatrice de Gelder
Dr. Geert van Boxtel

Hertie Institute for clinical brain research:
Prof. Martin Giese
Prof. Peter Thier
Prof. Wolfgang Strasser

Eidgenössische Technische Hochschule Zürich
Prof. Luc van Gool
Dr. Beat Fasel
Dr. Konrad Schindler

Weizmann Institute, IL
Prof. Tamar Flash
Dr Talma Hendler
Prof. Rafael Malach
Prof. Michal Irani

CNRS - Collège de France
Prof. Alain Berthoz
Dr. Julie Grèzes

Communication with Emotional Body Language

COmmunication with Emotional BOdy Language (COBOL) was launched in 2006 (3 year project) by the the European Commission as project, in the 6th EU framework programme.

Hertie Institute for clinical brain research:
Prof. Martin Giese
Prof. Peter Thier
Prof. Wolfgang Strasser

Hertie-Institut
für klinische Hirnforschung

Some Objectives

Workpackage 1
Description and analysis of the kinematic and dynamical structure of EBL
Workpackage 2
Development of EBL avatars and measurement of EBL perception and recognition
Workpackage 3
The cognitive basis of EBL
Workpackage 4
Coordinating social interactions

Blind shift invariant feature learning with applications to image processing and extraction of emotion from gait





Lars Omlor
Claire Roether
Aee-Ni Park
Albert Mukovskiy
Martin A. Giese

Hertie Institute for Clinical Brain Research, University Clinic Tübingen, Germany

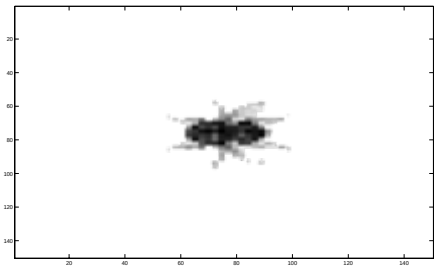
Overview:

1. Introduction to shift invariant feature extraction in image processing
   - The anechoic mixture model
   - Derivation of the algorithm (technical part)
2. Analysis of human motion
   - Motivation
   - Extraction of emotion specific features
3. Synthesis of human motion
   - Applications in computer graphics

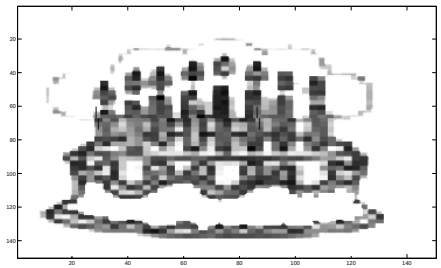
Part I: Shift invariant feature extraction and image processing

What is shift invariant feature extraction?

Generation of simple example:



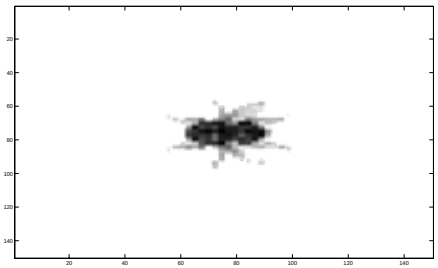
Picture I



Picture II

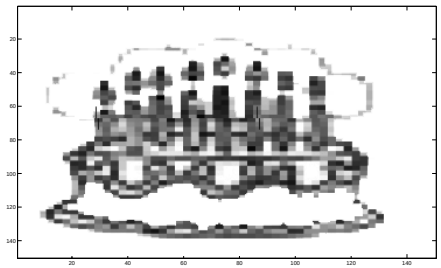
What is shift invariant feature extraction?

Generation of simple example:



Picture I

+



Picture II

=

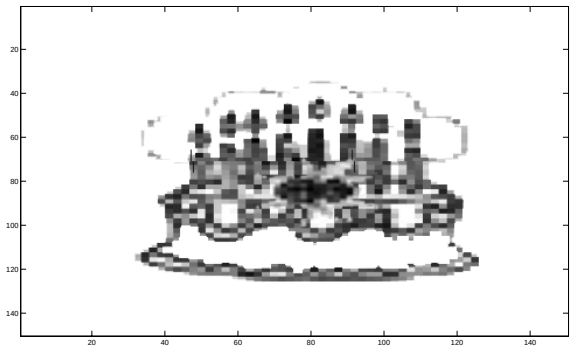
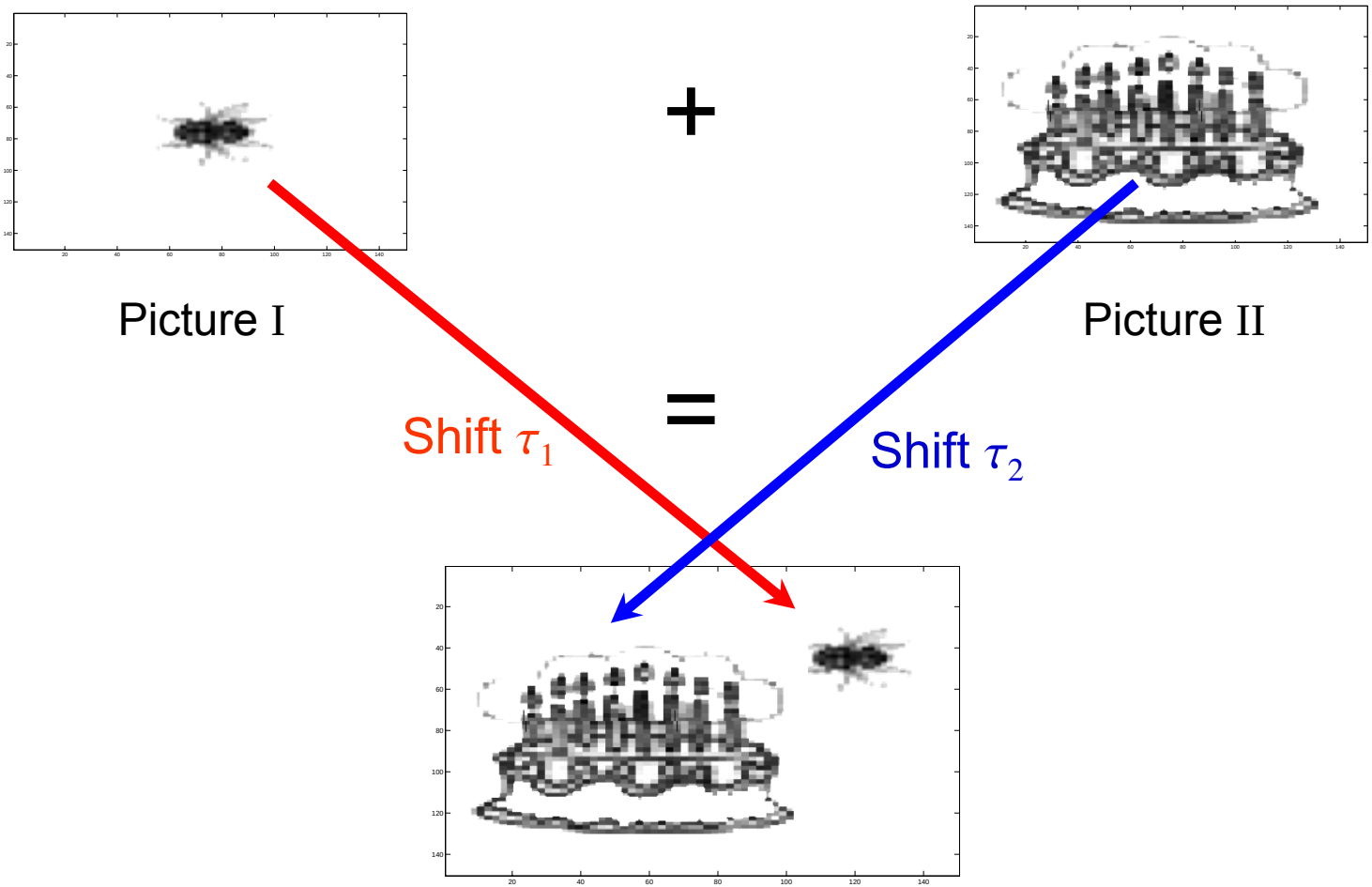


Image mixture

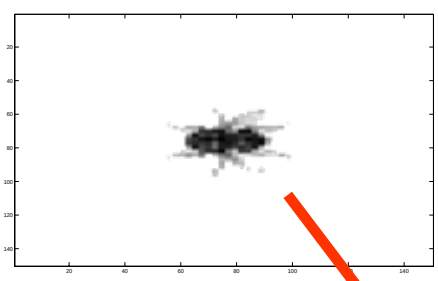
What is shift invariant feature extraction?

Generation of simple example:



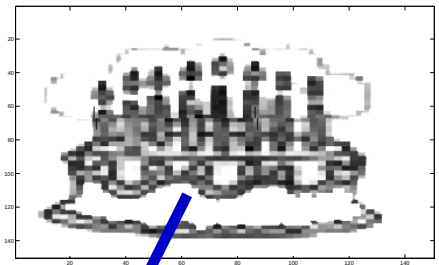
What is shift invariant feature extraction?

Generation of simple example:



Picture I

+



Picture II

=

Shift τ_1

Shift τ_2

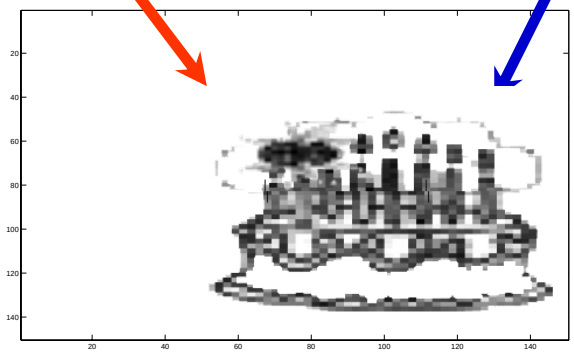
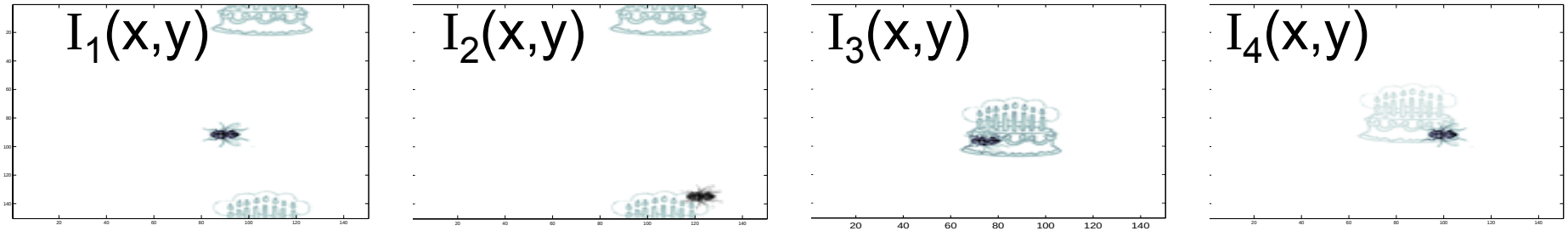


Image mixture
(alternative location)

What is shift invariant feature extraction?

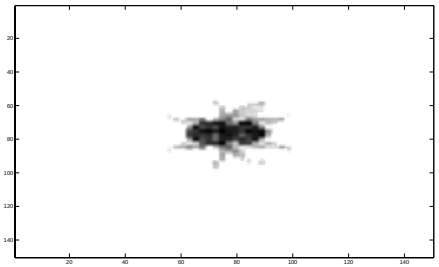
Task:

Input

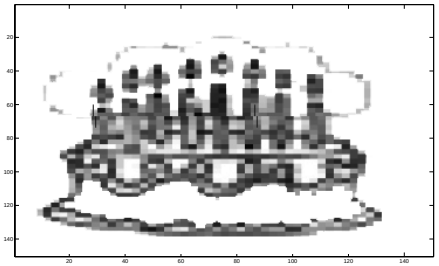


Change in absolute and relative position!

Unsupervised feature extraction



Picture I



Picture II

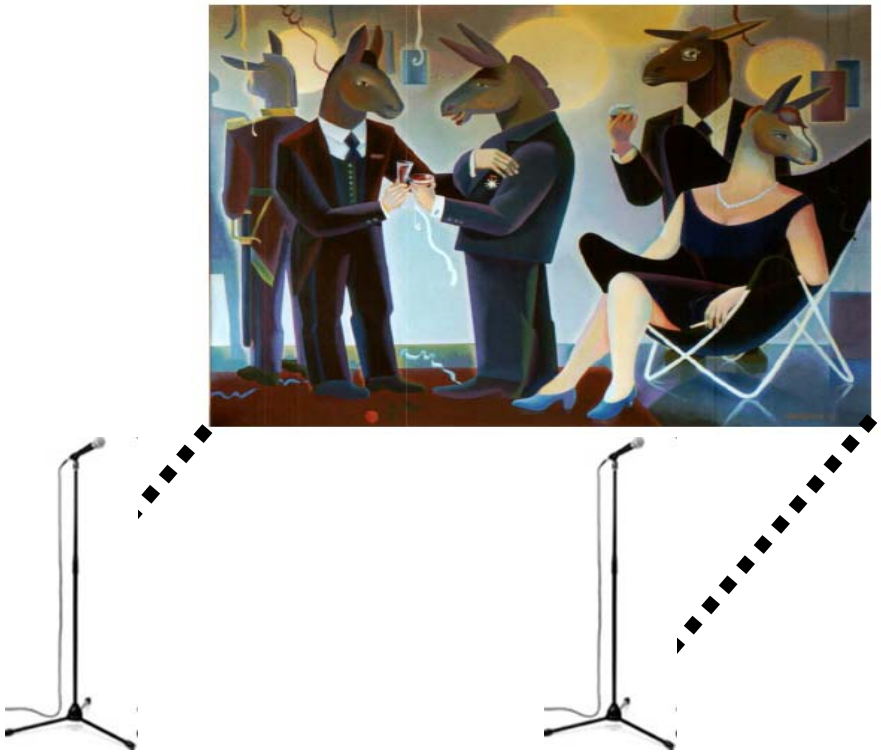
Classical field of application:
“Cocktail Party problem”



Several people are talking simultaneously (cocktail party), and one is trying to separate all of the discussions!

- Typical blind source separation problem
- Modeled with several possible degrees of realism

Classical field of application:
“Cocktail Party problem”



Simplest model:
All sounds reach the microphones simultaneously

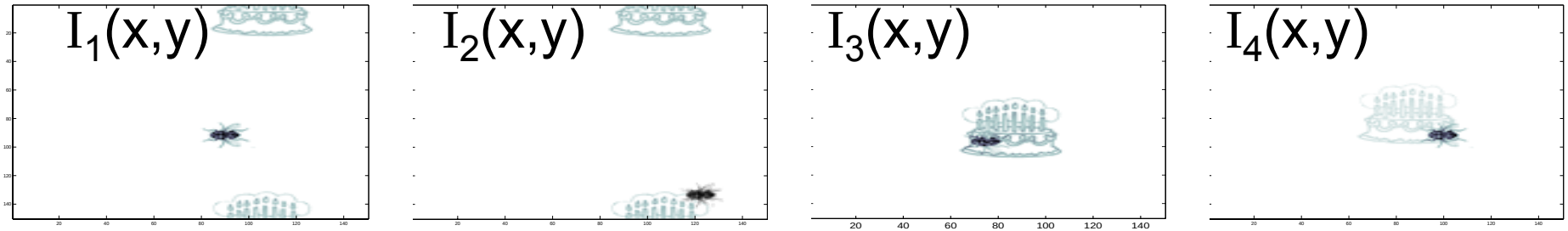
$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(t)$$

instantaneous mixture model (e.g. PCA)

What is shift invariant feature extraction?

Application of standard method (PCA):

Input



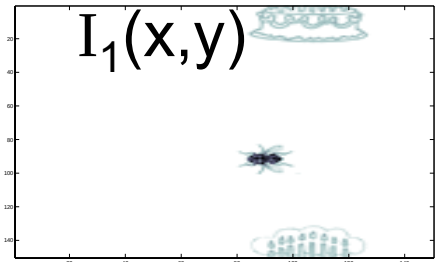
PCA

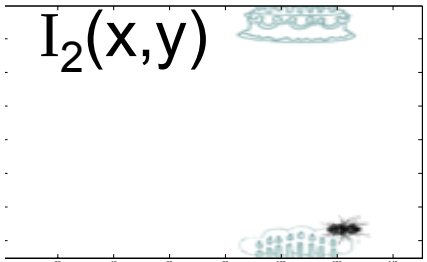
$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x,y)$$

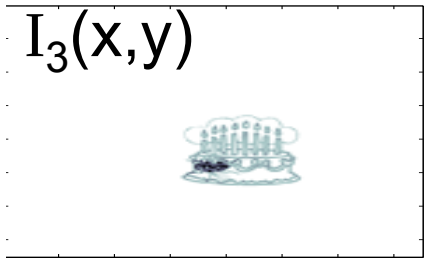
What is shift invariant feature extraction?

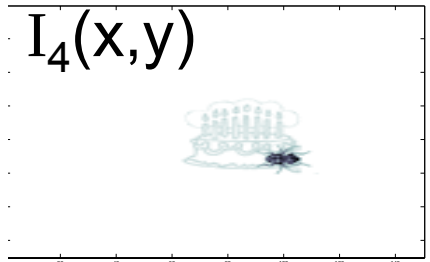
Application of standard method (PCA):

Input





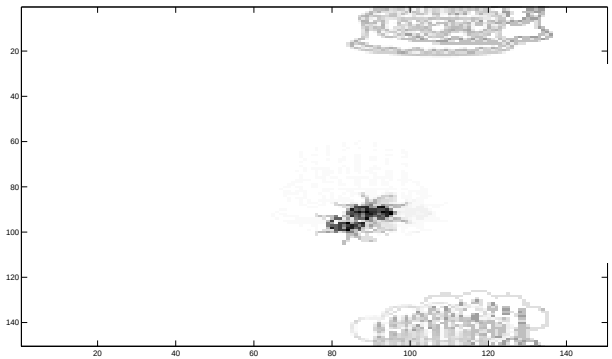


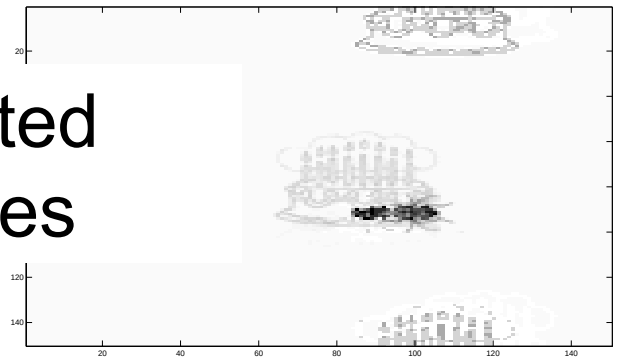


PCA

$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x,y)$$

Extracted features





Why is the PCA result so bad?

Assumed model (matrix notation):

$$I = AS$$

Why is the PCA result so bad?

Assumed model (matrix notation):

$$I = AS$$

$$\Rightarrow S_{est} = A_{est}^{-1} I$$

Extracted features are just linear combinations of the input

But we used shifts in the construction !

Why is the PCA result so bad?

