July 16th - August 10th, 2007 - Istanbul, Turkey
eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

Audio-Driven Human Body Motion Analysis and Synthesis

Project #5

Midterm Project Presentation

hosted by Boğaziçi University, Multimedia Group

Outline

- Project Overview
- Audio Modality : Beat Detection
- Video Modality : Body Motion Capture
- Analysis: HMM on Audio & Video Modalities
- Animation: Full Body Avatar
- Where are we?
- What is left?

Project Overview

eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

- Audio-driven human body motion analysis towards automatic and realistic audio-driven avatar synthesis
  - In the context of a dance performance,
    - Analyze the relations between the music and the body movements à correlation model
    - Synthesize the correlated dance movements when driven with any musical piece of the same genre
    - Animate a dancing avatar

Database Preparation

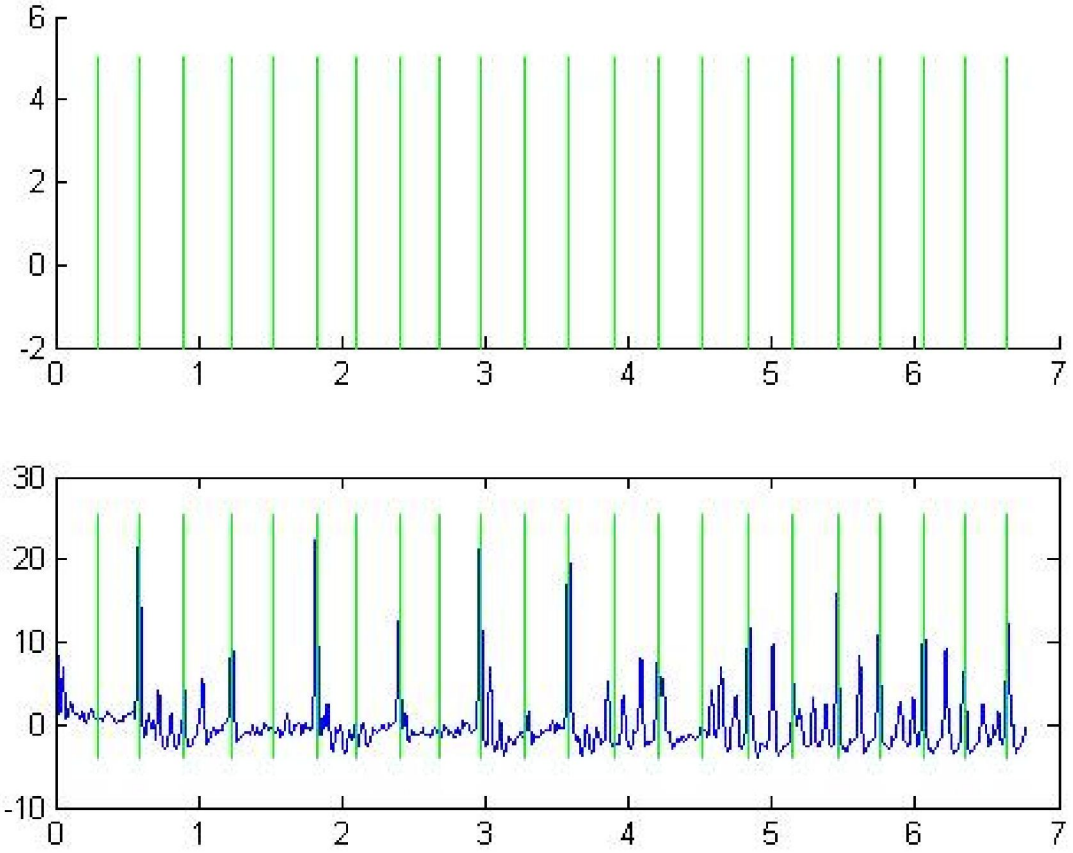
- Audiovisual multicamera recordings
  - *Salsa*
  - *Belly dance*
  - *Isa dance from Canary Islands*
  - *Zeybekiko*
  - *Ispanyol Kasabı*
- Each recording ~ 3 mins.
- One or two dance figures at most