Assumed model (matrix notation):

$$I = AS$$

$$\Rightarrow S_{est} = A_{est}^{-1} I$$

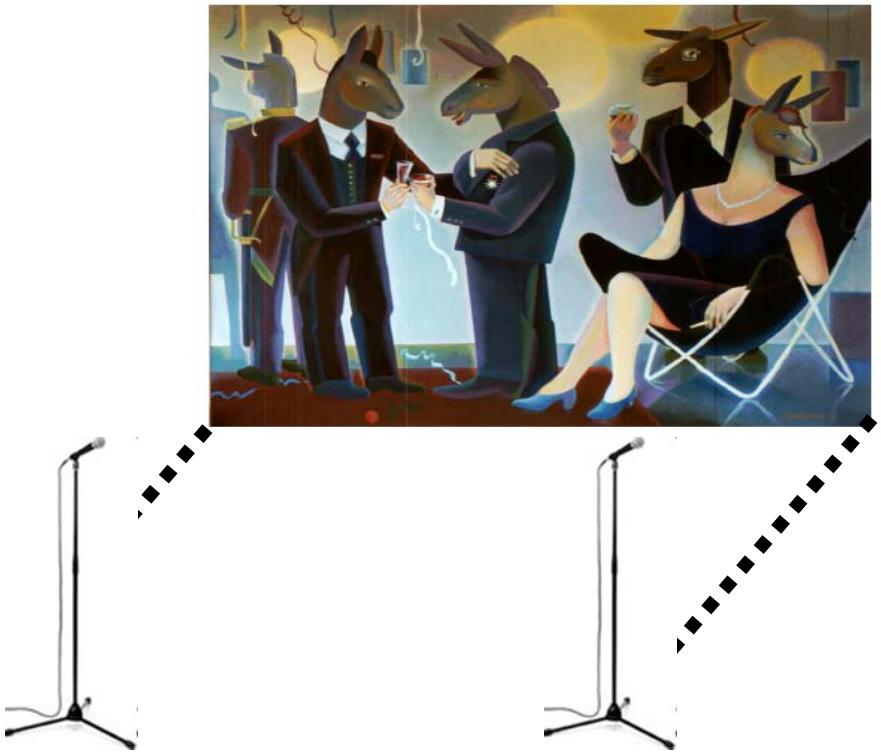
➡ Extracted features are just linear combinations of the input

But we used shifts in the construction !

➡ Better model:

$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x-\tau_{i1}, y-\tau_{i2})$$

Classical field of application:
“Cocktail Party problem”



Simplest model:
All sounds reach the microphones simultaneously

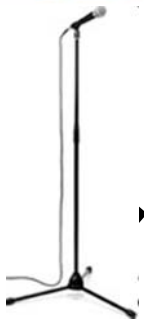
$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(t)$$

instantaneous mixture model

Classical field of application:

“Cocktail Party problem”







More realistic model:
Sounds reach the microphones with different traveling time, due to different distance to the speakers, but no reverberations

Classical field of application:
“Cocktail Party problem”



More realistic model:
Sounds reach the microphones with different traveling time, due to different distance to the speakers, but no reverberations

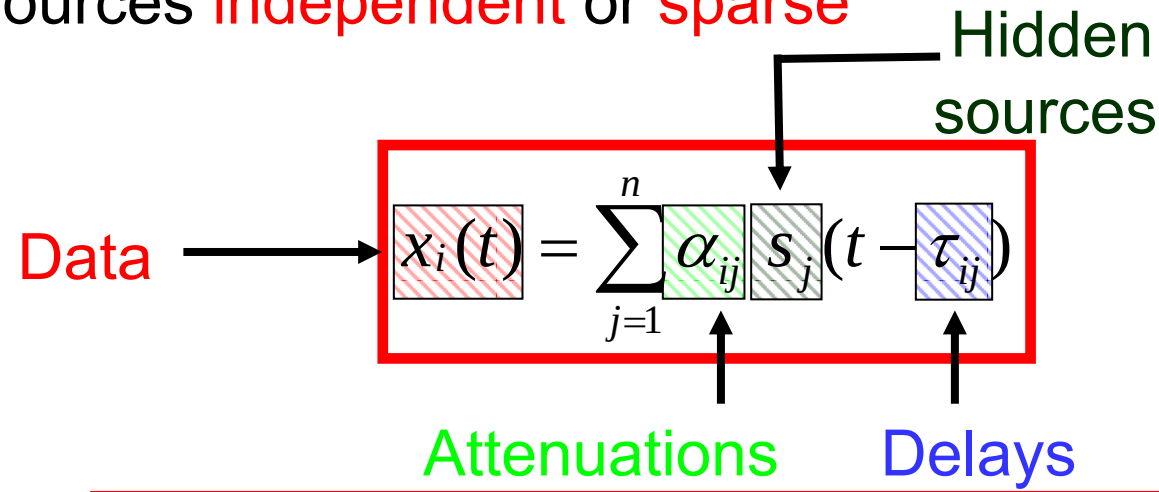


$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(t - \tau_{ij})$$

anechoic mixture model

A quick look at anechoic mixtures

- Linear unsupervised learning model
- Data is superposition of delayed and attenuated sources
- Classically less data then contributing sources (underdetermined)
- Sources independent or sparse



univariate

$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x - \tau_{i1}, y - \tau_{i2})$$

multivariate

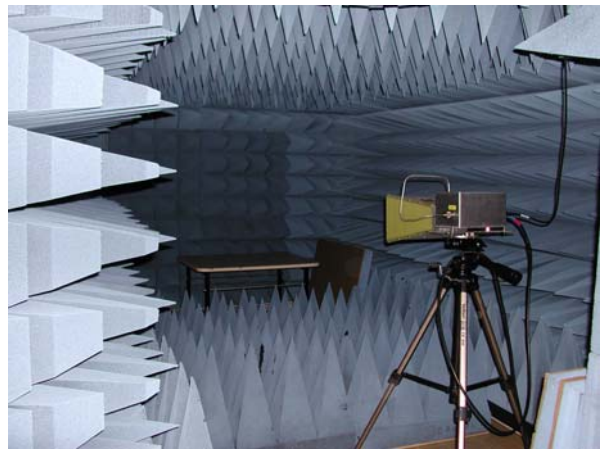
Classical anechoic mixtures: Simplified blind source separation model without reverberations

Underdetermined Problems:

- Sound separation

  Simple model for sound mixtures under the reverberation (echo) free condition.

  B. Emile, P. Comon (1998) Signal Processing 69
  K. Torkkola (1996) Proc. IEEE Int. Conf. Acoust., Speech, Sig. Proc. 96
  A. Yeredor (2003) Acoustics, Speech, and Signal Processing 5
  O. Ylmaz, S.Rickard (2004) IEEE Transactions On Signal Processing 52
  Bofill, P. (2003) Neurocomputing Vol. 55 627–641.



- Electrical engineering

  Signals from multiple antennas are received asynchronously

  A. Swindelhurst (1998) IEEE Trans. on Sig. Proc. no. 6



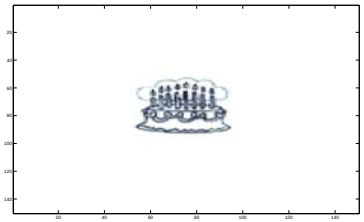
Overdetermined Problems:

- Human motion analysis

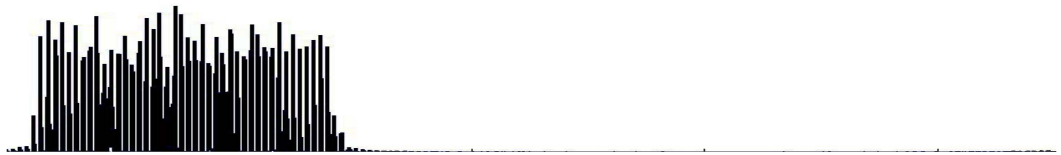
  L. Omlor, M. A. Giese (2007) Neural Information Processing Systems 19

How to deal with multivariate data?

Common trick in multivariate instantaneous mixtures:



vectorize

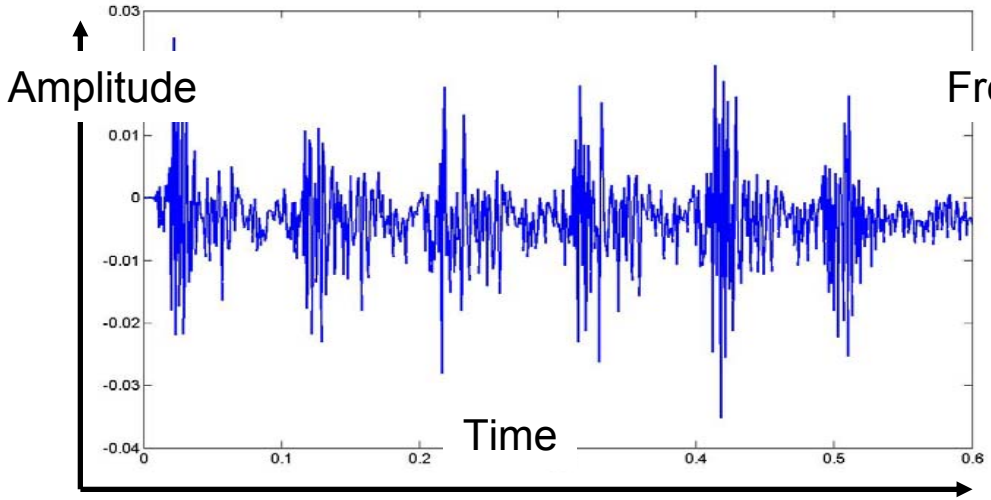


Does not work the anechoic case , because it messes up the order of the data !

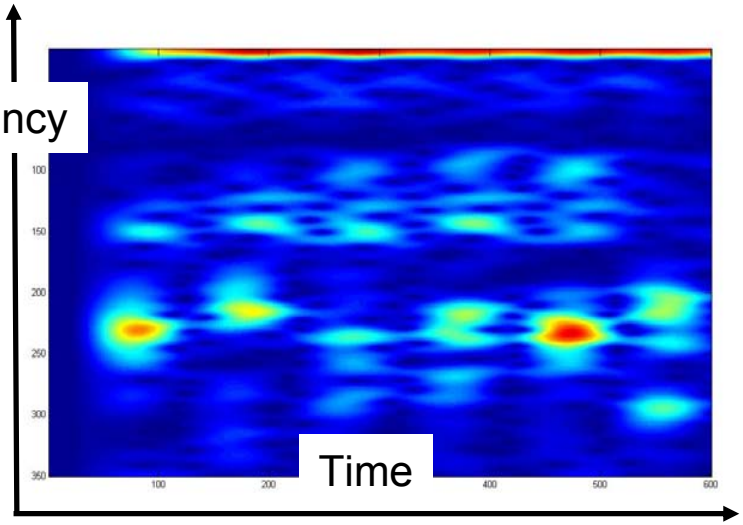
Idea: Find a transformation , which changes the multivariate problem into several univariate problems!

Short introduction into time-frequency methods:

Signal (Wave)

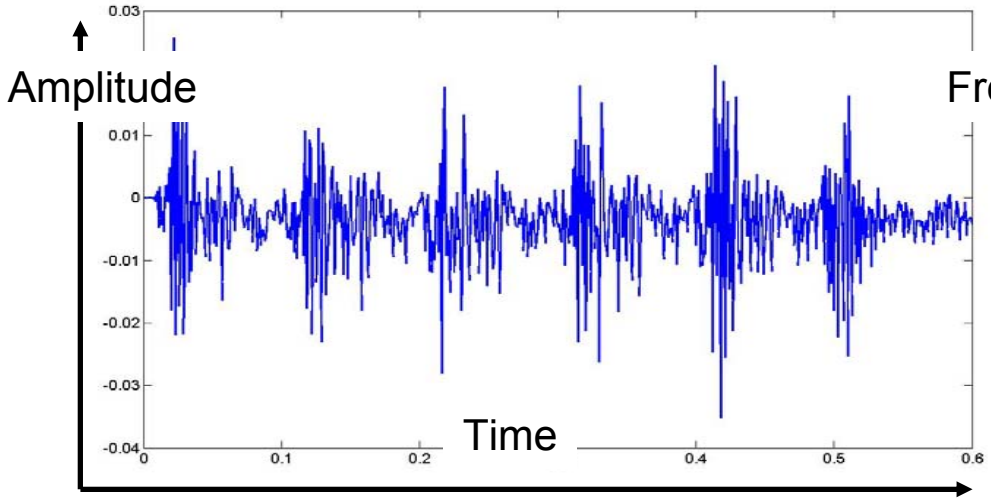


Time-Frequency representation

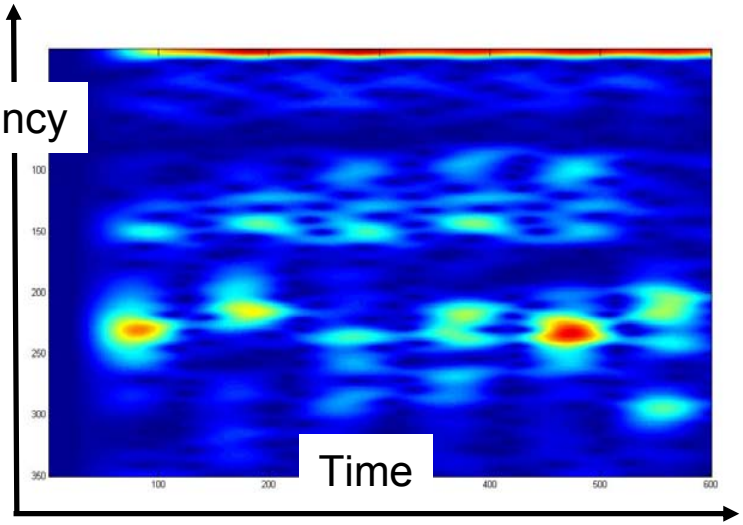


Short introduction into time-frequency methods:

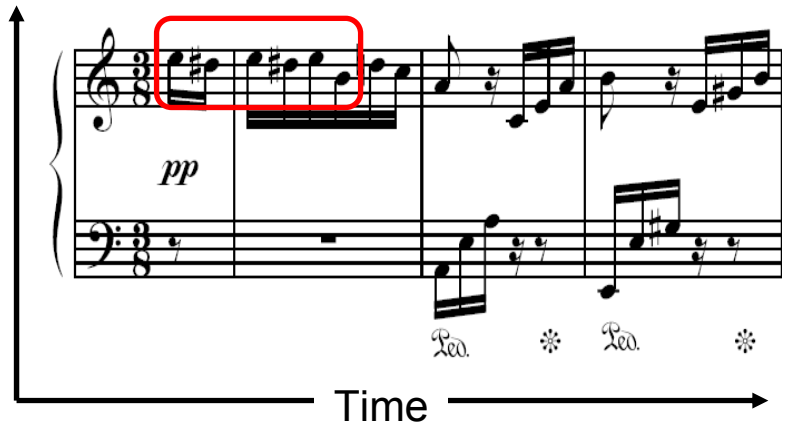
Signal (Wave)



Wigner-Ville-spectrum



Ludwig van Beethoven : For Elise



Short introduction into time-frequency methods:

Very powerful bilinear time-frequency distribution:

$$W_x(t, f) := \int_{\tau} E\{x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})\}e^{-2\pi i f\tau} d\tau$$

Wigner-Ville spectrum

"Loosely interpretable as a time-frequency distribution of the mean energy of x(t)"[1]

[1]G. Matz, F. Hlawatsch (2003) Wigner distributions (nearly) everywhere. Signal Processing 83, 1355-1378.

Short introduction into time-frequency methods:

Very powerful bilinear time-frequency distribution:

$$W_x(t, f) := \int_\tau E\{x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})\}e^{-2\pi i f\tau}\,d\tau$$

Wigner-Ville spectrum

Many mathematical properties:

- Marginal properties: $\int_t W_x(t, f)dt = E\{|Fx|^2\}$ $\int_f W_x(t, f)df = E\{|x|^2\}$
- Time-frequency shift covariance :

$$\tilde{x}(\tau) := S_{t_0, f_0}x(\tau) = x(\tau - t_0)e^{2\pi i f_0\tau} \Rightarrow W_{\tilde{x}}(t, f) = W_x(t - t_0, f - f_0)$$

- Symplectic covariance : $W_x\left(A\begin{pmatrix} t \\ f \end{pmatrix}\right) = W_{\tilde{x}}(t, f) \Leftarrow \mu(A)x(\tau) =: \tilde{x}(\tau)$

A: symplectic matrix ; μ: metaplectic representation

Application of the Wigner transformation on the anechoic problem:

$$x_i(\vec{t}) = \sum_{j=1}^{n} \alpha_{ij} s_j(\vec{t} - \vec{\tau}_{ij})$$

Wigner-Ville-spectrum ⬇ Independence of the s_j

$$W_{x_i}(\vec{t}, \vec{f}) = \sum_{j=1}^{n} |\alpha_{ij}|^2 W_{s_j}(\vec{t} - \vec{\tau}_{ij}, \vec{f})$$

Symplectic covariance ⬇ Special case 2D

1D-Problem (suitable choice of parameters) !!:

$$\left|F^{\gamma} x_i(f)\right|^2 = \sum_{j=1}^{n} |\alpha_{ij}|^2 |F^{\gamma} s_j(f - \cos(\gamma)\tau_{ij})|$$

Solution of the resulting univariate problem

$$\left|F^{\gamma} x_i(f)\right|^2 = \sum_{j=1}^{n} |\alpha_{ij}|^2 |F^{\gamma} s_j(f - \cos(\gamma)\tau_{ij})|$$

Of the form: X= A S with X≥0 and S≥0

➡ Non-negative matrix factorisation problem (NMF)[1,2]

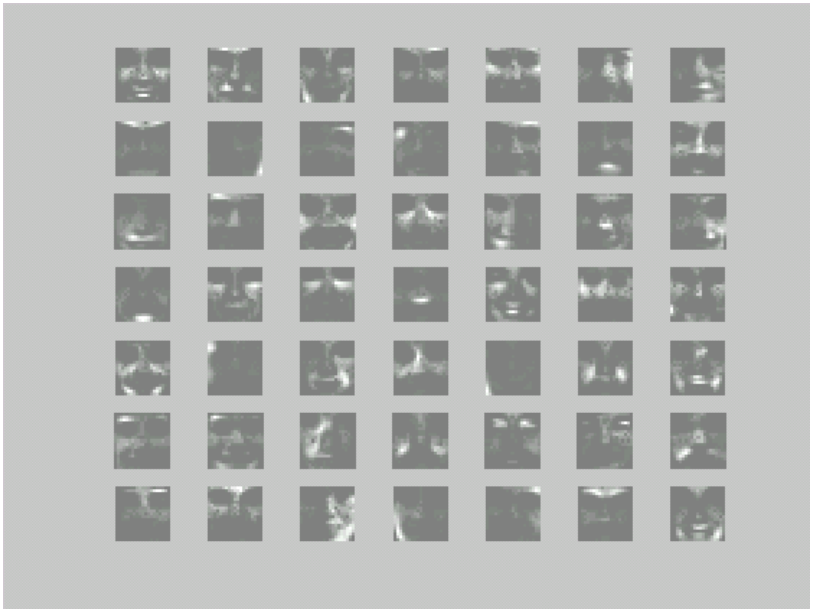
- Iterative algorithm
- Parts based representation (not holisitc like PCA)

[1]D.D. Lee, H.S. Seung (1999) Learning the parts of objects by Non-Negative Matrix Factorization. Nature 401, 788-91.

[2]A. Cichocki, et al (2006) New Algorithms for Non-Negative Matrix Factorization in Applications to Blind Source Separation. ICASSP-06, 621-625.

Example for NMF





Matrix factorization: $V \approx WH$

V: $n \times m$ matrix. Each column of which contains n nonnegative pixel values of one of the m facial images.

W: $(n \times r)$: r columns of W are called basis images.

H: $(r \times m)$: each column of H is called encoding.

Example for NMF

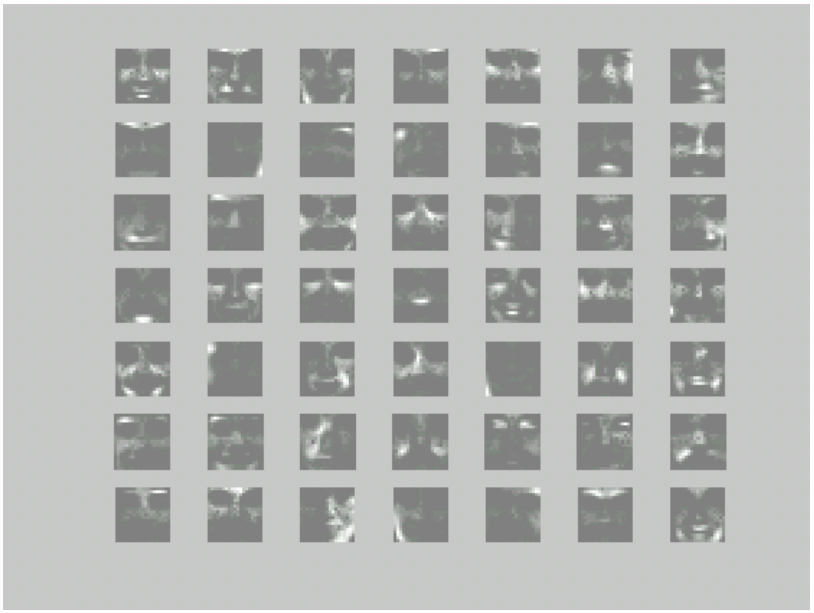


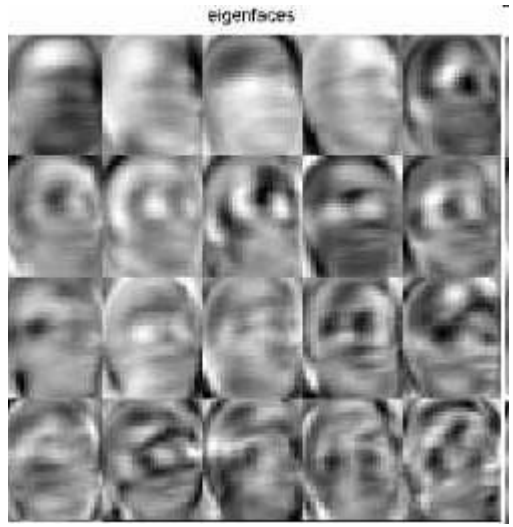
Matrix factorization: V≈WH

V: n×m matrix. Each column of which contains n nonnegative pixel values of one of the m facial images.

W: (n ×r): r columns of W are called basis images.

H: (r ×m): each column of H is called encoding.

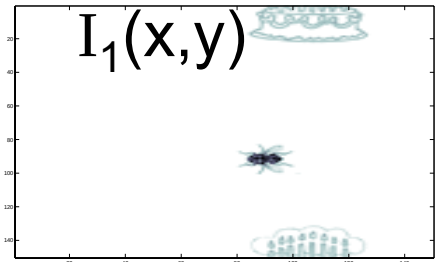


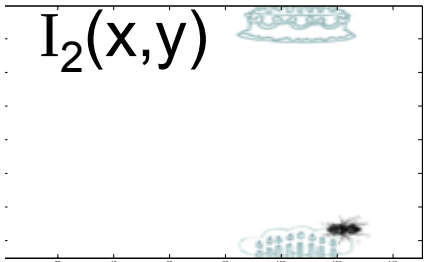


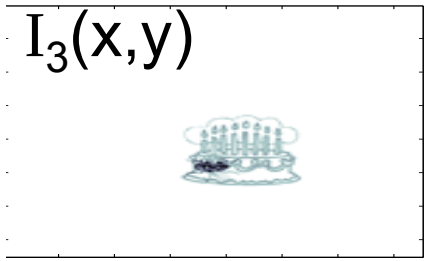
Comparison: PCA Eigenfaces

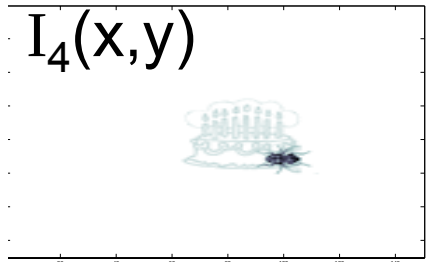
Application to introductory example

Input

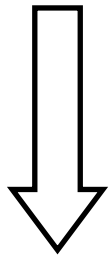








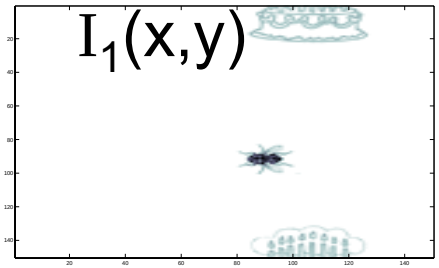
Anechoic demixing

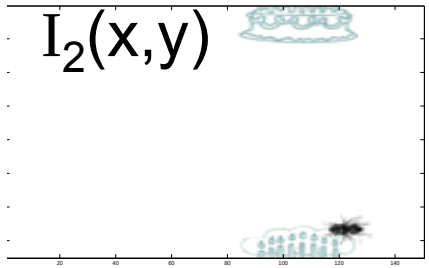


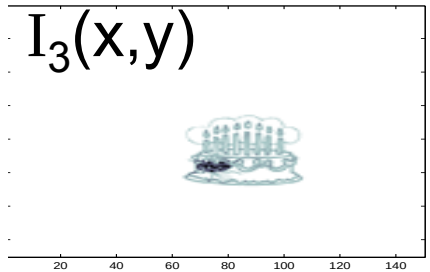
$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x-\tau_{i1}, y-\tau_{i1})$$

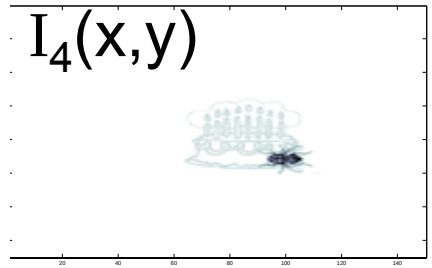
Application to introductory example

Input

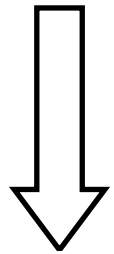




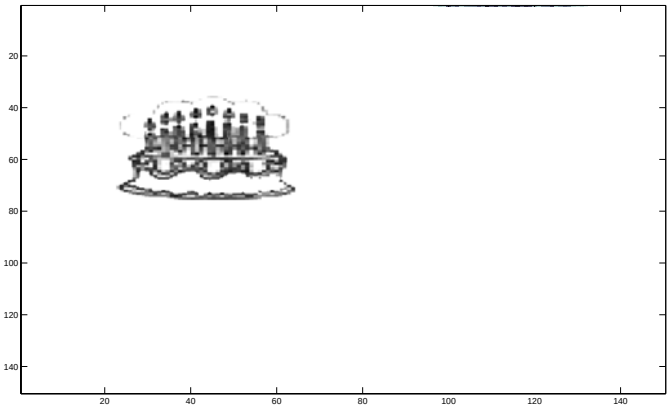


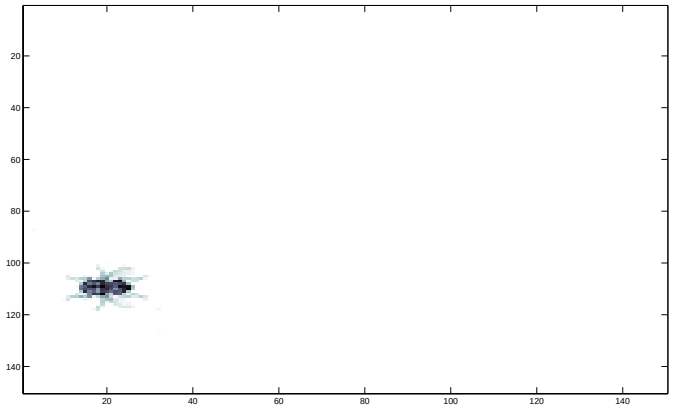


Anechoic demixing

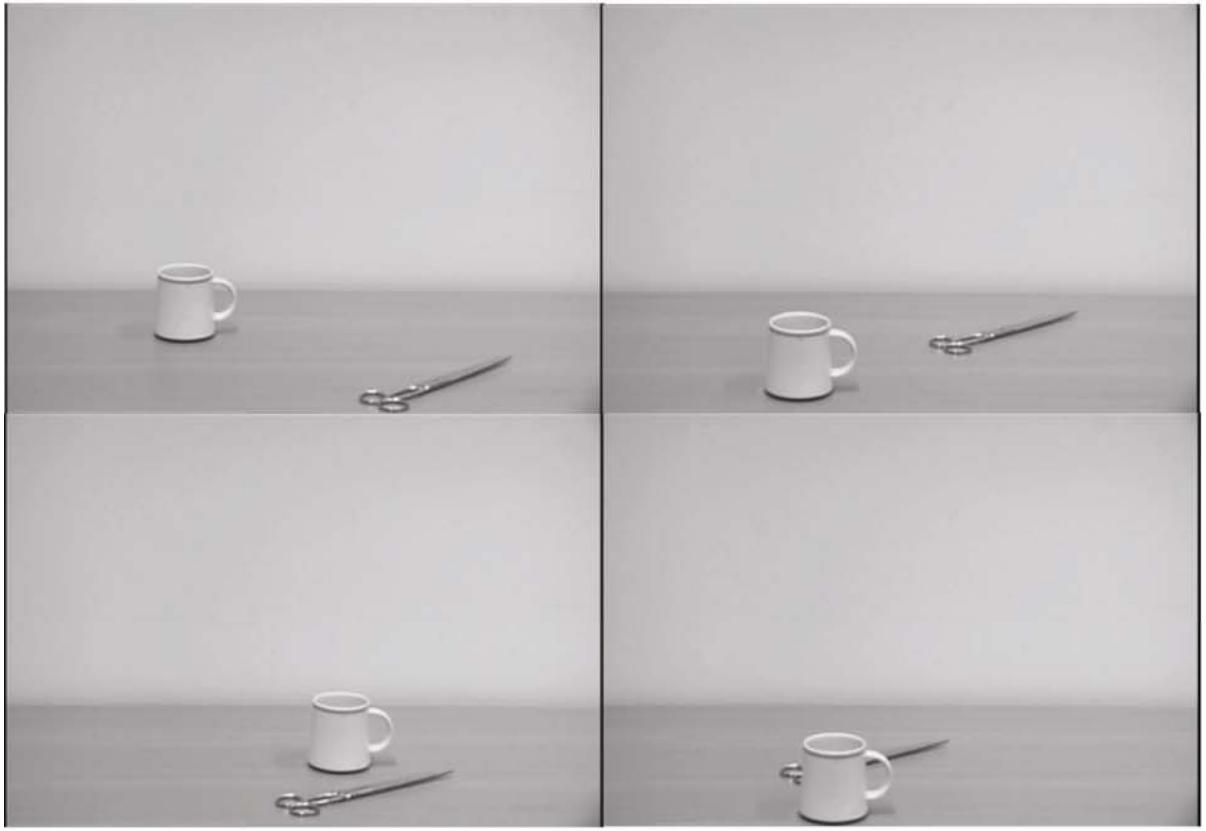


$$I_i(x,y) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(x-\tau_{i1}, y-\tau_{i1})$$



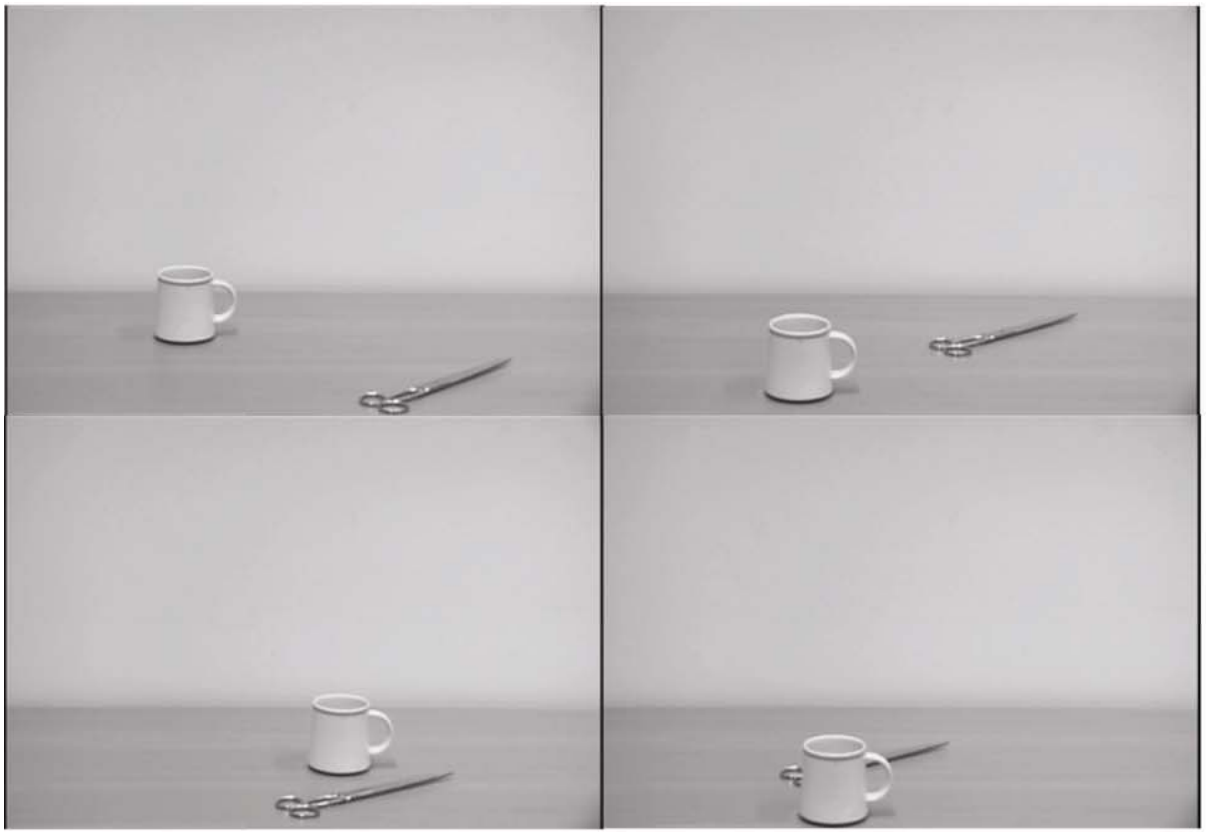


Real application: Natural images

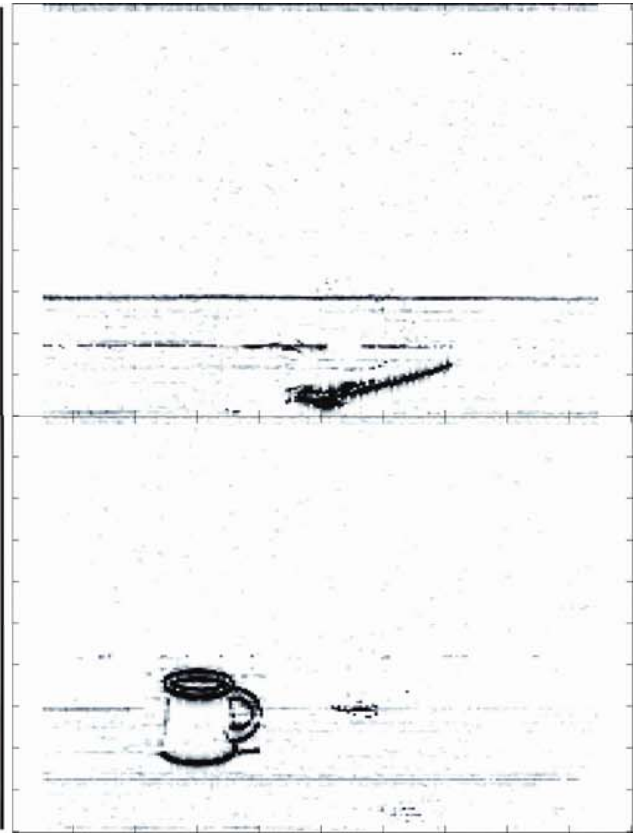


Original grayscale images (200 x 200)

Real application: Natural images



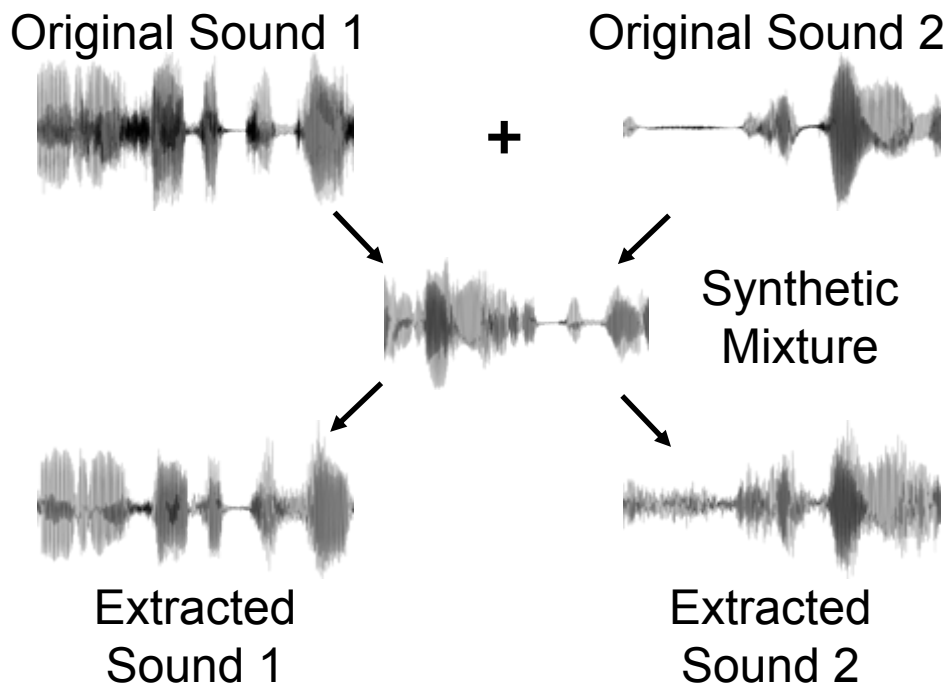
Original grayscale images (200 x 200)



Extracted features
(after whitening)

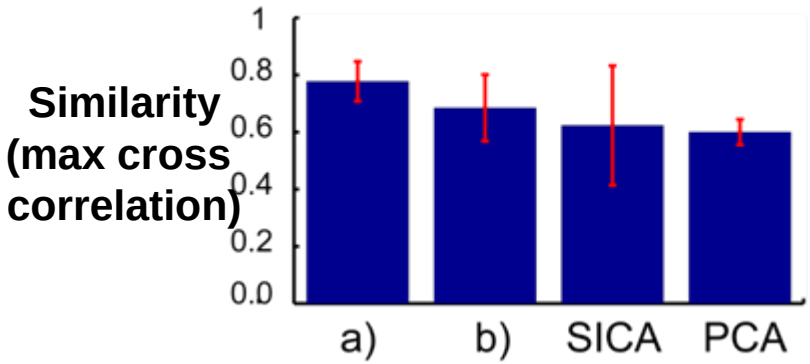
Application: “Cocktail party problem”





- Test with several kinds of synthetic anechoic sound mixtures and two different implementations (a) and b))
- Average match of 80% between extracted sources and original sounds
- Better than existing methods (SICA/PCA)

Average result for sound demixing



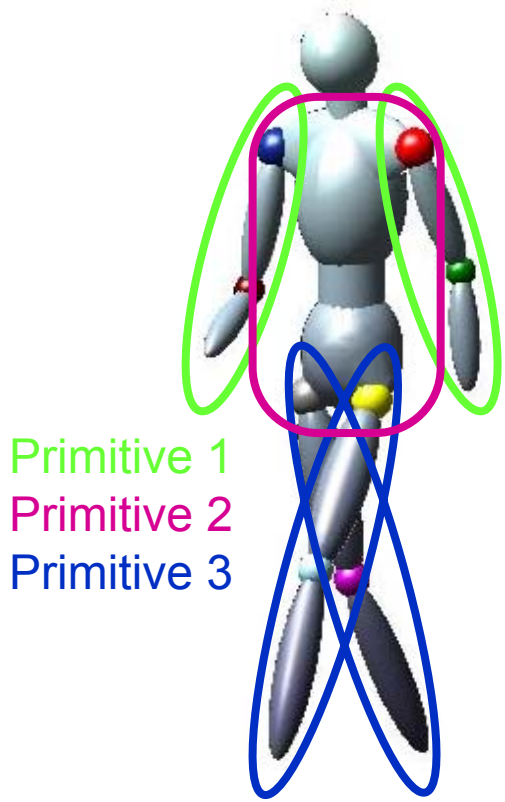
Part II: Analysis of human motion and application to COBOL

Part II: Analysis of human motion and application to COBOL

Question: Can this unsupervised learning technique also be used to learn motion components?