Audio Modality: Beat Detection

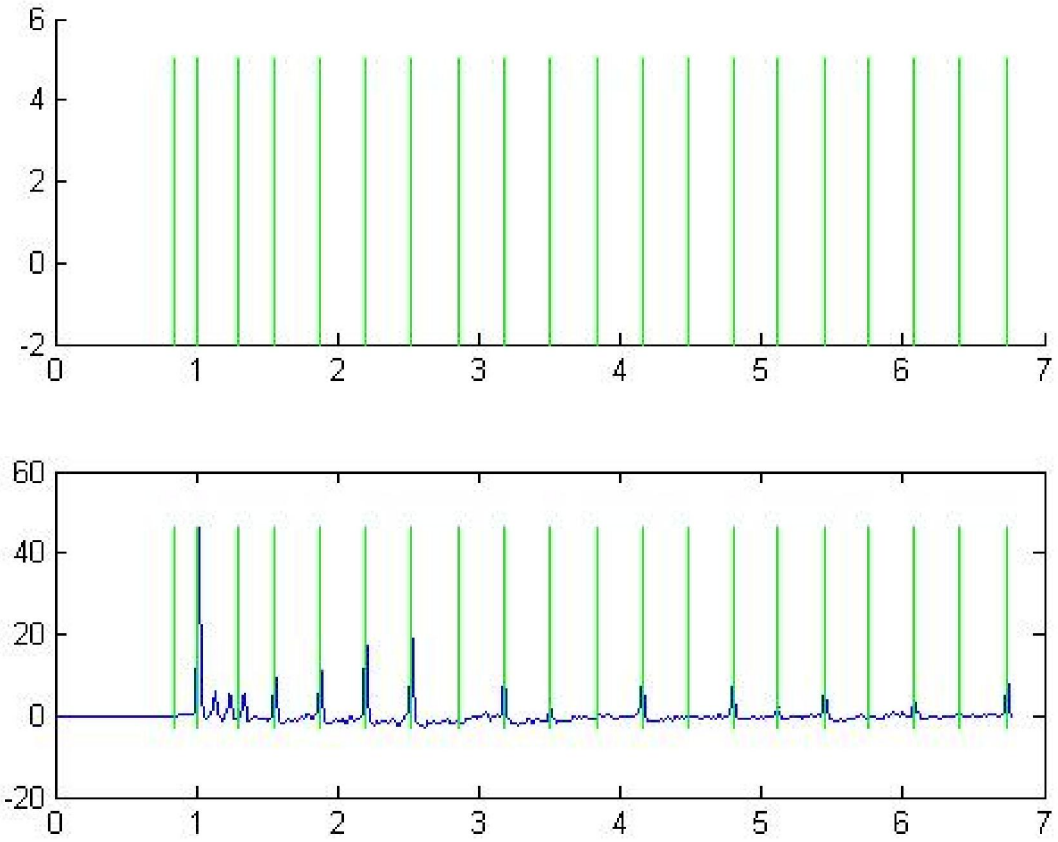
eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

- *Onset Detection*: Computing a detection function based on the spectral energy flux of the input audio signal.
- *Periodicity Estimation* : The periodicity of the detection function is estimated using autocorrelation function
- *Beat Location Estimation* : The position of the corresponding beats is obtained from the cross-correlation between the detection function and an artificial pulse train.

Audio Modality:
Beat Detection for Belly Track



Audio Modality:
Beat Detection for Salsa Track



Video Modality: Body Motion Capture

eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

§ Two approaches are studied in these project:

Technologies Comparison

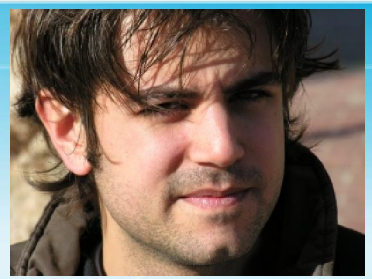
- **Marker based approach:** A technologies widely used in the cinema industry based on placing landmarks on the body.
  - **Pros:** It may produce very accurate results
  - **Cons:** It requires intrusive markers
- **Markerless approach:** No markers are used to extract.
  - **Pros:** It is not intrusive since it does not require any marker
  - **Cons:** May produce not so accurate results