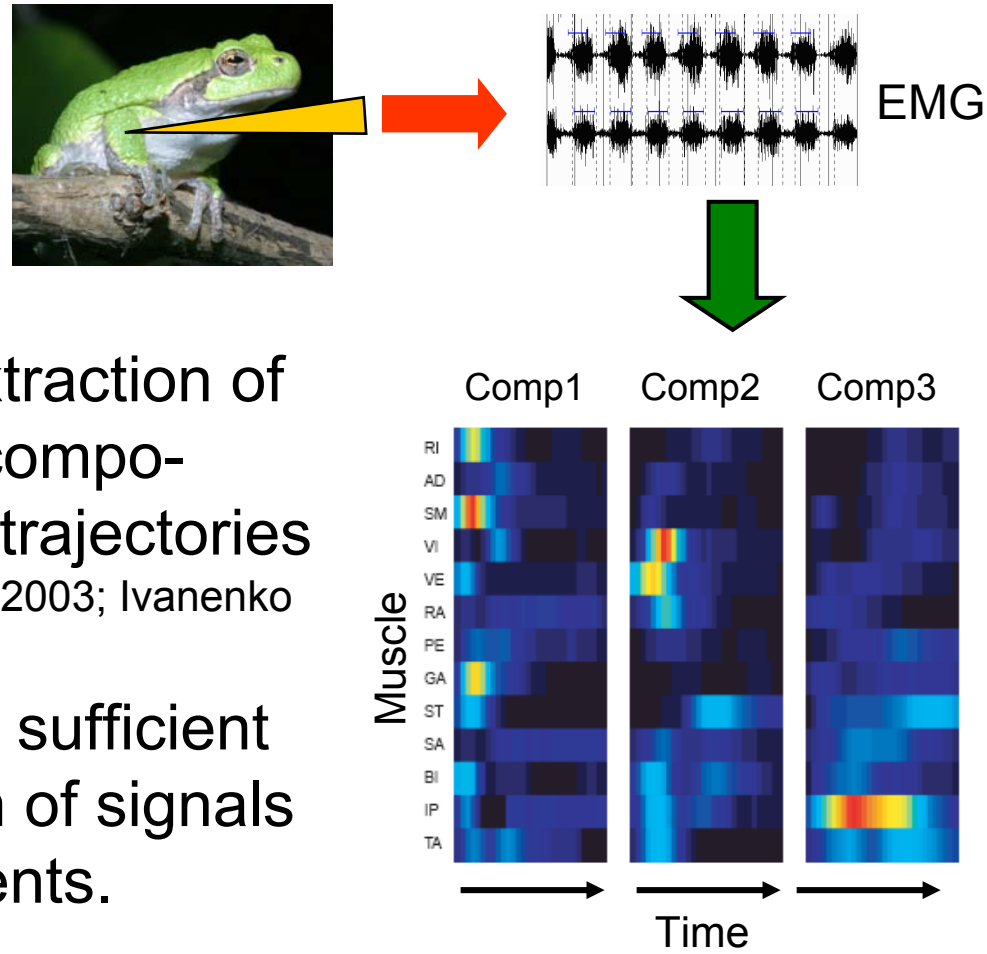
Analysis of human movements

- Degree of freedom problem in the planning of complex motor behaviour
- Motor control: Spatial primitives **(synergies)** as modules for the control of complex movements
  (e.g. Bernstein, 1967; Flash & Hochner, 2005).
- Synergies often encompass only limited number of degrees of freedom (DoFs).



Extraction of spatial primitives

- Synergies corresponding to jointly activated muscle groups.
- Supporting evidence by extraction of common spatio-temporal components from EMG data and trajectories (e.g. Santello et al. 2002; d'Avella et al. 2003; Ivanenko et al. 2005).
- Often few components are sufficient for accurate reconstruction of signals for many different movements.



(d'Avella, Saltiel & Bizzi, 2001)

Extraction of spatial primitives

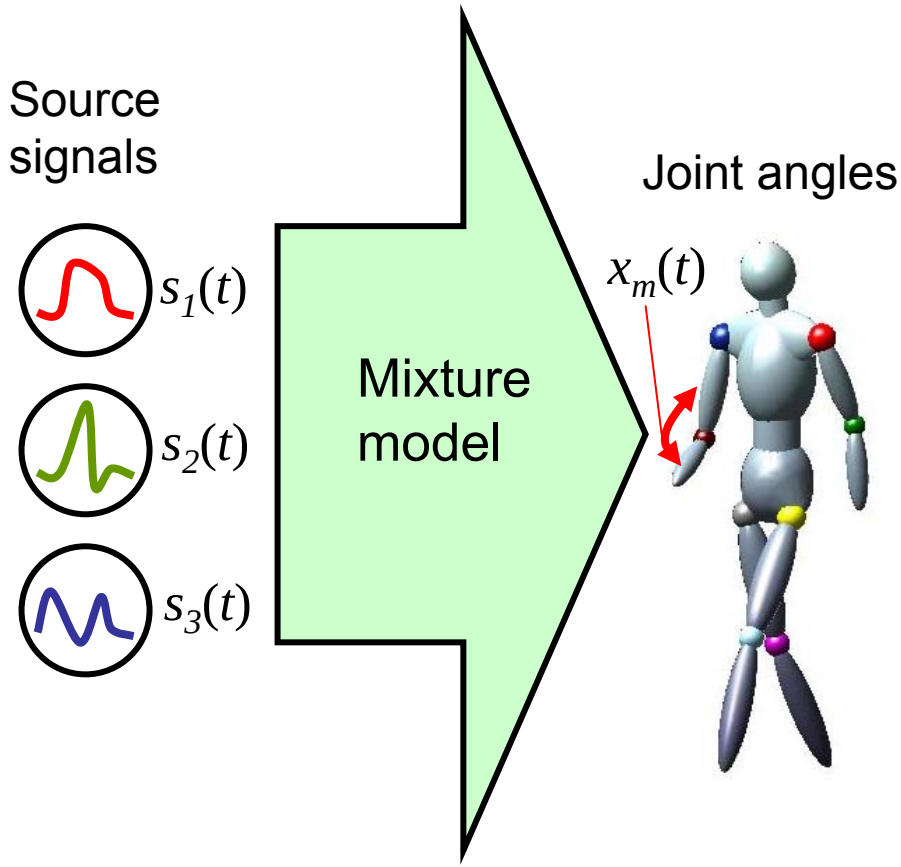
Classical approach

- Approximation of joint angle trajectories by mixture of source signals.
- Simple linear mixture model (PCA, ICA):

$$x_m(t) = \sum_n w_{mn} s_n(t)$$

Joint angle (known)

Mixing weights



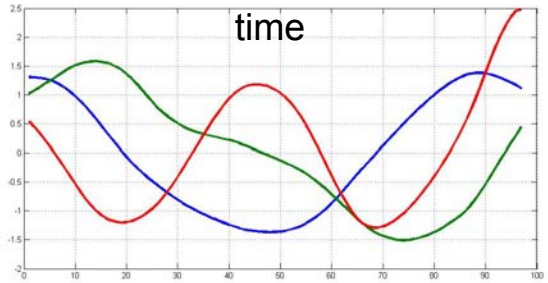
Extraction of spatial primitives

Joint by Joint analysis

Idea:

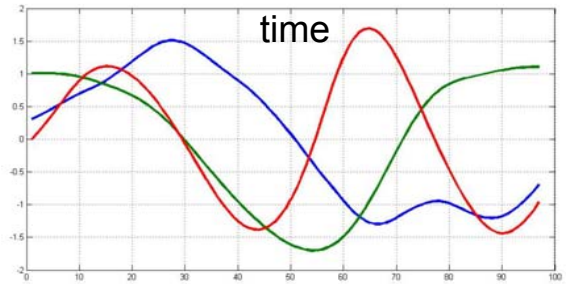
Application of ICA to single joints separately should give better results since those trajectories are more similar.

Exemplary source signals



Left knee

amplitude



Right elbow

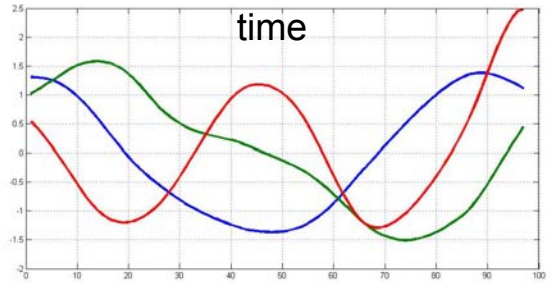
Extraction of spatial primitives

Joint by Joint analysis

Idea:

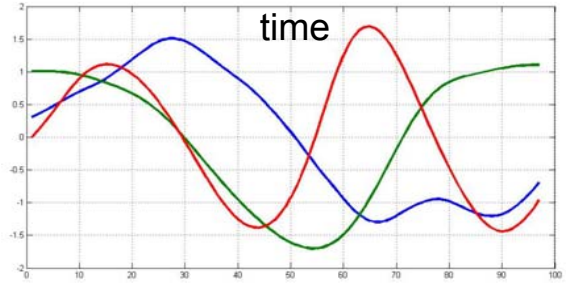
Application of ICA to single joints separately should give better results since those trajectories are more similar.

Exemplary source signals



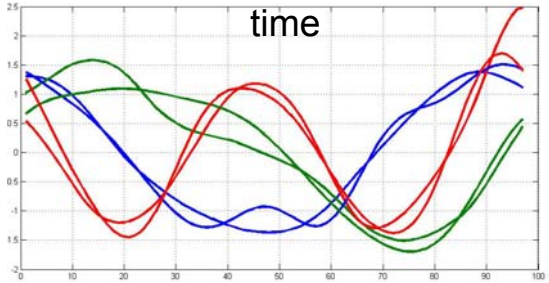
Left knee

amplitude



Right elbow

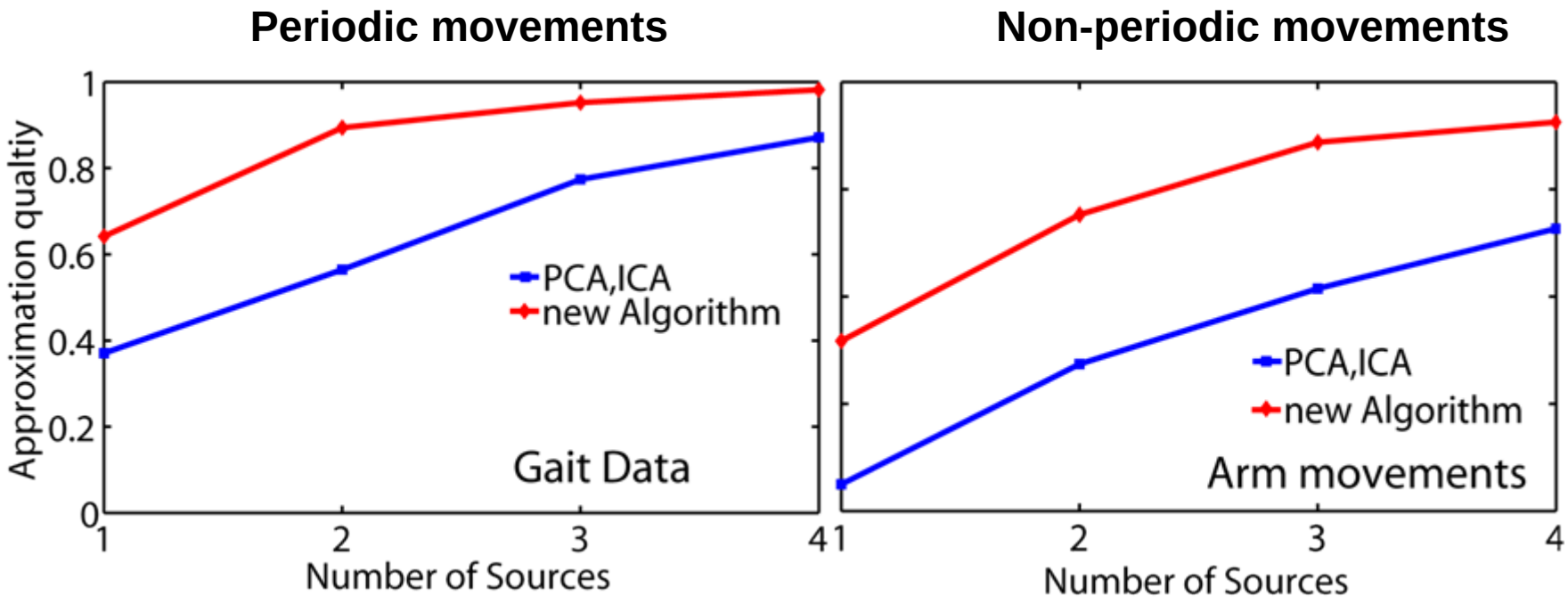
(realigned)



Better Model includes shifts: $x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\ s_j(t - \tau_{ij})$

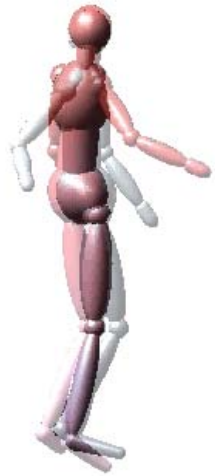
The anechoic model should be much better suited to describe the motion data!

Comparison between different models



- Anechoic demixing outperforms standard methods (PCA, ICA, …).
  - -Very accurate approximation of all joint movements with only 3-4 sources.
- Similar result for non-periodic movements.





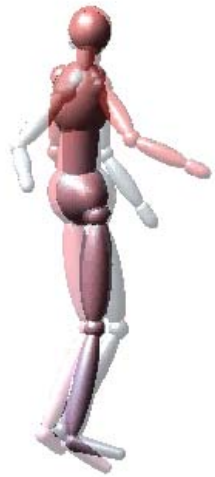
New algorithm
(3 sources, 606 parameters)

Ground truth

PCA
(4 sources, 604 parameters)

Anechoic mixtures allow a highly accurate and compact representation of complex human motion!





New algorithm
(3 sources, 606 parameters)

Ground truth

PCA
(4 sources, 604 parameters)

Anechoic mixtures allow a highly accurate and compact representation of complex human motion!

Question: Can this representation be used to identify emotion specific features ?

Motion capture database

Emotion expression in gait

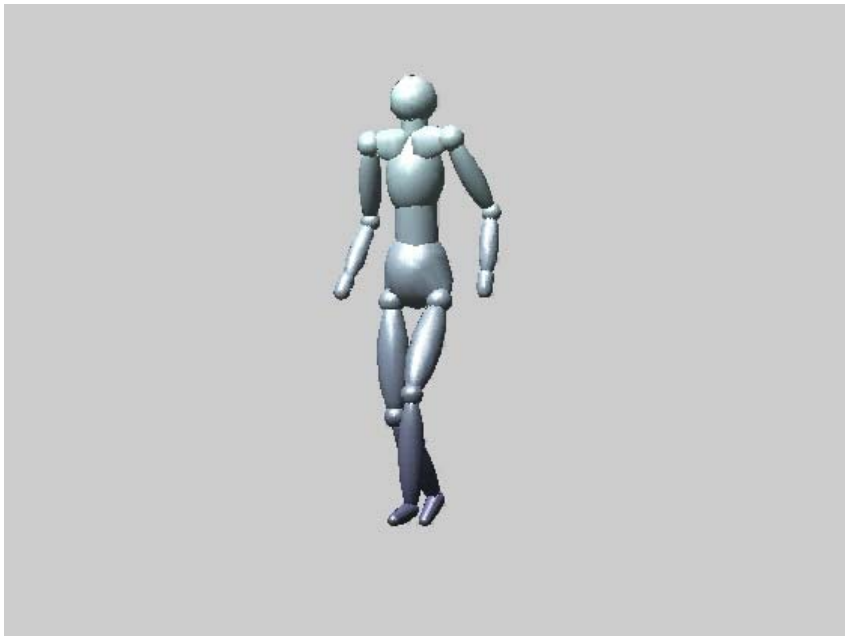
Motion-capture database

- Five basic emotions

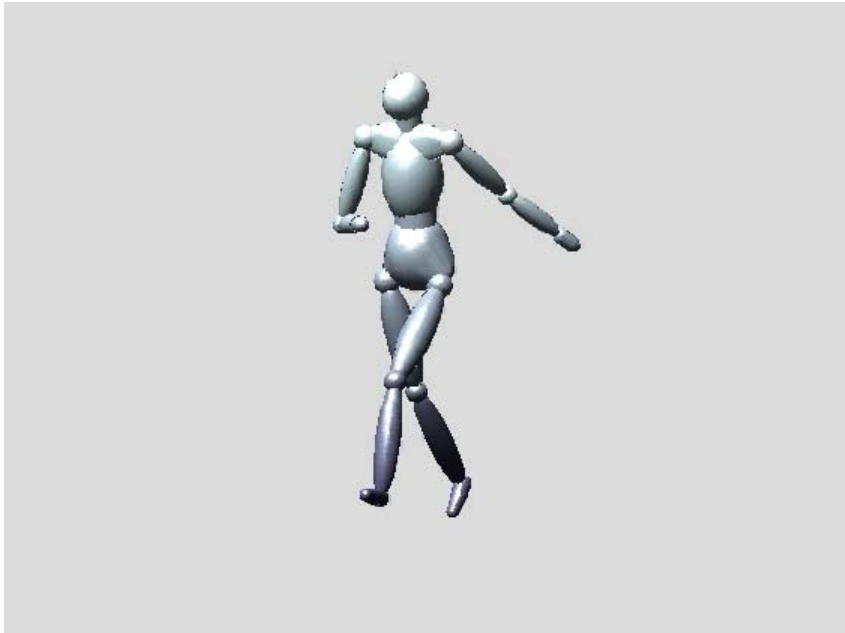
  (Anger, happiness, sadness, slow fear)
- Mood induction
- 25 different lay actors
- 13 right and 12 left handed people



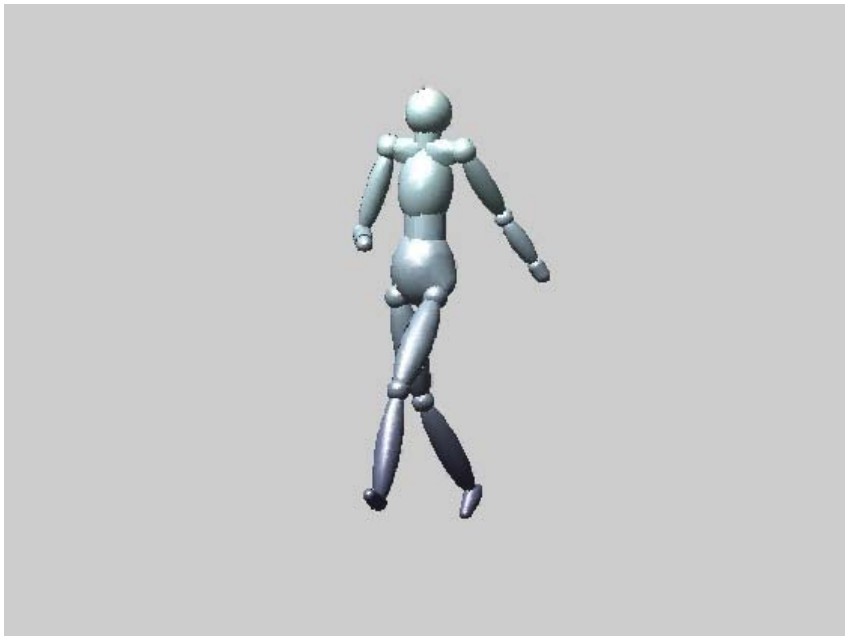
Example:Fear



Example: Anger

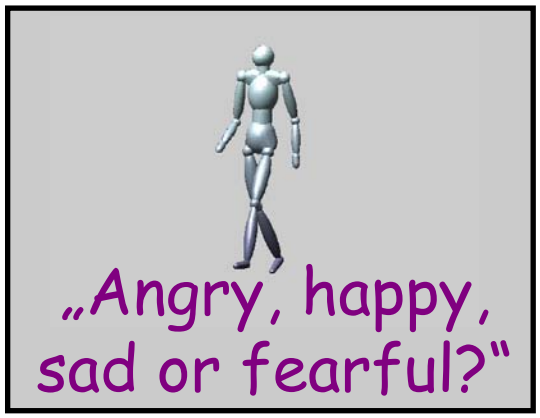


Example: Happiness



Evaluation of our Database
Classification of emotional gait

Two experimental tasks





Stimulus Response

Affect	Angry	Happy	Fear	Sad
Angry	70.3	15.6	3.2	1.0
Happy	23.2	75.1	1.9	1.2
Fear	4.7	6.6	77.1	8.0
Sad	1.8	2.7	17.9	89.8

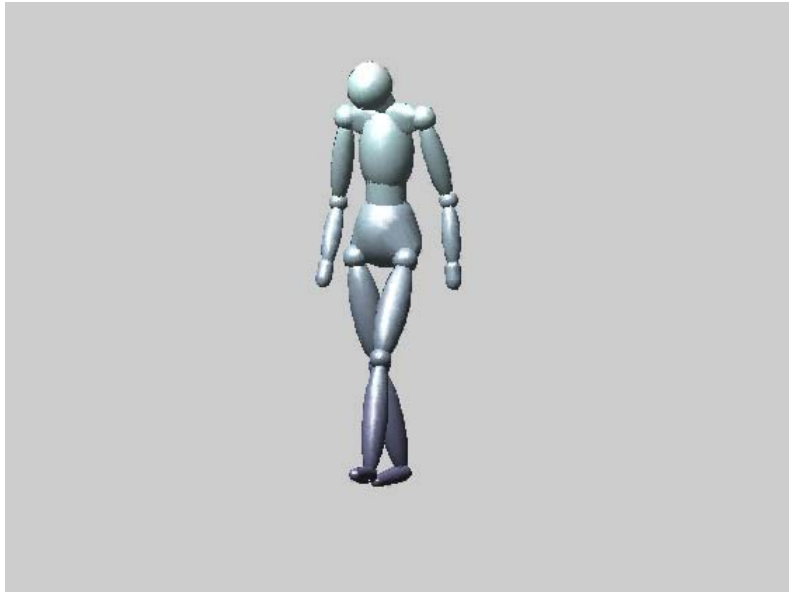
- Expressions categorised with high accuracy
- Confusions: velocity seems an important cue

Question: Expressive features?

Previous Research:

- Subjective ratings of movement features
- Physical measures
- Identified features:

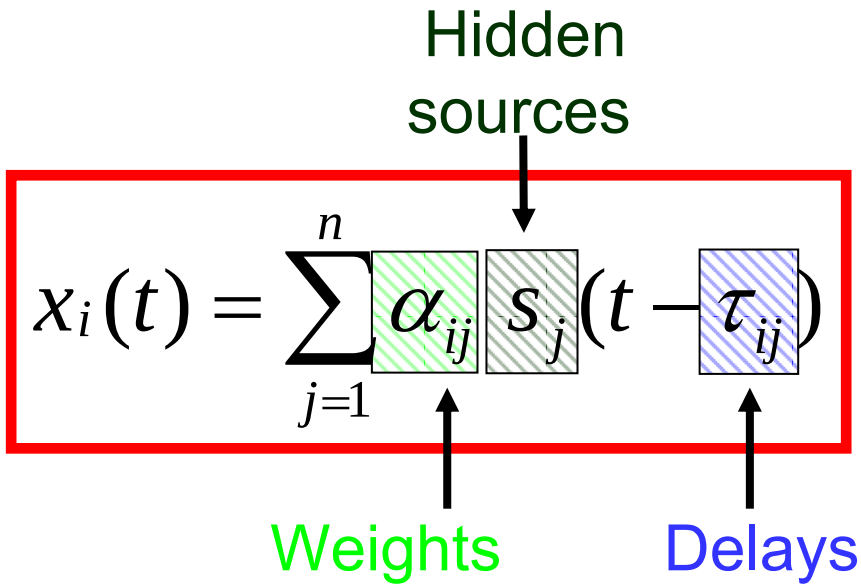
Amount of arm swing
Head inclination
Velocity …



(e.g. Montepare et al. 1987, Wallbott 1998, Atkinson et al. 2004, 2007, Pollick 2001)

Question: Expressive features?

Generative model:



Possible emotion dependent changes in sources, delays or weights

Question: Expressive features?

Generative model:

Average cross correlation between sources extracted from separate emotions is over 97%

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\, s_j(t - \tau_{ij})$$

➡ Sources almost independent of emotion!

Question: Expressive features?

Generative model:

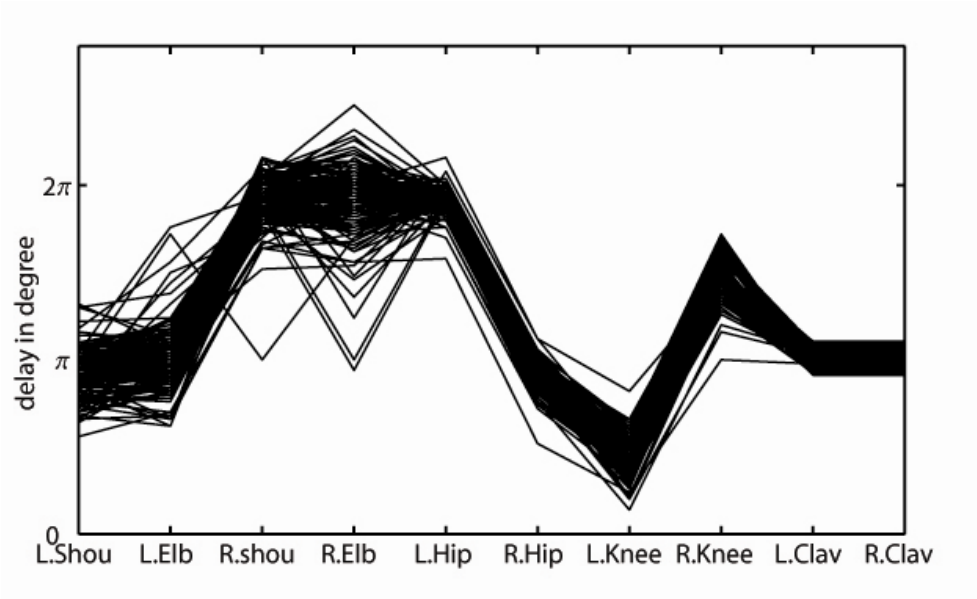
Average cross correlation between sources extracted from separate emotions is over 97%

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij}\, s_j(t - \tau_{ij})$$

➡ Sources almost independent of emotion!

Analysis of Delays:

- Main effect: Joint not emotion
- Reflecting gait behaviour
- Very small significant variation with emotion (<1%)



➡ Delays almost independent of emotion!

Question: Expressive features?

Change in Weight?

- Decomposition of mixing weights:

$$\alpha_{mn} \approx \alpha_{mn,\text{neutral}} + \Delta\alpha_{mn,\text{ Emotion } k}$$

Weight for neutral movement

Emotion-specific part

- Estimation of significant 'expressive' emotion-specific components by **sparse regression;** minimize:

$$V(\Delta\alpha_{\text{Emotion } k}) = \sum_l \left\| \alpha^l_{\text{Emotion } k} - \alpha^l_{\text{Neutral}} - \Delta\alpha_{\text{Emotion } k} \right\|^2_F$$

$$+ C \underbrace{\sum_{k,m,n} \left| \Delta\alpha_{mn,\text{ Emotion } k} \right|}_{L_1 \text{ norm of } \Delta\alpha_{\text{Emotion,k}}}$$

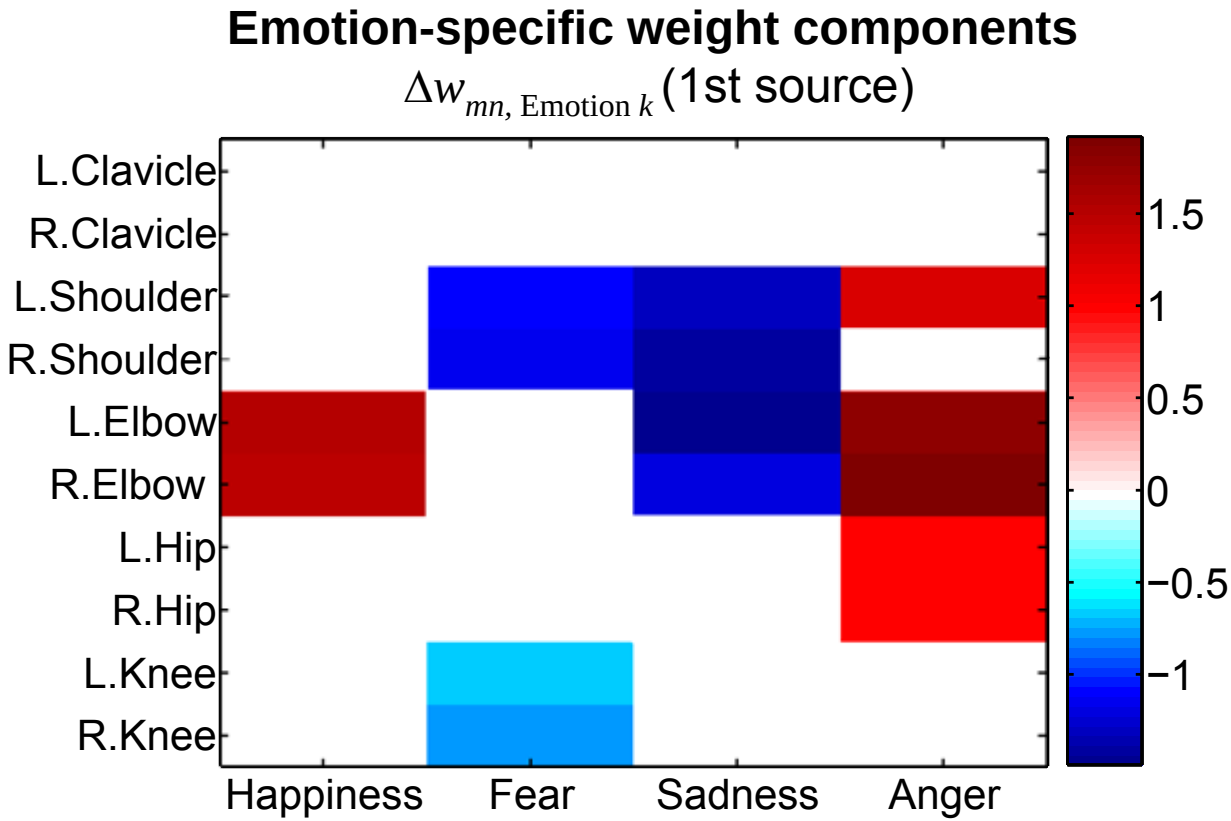
L_1 regularization $\Rightarrow$ many zero elements.

l: actor / trial number

Question: Expressive features?

Change in Weight?

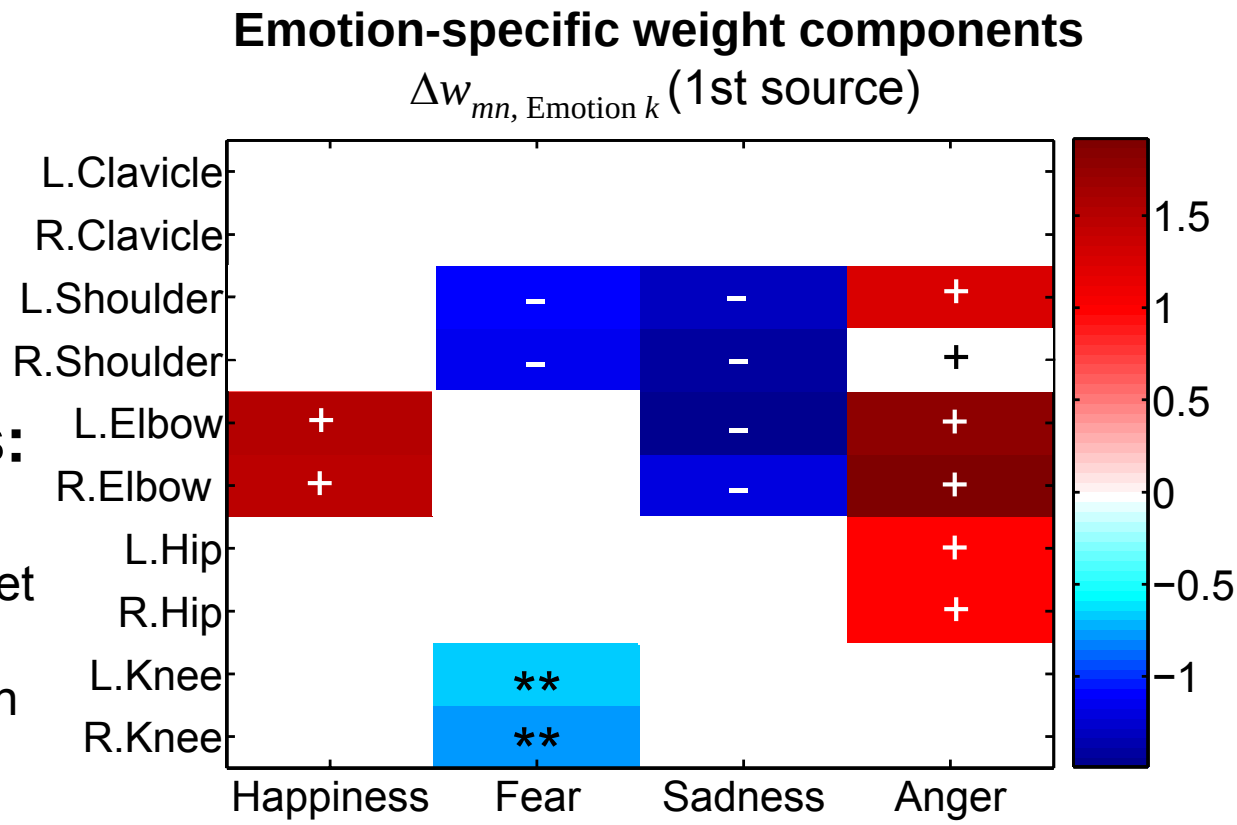
- Distinctive set of emotion-specific spatial components.
- Increase of amplitude for anger and happiness.
- Decrease for fear and sadness.



Question: Expressive features?

Change in Weight?

- Close match with emotion-specific components for gait perception.



Perception features:

+/-: Literature results (cf. Meijer, 1989; Goldstein et al. 1987; Walbott, 1998)

**: Feature found in our own psychophysical experiments

Question: Expressive features?

Change in Weight?

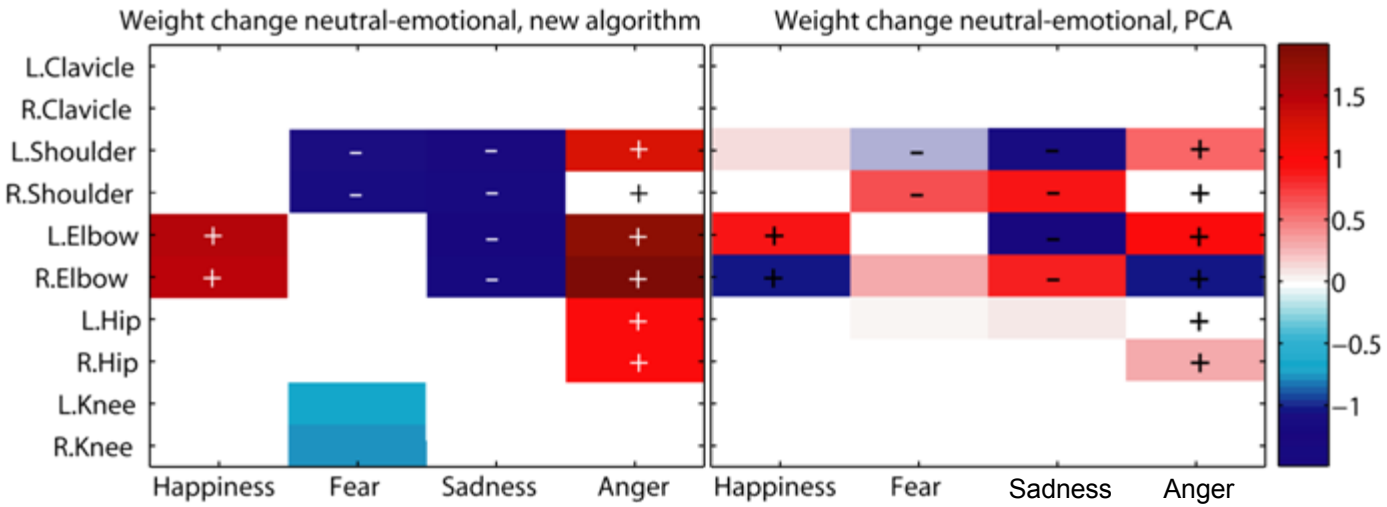
Comparison with other methods:

- Same statistical analysis for sources extracted with PCA, Fourier,…
- No obvious relationship with psychophysical results for other methods.

➔ Compact model results in more interpretable features.