Video Modality: Body Motion Capture

eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

- Tracking WITH markers
  - Learn a Gaussian color model for the markers
  - Track the markers with Kalman filtering

Video Modality: Body Motion Capture



eNTERFACE '07
The SIMILAR NoE Summer Workshop on Multimodal Interfaces

- Tracking WITHOUT markers
  - Based on a 3D reconstruction of the scene as the input data for the analysis system
  - An annealing Monte-Carlo strategy is employed
  - Simulations are in the oven

Analysis:
Audio & Video Modalities

eNTERFACE '07
The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

§ Testing HTK for non-audio files

- ú Convert any kind of input files into HTK format (Matlab)
- ú Define HMM structure (left-to-right)
- ú Train with any kind of input data: .batch script
  - Test Scenario:
    - Defined left-to-right HMM with 3 emitting states, 2 data parameters, each one modeled by one Gaussian distribution
    - Generate data sequences according to this model
    - Train a 3-stage HMM using this data with HTK
    - Compare the values of the parameters of the original HMM and those of the trained HMM
- ú worked very well with synthetic data

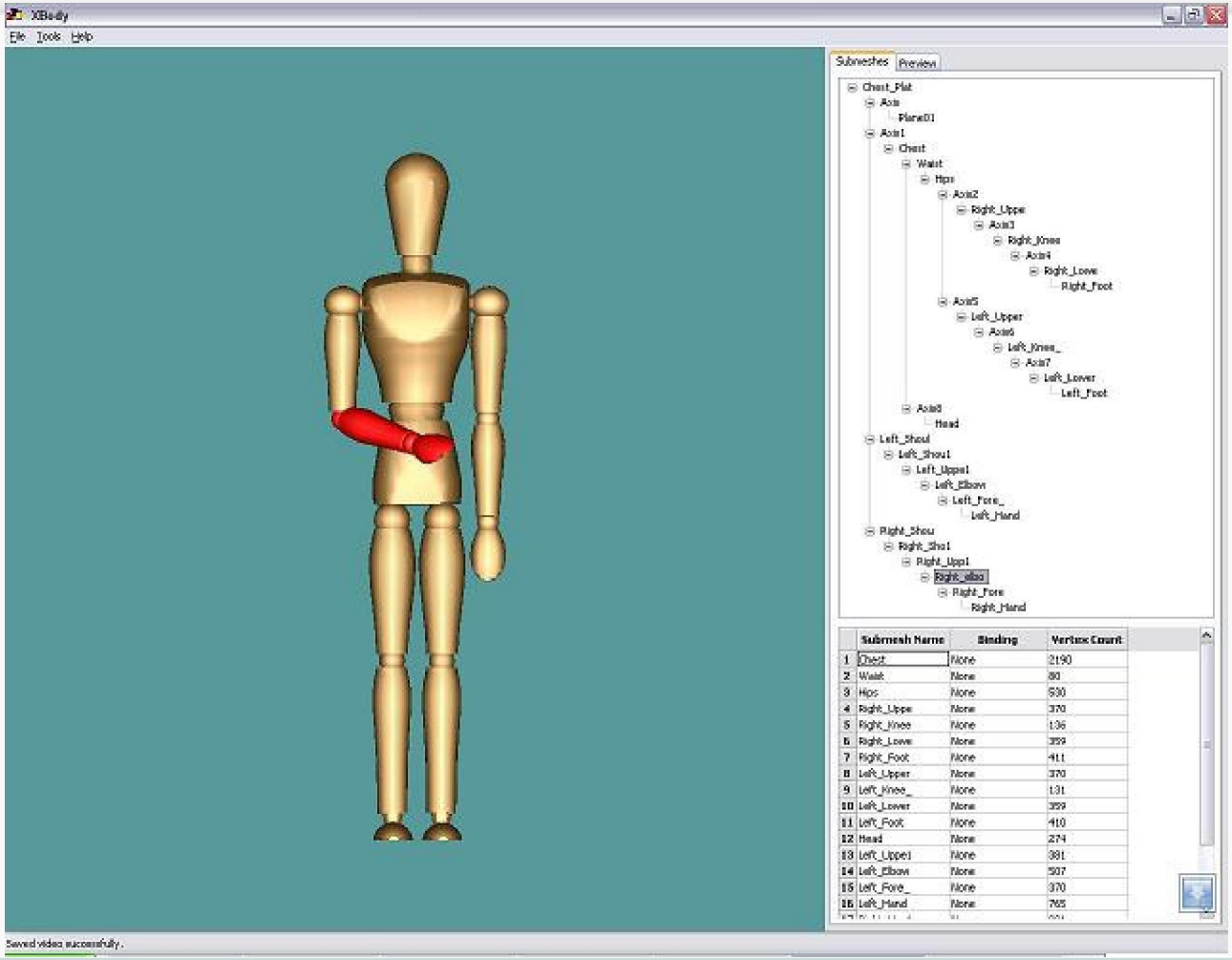
  à use it with “dance data”!