Emotion-specific weight components

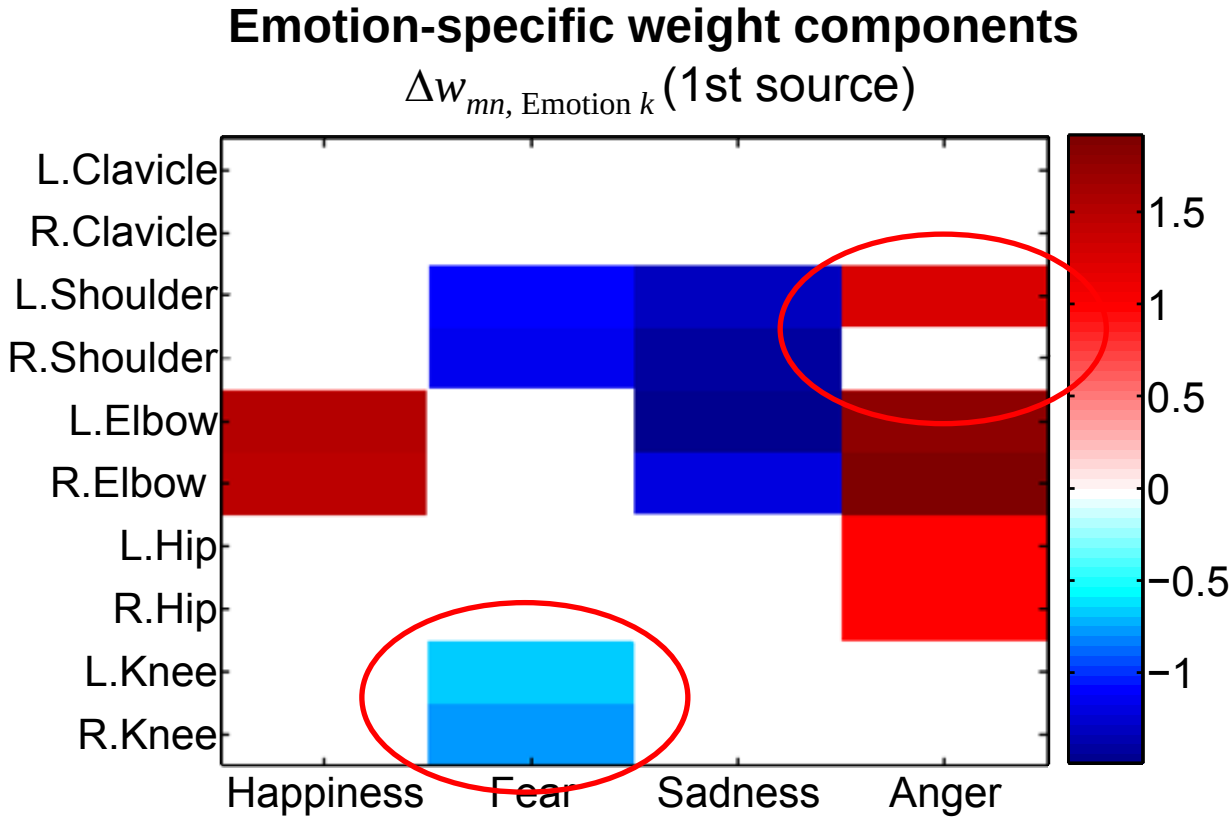
$\Delta w_{mn,\,\text{Emotion } k}$ (1st source)



New method PCA

Question: Expressive features?

Interesting observation!

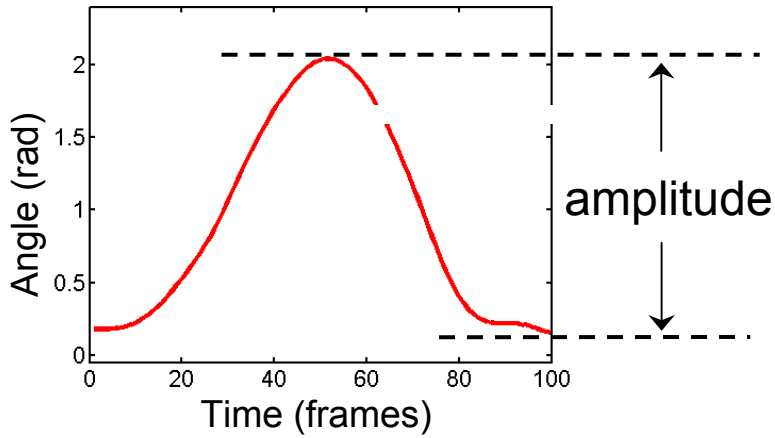


Unequal weights between left and right body half!

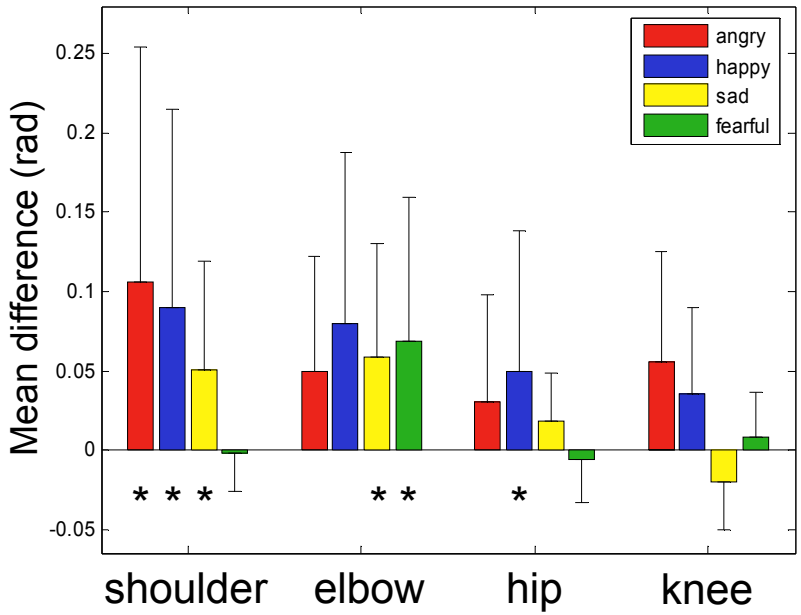
➡ Emotion specific asymmetry

Quantification of asymmetry

Simplest physical measure: amplitude

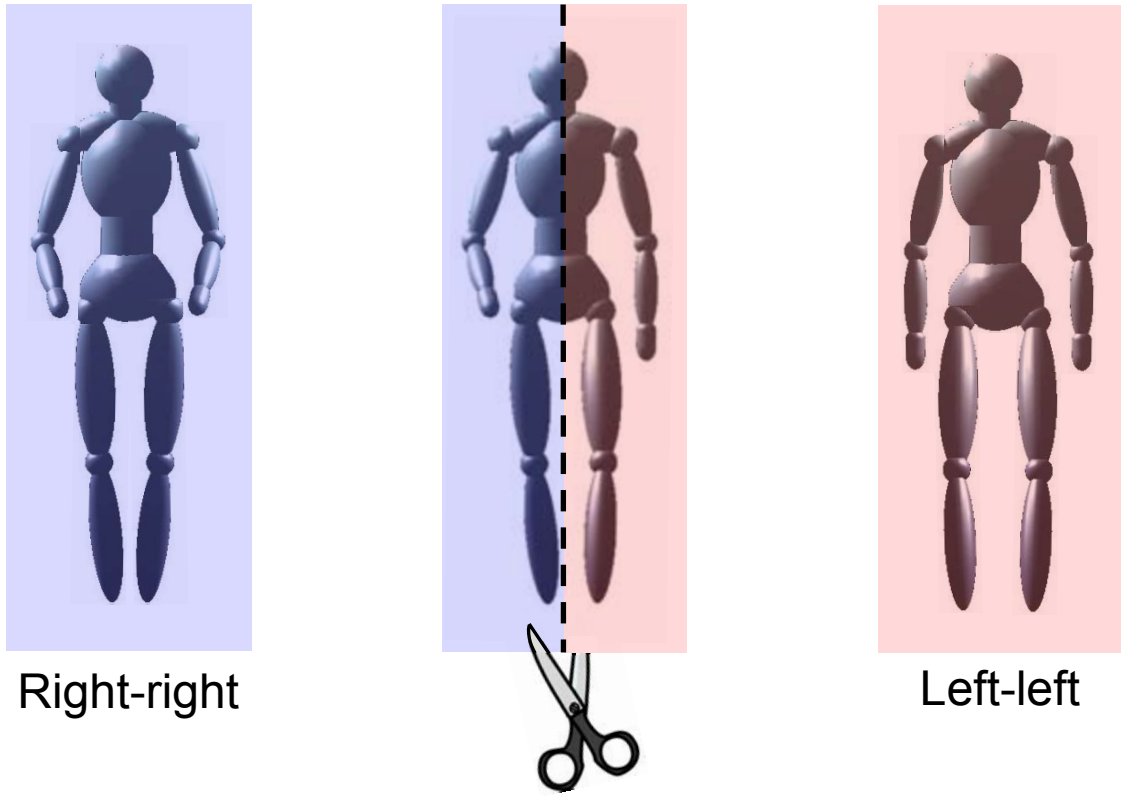


Left-right amplitude difference

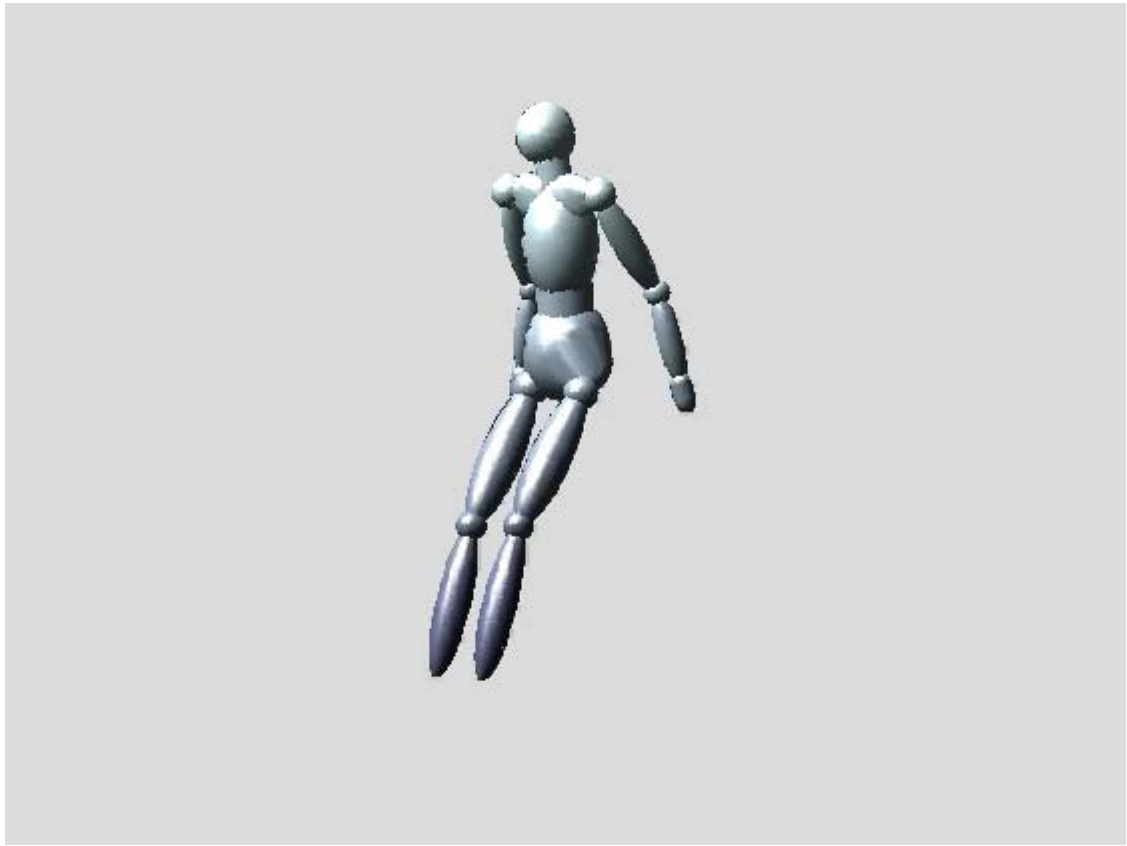


Higher movement amplitudes for left side of body

Asymmetry of emotional body movement: perception

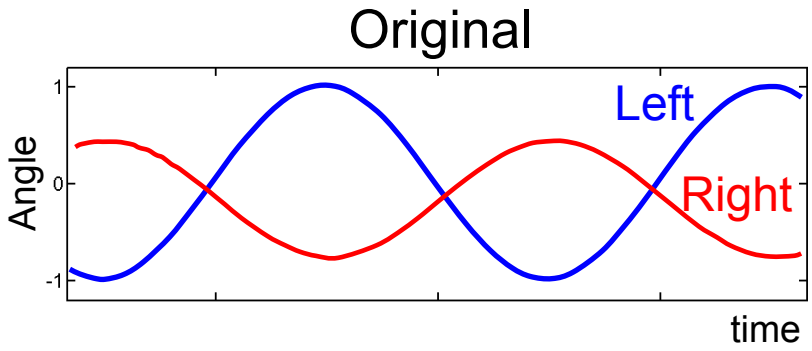


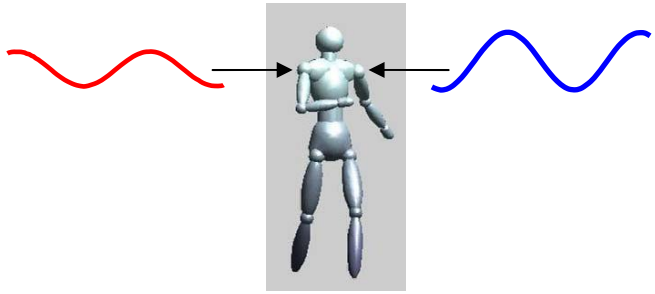
Creation of a ‘chimeric walker’

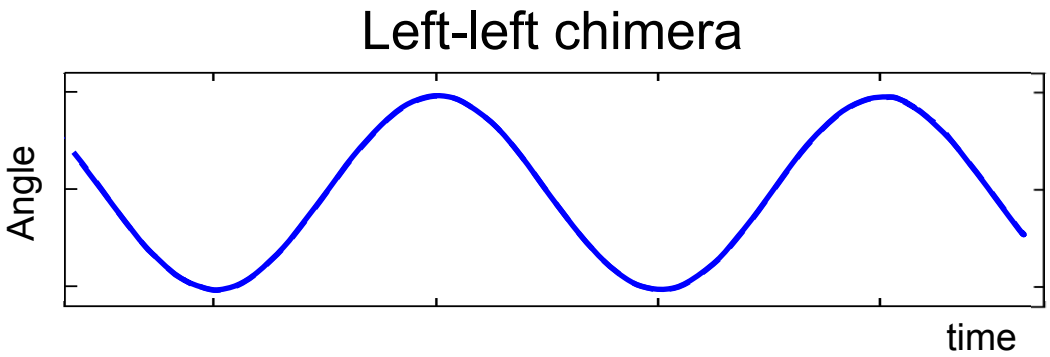


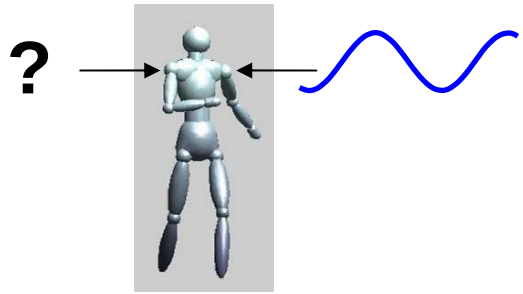
Naive method results in weird locomotion pattern