Animation

eNTERFACE '07
The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

- Implementation of a 3D dancing avatar.
  - Forward Kinematics rendering of angular data from analysis & HMM based synthesis.
  - OpenGL and wxWidgets based platform independent standalone application.
  - Using Douglas F. Woodward as 3D model, but can use any 3D model in 3DS format.
  - 70% complete:
    - 3DS file importing is done
    - Scene graph implementation (forming scene tree and transform & rendering pipeline)
    - Selection of submeshes and previewing simple animations
    - Animation can be saved as video (in AVI format)

Animation


XBody
File Tools Help
Submeshes
Preview
Chest_Plat
Axis
Plane01
Axis1
Chest
Waist
Hips
Axis2
Right_Upper
Axis3
Right_Knee
Axis4
Right_Lower
Right_Foot
Axis5
Left_Upper
Axis6
Left_Knee_
Axis7
Left_Lower
Left_Foot
Axis8
Head
Left_Shoul
Left_Shou1
Left_Upp1
Left_Elbow
Left_Fore_
Left_Hand
Right_Shou
Right_Shoi
Right_Upp1
Right_elbo
Right_Fore
Right_Hand
Submesh Name
Binding
Vertex Count
1 Chest None 2190
2 Waist None 80
3 Hips None 530
4 Right_Upper None 370
5 Right_Knee None 136
6 Right_Lower None 359
7 Right_Foot None 411
8 Left_Upper None 370
9 Left_Knee_ None 131
10 Left_Lower None 359
11 Left_Foot None 410
12 Head None 274
13 Left_Upper1 None 381
14 Left_Elbow None 507
15 Left_Fore_ None 370
16 Left_Hand None 765
Saved video successfully.

Where are we?

	Week 1							Week 2							Week 3							Week 4				
Work Packages	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10
WP1 : Data Acquisition																										
WP2 : Body Motion Tracking																										
WP3 : Body Motion Feature Extraction																										
WP4 : Audio Feature Extraction																										
WP5 : Motion Analysis																										
WP6 : Audio Analysis																										
WP7 : Joint Analysis																										
WP8 : Synthesis and Animation																										
Report																										

	Week 1							Week 2							Week 3							Week 4				
Work Packages	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10
WP1 : Data Acquisition																										
WP2 : Body Motion Tracking																										
WP3 : Body Motion Feature Extraction																										
WP4 : Audio Feature Extraction																										
WP5 : Motion Analysis																										
WP6 : Audio Analysis																										
WP7 : Joint Analysis																										
WP8 : Synthesis and Animation																										
Report																										

What is left?

- Analysis:
  - Supervised training of audio and body motion
    - Determine and label the meaningful video segments by inspection
    - Learn audio and body motion patterns
- Synthesis:
  - VQ based clustering of input audio
  - Decide on audio pattern
  - Generate dance figures according to the specified audio class
- Animation:
  - Parse the results coming from synthesis and let them drive the animation

The End