Creation of a chimeric walker

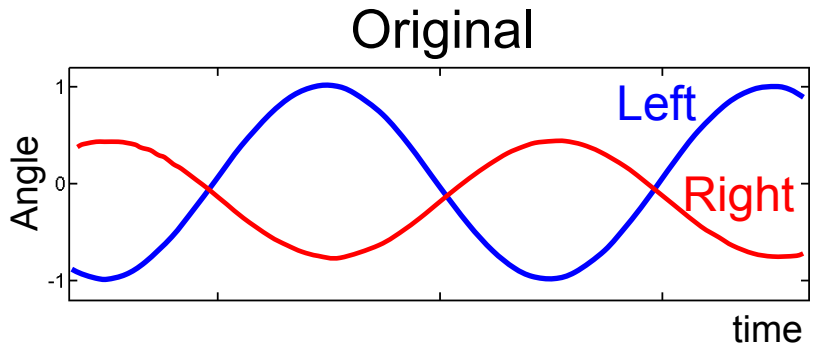


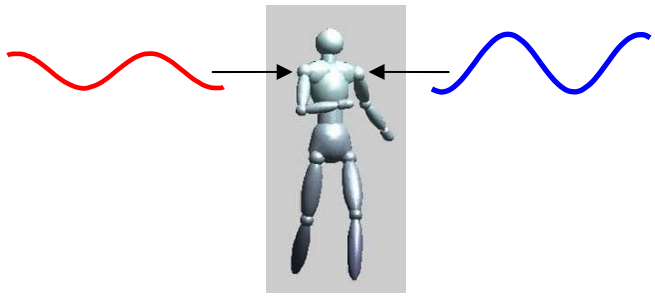


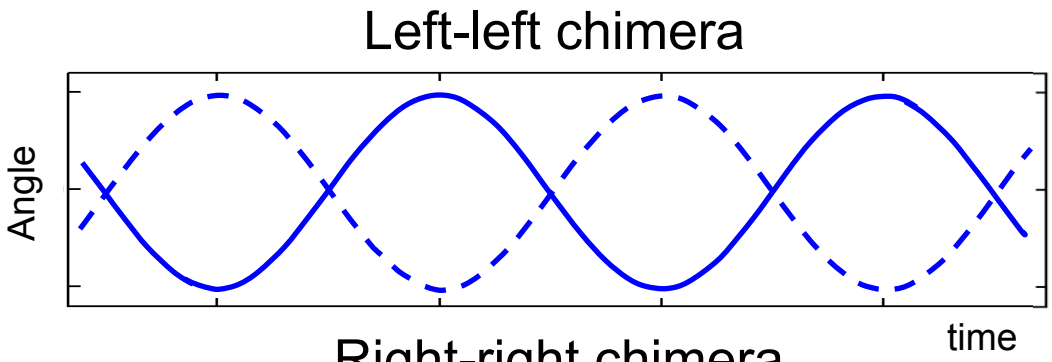


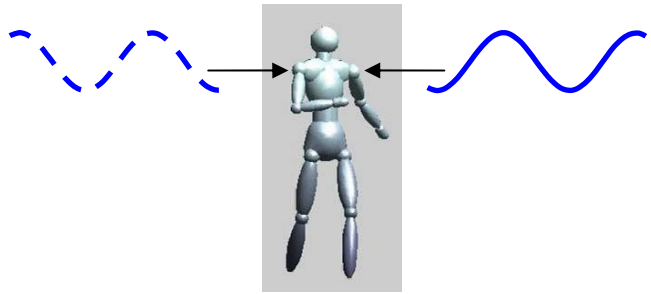


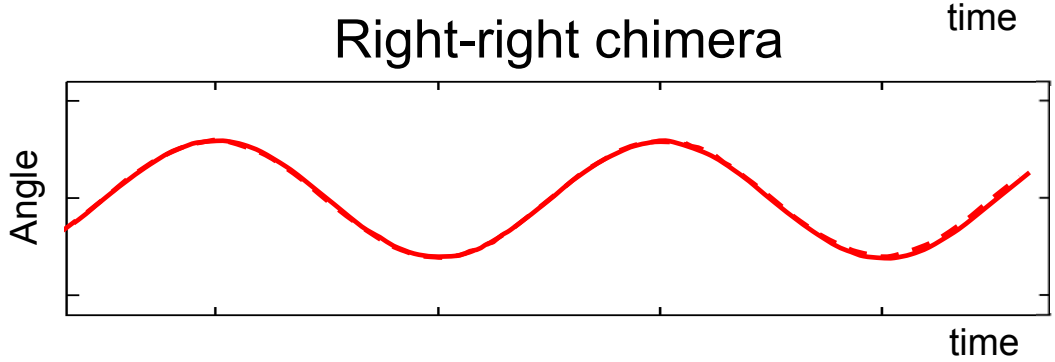
Creation of a chimeric walker

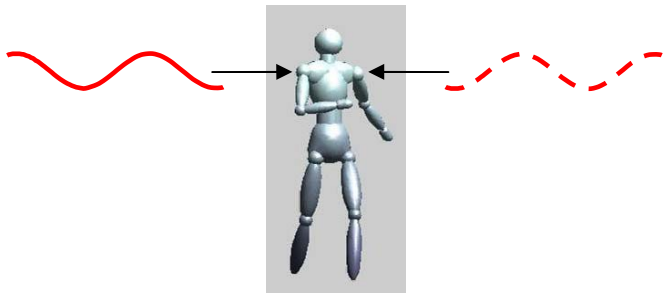












Chimeric walkers: examples

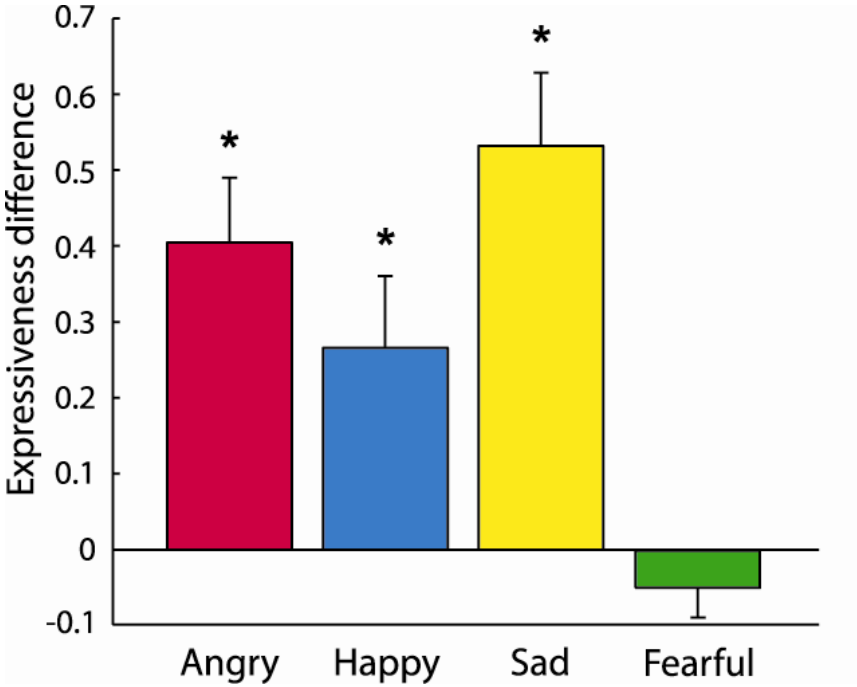
Left-left chimera



Right-right chimera



Asymmetrie in perception



- Left-left chimeras more expressive than right-right chimeras
- Significant for three emotions

Part III: Synthesis of motion and application in computer graphics

Requirements for computer animation

1. High approximation quality

2. Sensitivity to subtle style changes

3. Modifiability and interactive behavior

Requirements for computer animation

1. High approximation quality
   3 Sources explain 99% of the variance ✓

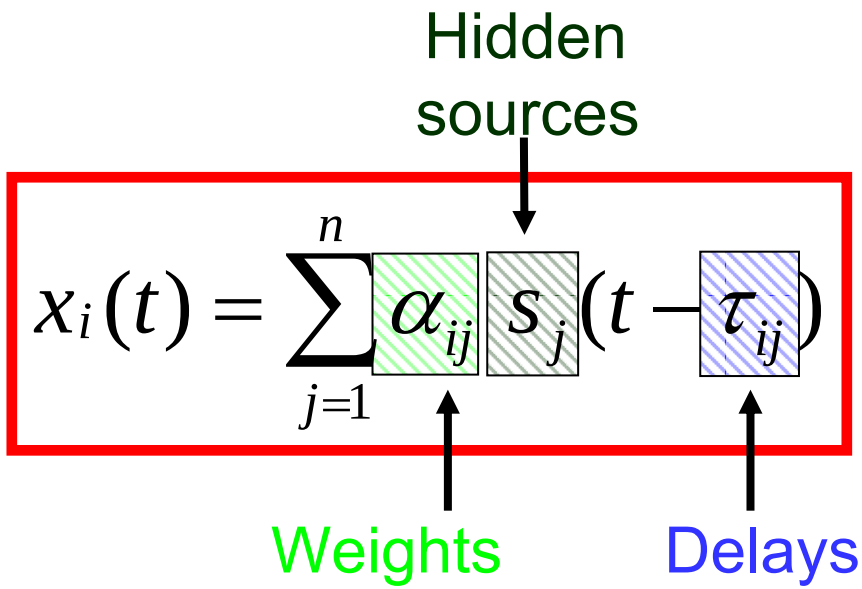
2. Sensitivity to subtle style changes
   perfect match with psychophysics for emotions ✓

3. Modifiability and interactive behavior

- Whole trajectory needs to be known in advance ⇨ cannot be synthesized in real-time / online
- Problem not interactive

The model again ;-)

Alterable parameters:

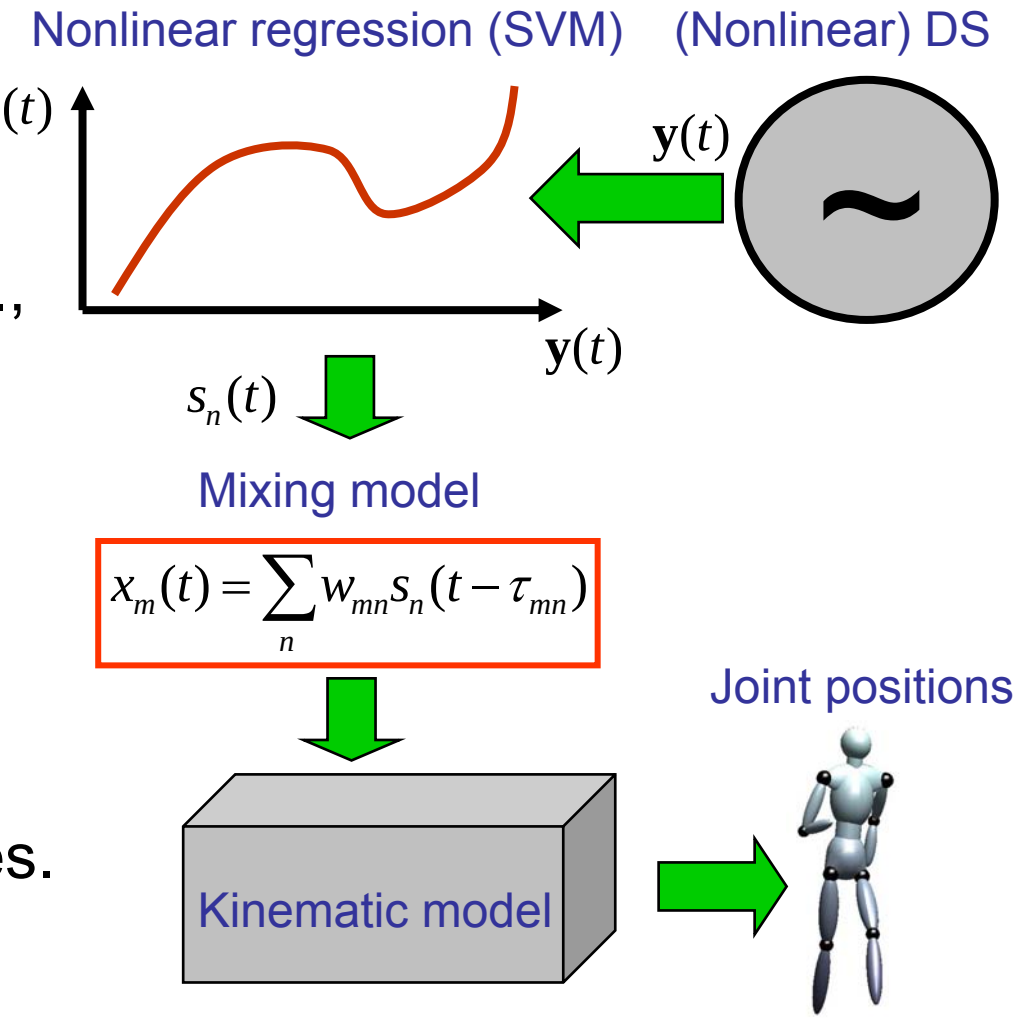


Change of sources: interactive behavior

- Mapping between source signals and solution of NL dynamical system.
- NL dynamical systems (e.g., van der Pol oscillator)

$$\ddot{y}(t) + \dot{y}(t)(y^2(t) - k) + \omega_0^2 y(t) = 0$$

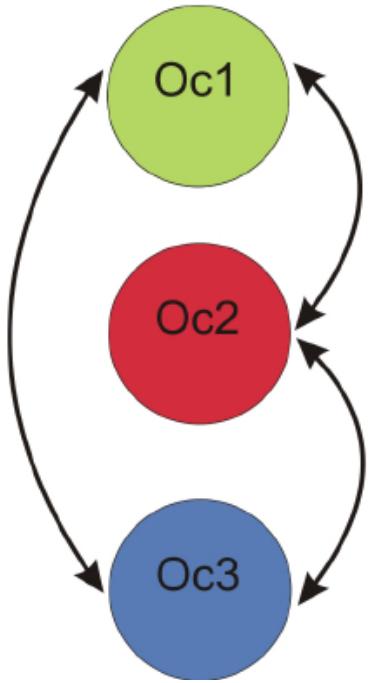
- State vectors $\mathbf{y}(t)$ mapped onto source signals $s_n(t)$.
- Learned generative model used to compute joint angles.



Dynamic couplings

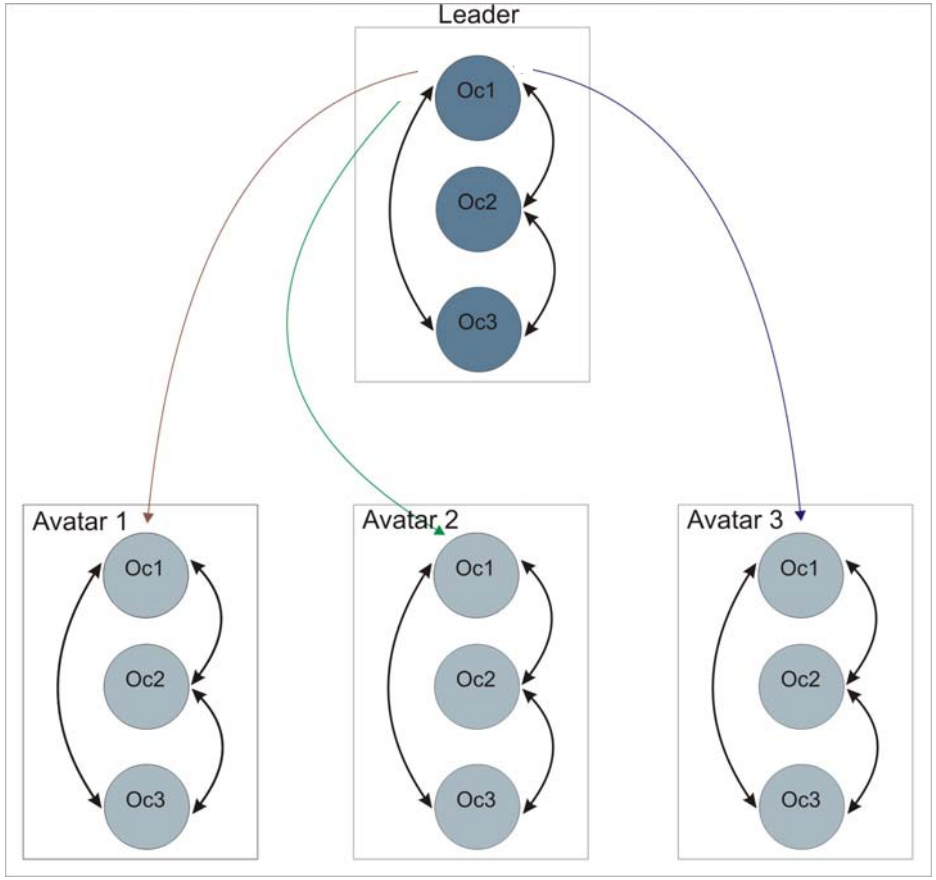
- Dynamic **coupling** between oscillators to stabilizes coordination.
- **Contraction theory** (Slotine & Lohmiller, 2000) $\Rightarrow$ Stable system dynamics for velocity coupling:

$$\ddot{y}_1 + \lambda\left(y_1^2 - k\right)\dot{y}_1 + \omega_0^2\, y_1 = \alpha\left(\dot{y}_2 - \dot{y}_1\right) + \alpha\left(\dot{y}_3 - \dot{y}_1\right)$$
$$\ddot{y}_2 + \lambda\left(y_2^2 - k\right)\dot{y}_2 + \omega_0^2\, y_2 = \alpha\left(\dot{y}_1 - \dot{y}_2\right) + \alpha\left(\dot{y}_3 - \dot{y}_2\right)$$
$$\ddot{y}_3 + \lambda\left(y_3^2 - k\right)\dot{y}_3 + \omega_0^2\, y_3 = \alpha\left(\dot{y}_1 - \dot{y}_3\right) + \alpha\left(\dot{y}_2 - \dot{y}_3\right)$$



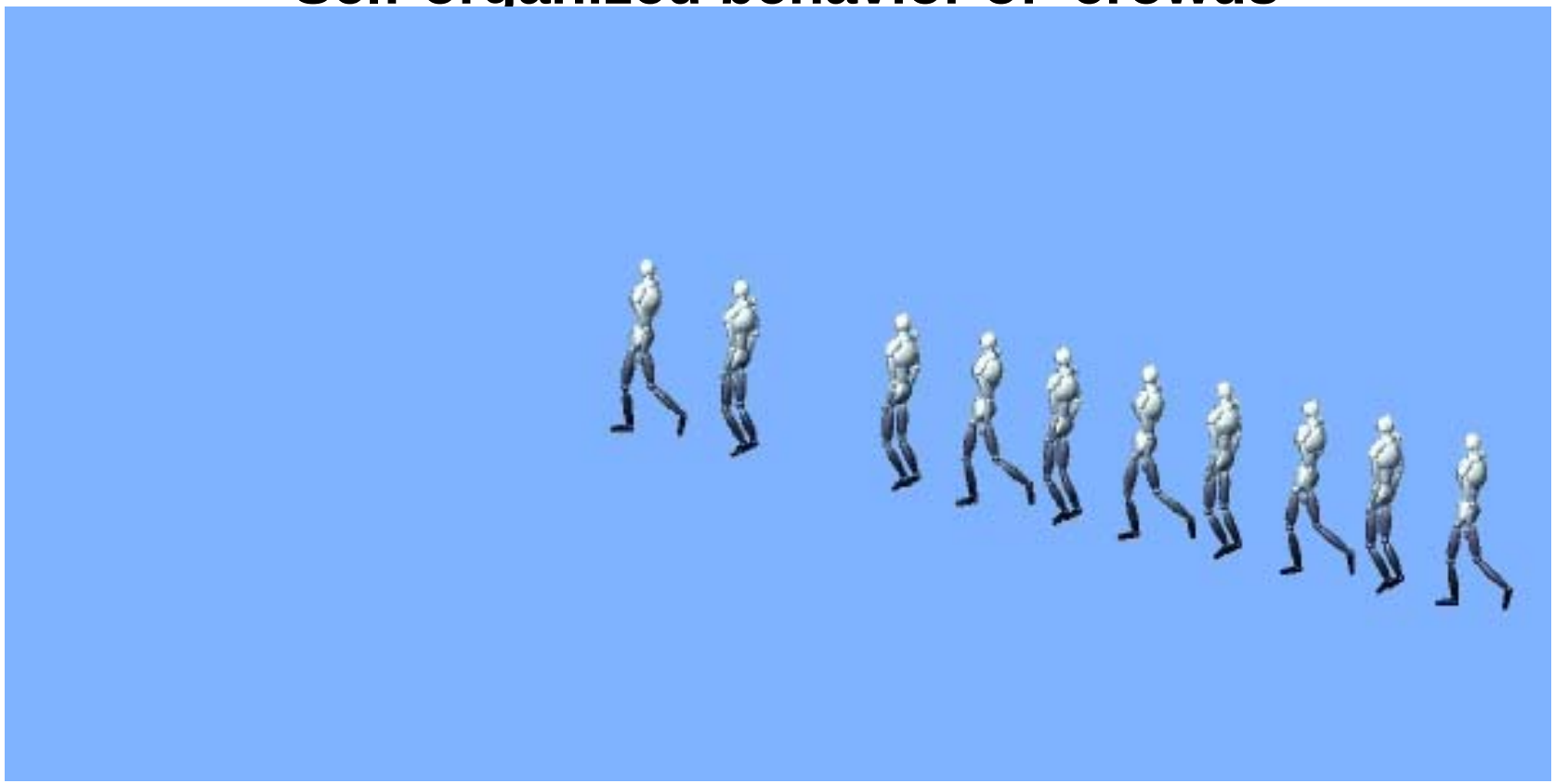
Coordinated crowds

- Introduction of similar couplings between multiple characters.
- ‘Leader’ that entrains movements of other characters.
- Simulation of coordinated behaviour of ‘crowds’.



Coordinated crowds

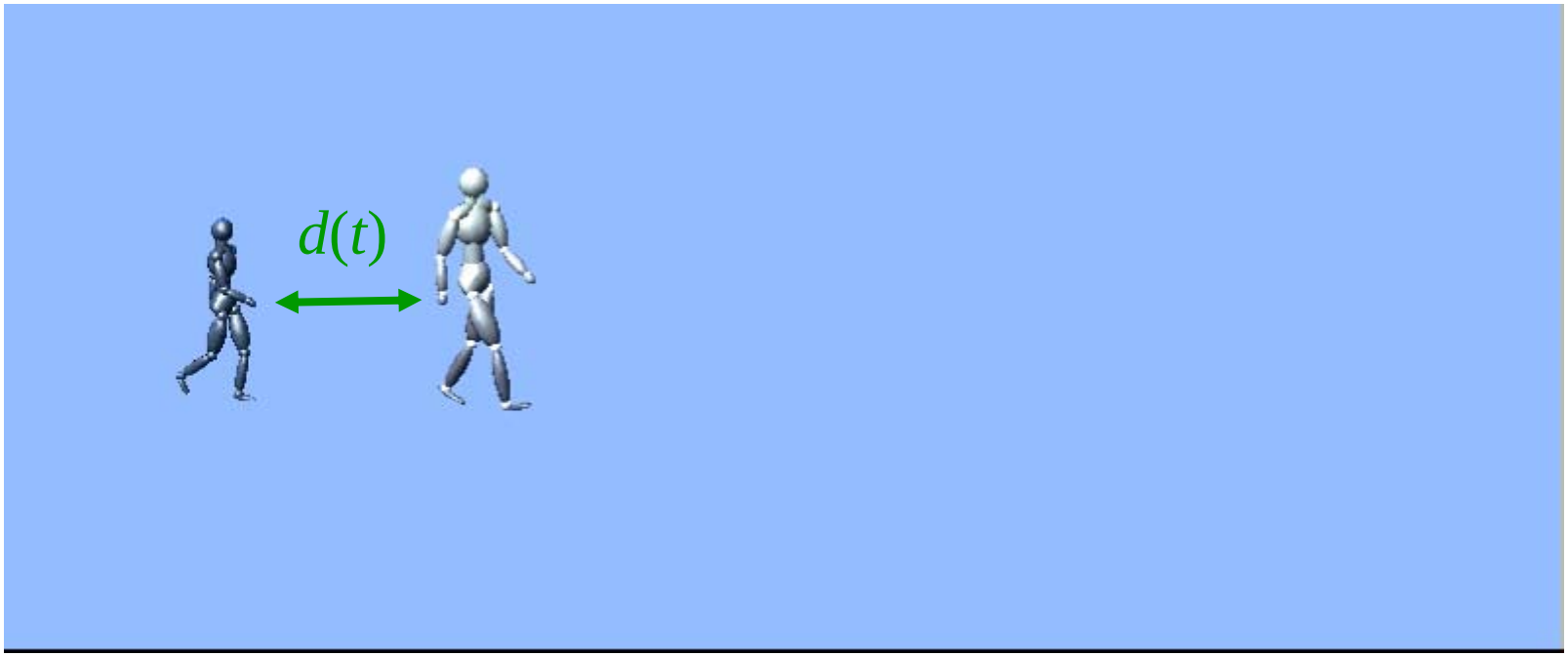
Self-organized behavior of ‘crowds’



- Body translation + rotation determined from foot contact constraints.

Reactive speed control

Following behavior



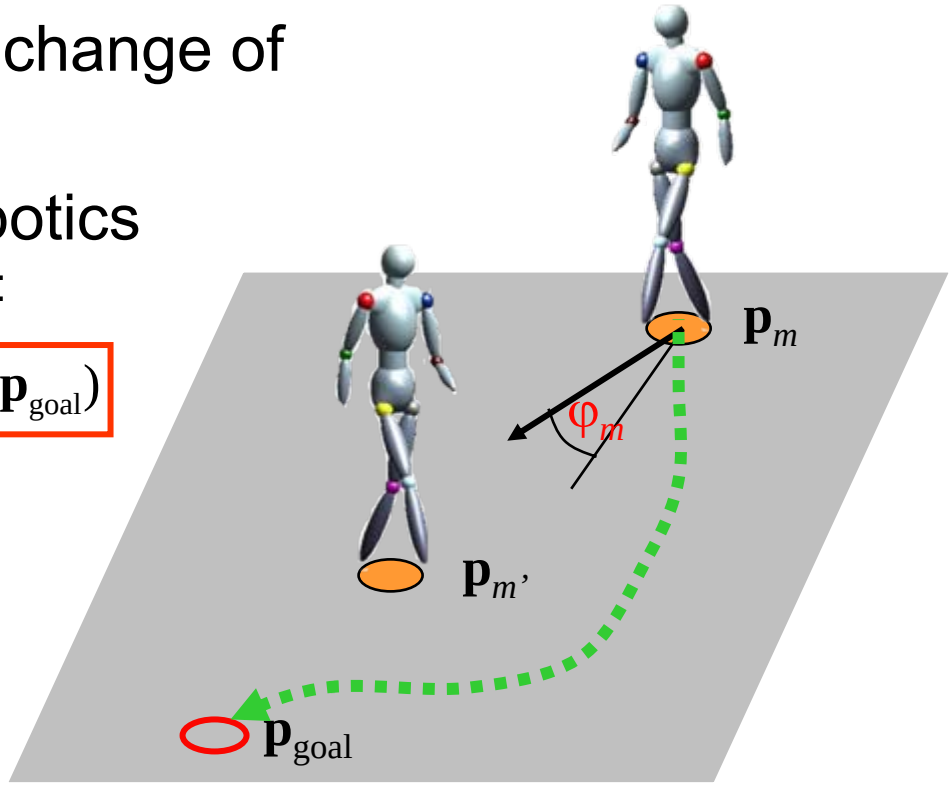
- Distance-dependent adaptation of eigenfrequency:

Change of weights: Navigation

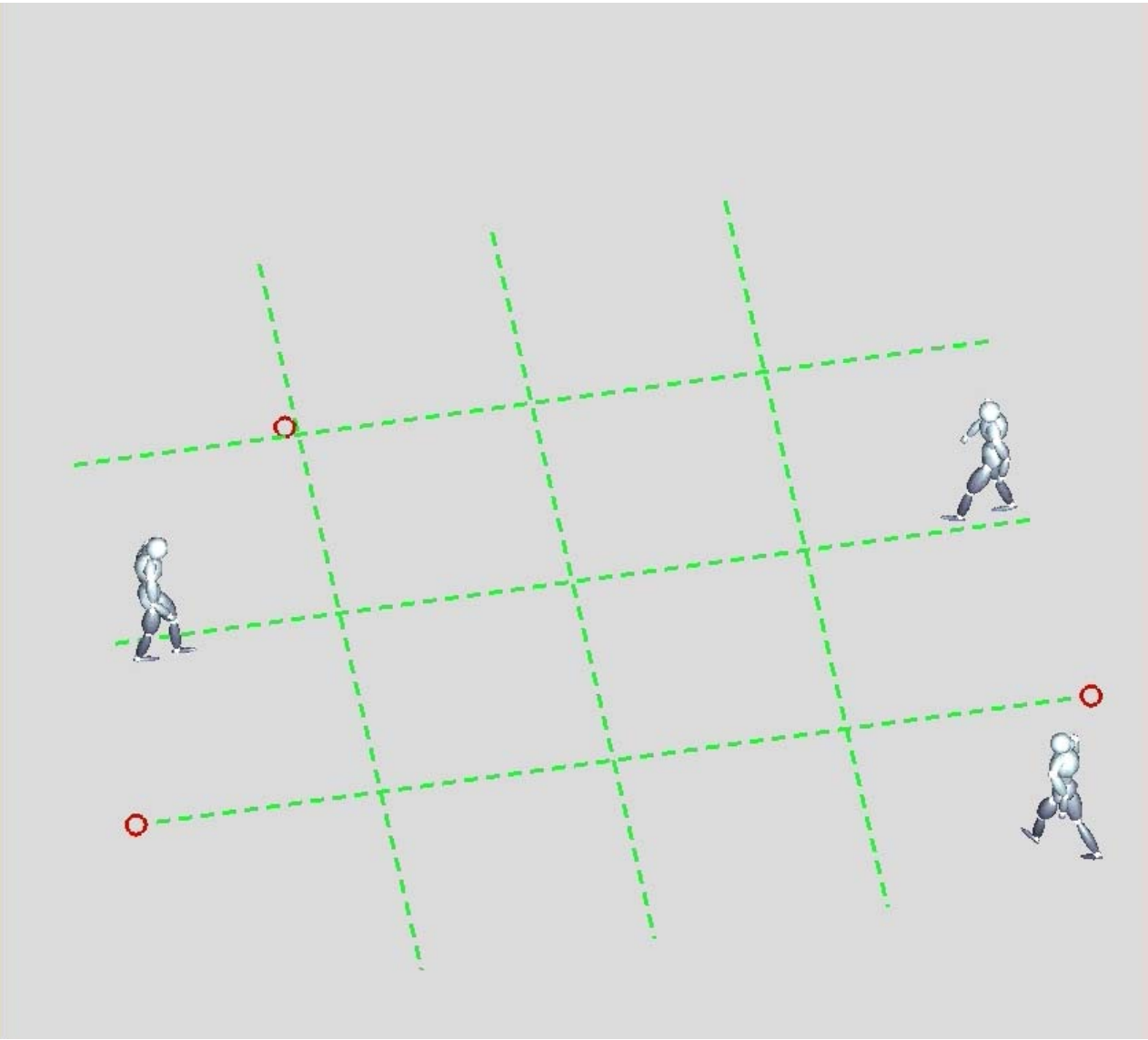
- Navigation by **morphing** between straight and curved gaits.
- Morphing weight depends on change of heading direction $\dot{\varphi}_m(t)$.
- Navigation dynamics from robotics (Schöner & Dose, 1995; Warren, 2006):

$$\dot{\varphi}_m(t) = h_{\text{obstac}}(\varphi_m, \mathbf{p}_1, \ldots, \mathbf{p}_M) + h_{\text{goal}}(\varphi_m, \mathbf{p}_m, \mathbf{p}_{\text{goal}})$$

- Combination with dynamic adaptation of emotional style.



Navigation+style control

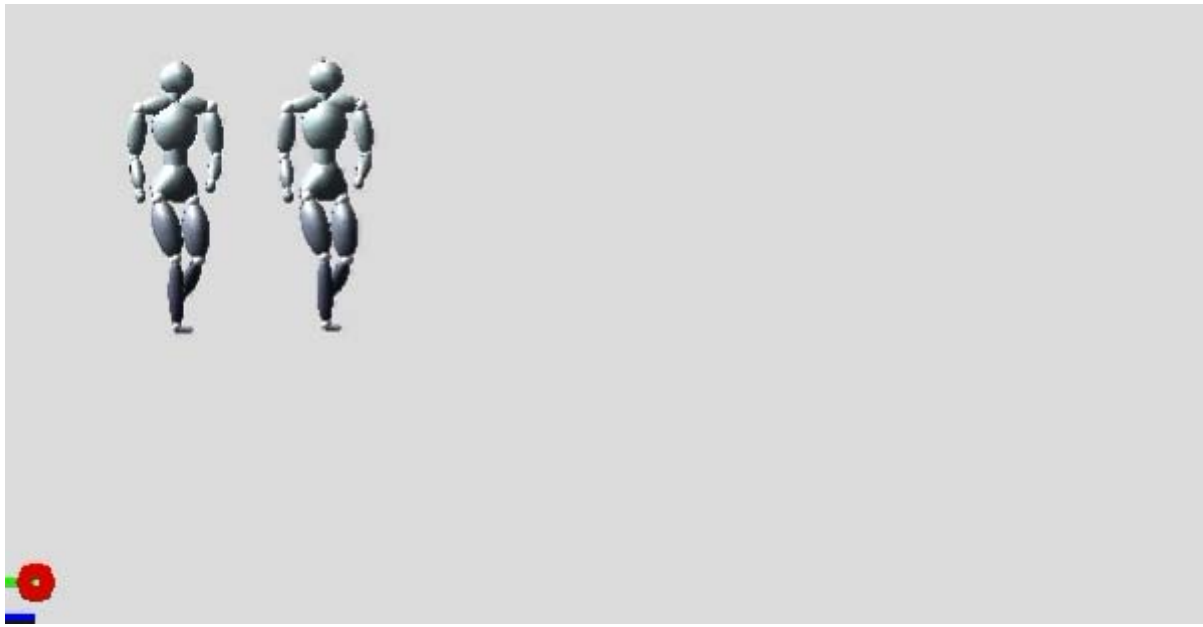


- Navigation towards goal points.
- Avoiding other avatars.
- Reactive change of emotional style (**style morphing**).

Goal point

Change of delays: Mixing periodic+nonperiodic primitives

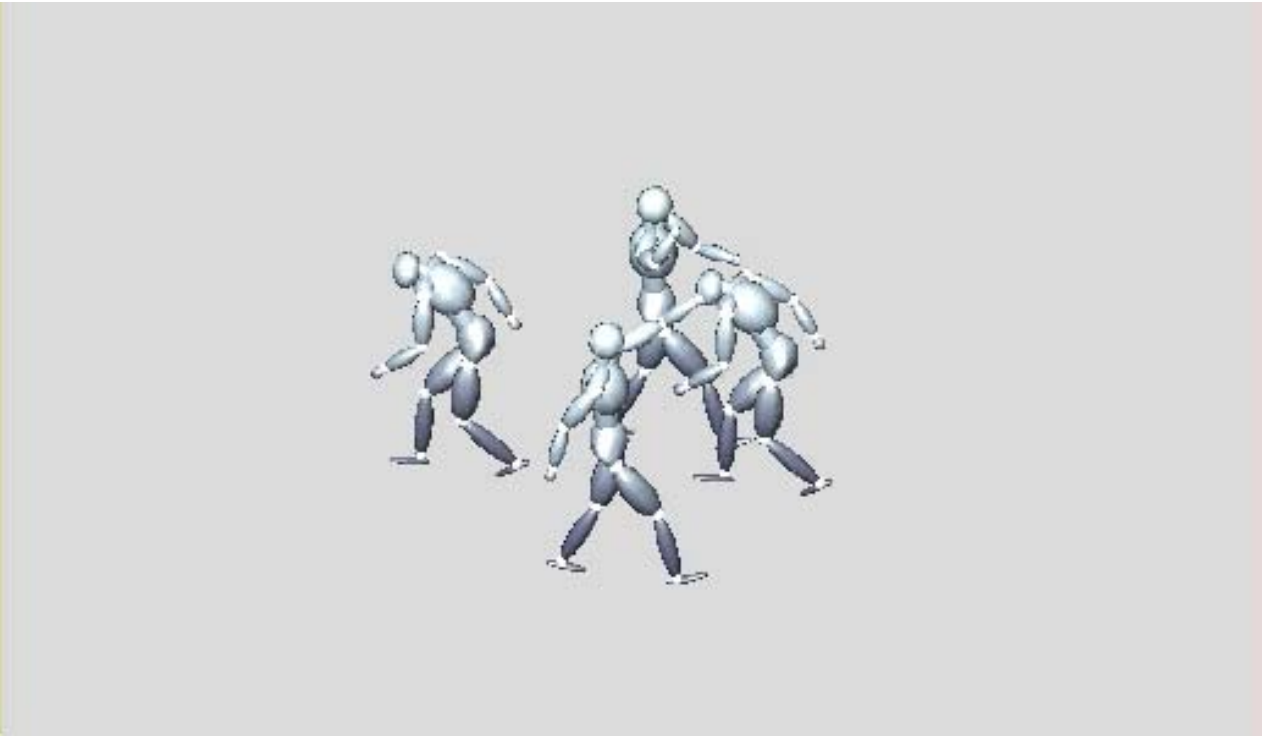
- Data basis with periodic gaits + non-periodic arm movements.
- Learning of periodic + non-periodic sources.
- Non-periodic behavior mapped onto attractor dynamics: $\dot{y}(t) = \eta\, y(t)(1 - y(t)) \qquad y \in [0,1]$



-- external signal

-- state var. $y(t)$

Mixing periodic+nonperiodic primitives

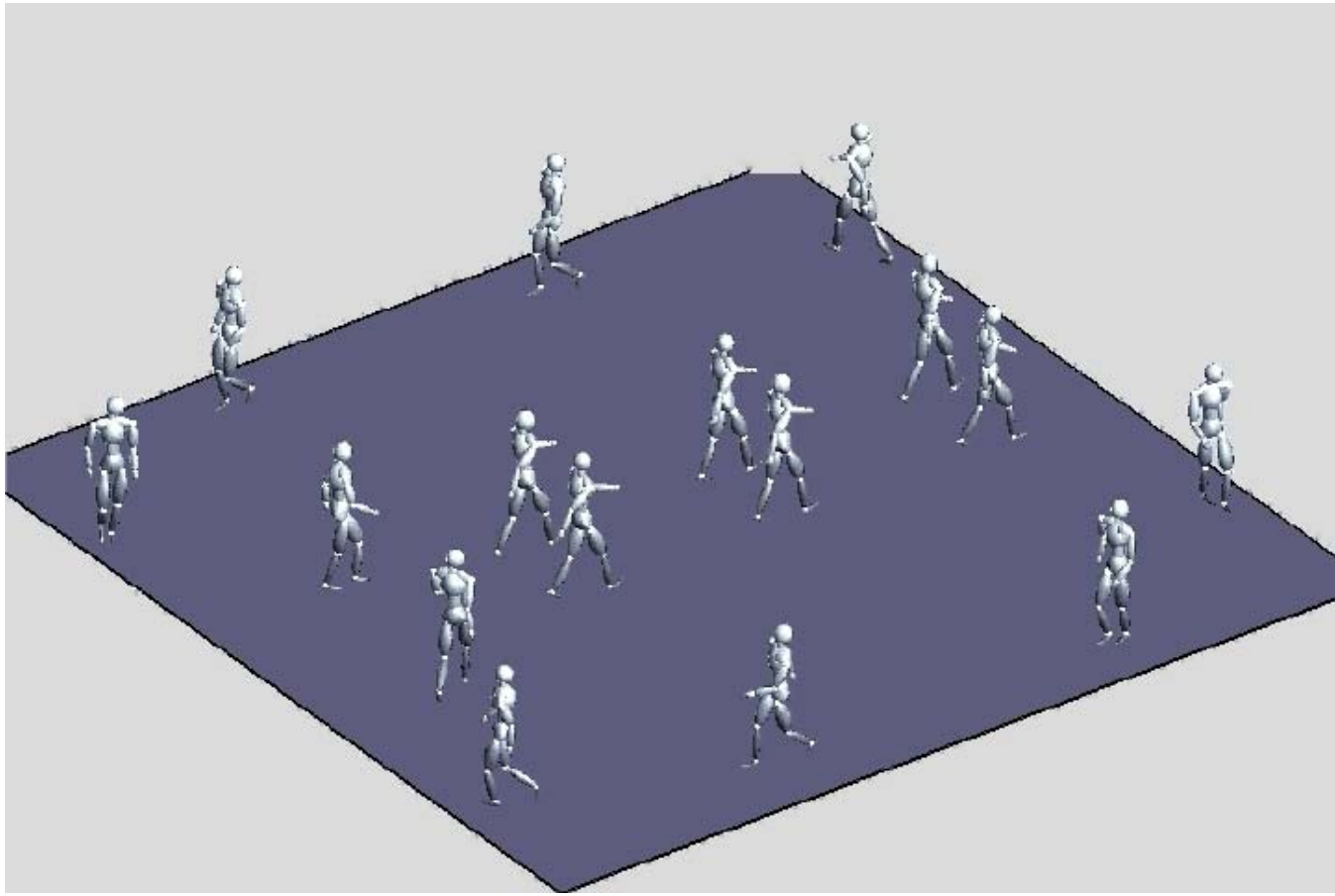


- Part of a dancing scene.
- Two couples taking turns (‘bridge’ / ‘passing through’)
- Self-organized sequence of steps.
- Synchronization with external rhythm (music).

Putting everything together: Complex scenarios

Self-organization of a Welsh folk dance





Summery

- Introduction into shift invariant feature extraction:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} \; s_j(t - \tau_{ij})$$

- New algorithm for anechoic demixing
- Powerful method for feature extraction (image processing, sound demixing)
- Compact representation useful for determining critical emotion specific motion features.
- Capable of realisitc motion synthesis

Thank You

Thank

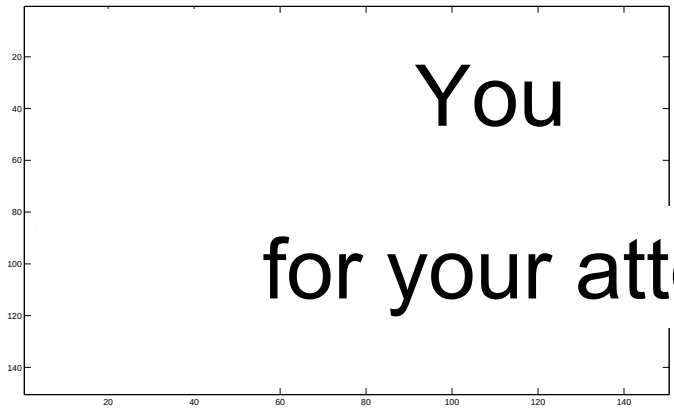
You

Thank

You

Thank
You





for your attention