I danced With Machine
Distributed Multimodal Interaction…

Guillaume Hutzler* — Bernard Gortais — Philippe Joly*** — Yann Orlarey**** — Jean-Daniel Zucker****

** Laboratoire de Méthodes Informatiques (LaMI)*
Université Evry-val d'Essonne
523, Place des Terrasses, 91000 Evry
hutzler@lami.univ-evry.fr

*** Laboratoire d'informatique de Paris 6 (LIP6)*
Université Paris 6
8, rue du Capitaine Scott, 75015 Paris
{bernard.gortais, jean-daniel.zucker}@lip6.fr

**** Institut de Recherche en Informatique de Toulouse*
Université Toulouse 3
118, route de Narbonne, 31062 Toulouse Cédex 4
joly@irit.fr

***** Grame*
9, rue du Garet, 69000 Lyon
orlarey@rd.grame.fr

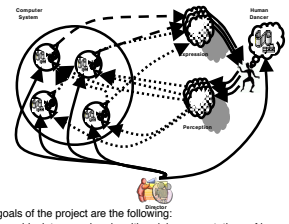
Multimodal Interaction between a Human Dancer and a Distributed Computer System

The global goal of the project is to put into place an experimental set to enable a dynamic multimodal dialog between a human dancer and a distributed computer system.

A dialog

The interaction would proceed this way:

1. the dancer proposes a choreographic sequence;
2. the system analyses the performance of the dancer and chooses an appropriate answer;
3. the system executes the chosen answer by means of graphical and musical expressions, and potentially by any type of action that the objects composing the system may be capable of;
4. the dancer perceives what the system is doing and reacts to it; dialog proceeds at step 1;



Scientific issues

The main scientific goals of the project are the following:

- developing hierarchical, temporal and multimodal representations of human behaviors based on the notion of quality, which is inspired by the analysis of various artistic fields of expression;
- developing the models and algorithms necessary to a hierarchical and real-time analysis of gestures using 2D video
- building a distributed platform that would enable
  - the dynamic interconnection of communicating objects
  - a collaborative distributed interpretation of a complex and dynamic situation
  - a collective multimodal reaction

A secondary goal deals with the production of multimodal performances by the system, in particular by means of musical and graphical expression

Experiments and validation

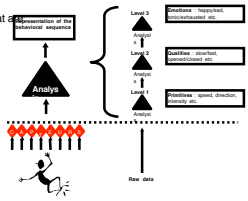
The experimental set with the dancer must be seen as a metaphor of a near future everyday life in which users will have to interact with a great number of electronic devices. Validation will be obtained by:

- carefully isolating the different steps of analysis
- designing controlled interaction situations

Hierarchical, Temporal and Multimodal Description of Behaviors

One of the aims of the project is to develop concepts and tools that would allow to design a ***multimodal ontology*** for the description of behaviors.

- Analysis, on a semantic basis, of the various languages that are manipulated by artists:
  - corporal and gesture languages
  - graphical language
  - musical language
- Description of behavioral sequences by means of ***primitives*** and ***qualities***:
  - primitive: a movement is measured to be done at a given speed (objective measure)
  - quality: depending on people, this movement will be rated as being slow or fast (subjective measure)
- Design of specific representations that are
  - temporal
  - multimodal
  - hierarchical



Real-time Analysis of movements

The goal is to perform a real-time analysis of the movements of a person

- in a daily environment,
- with the use of a single video camera
- without specific equipment on the person

This can be decomposed in 2 sub-goals:

- the definition of a human body model and the design of the corresponding algorithm that must enable to extract the parameters of the model from the 2D video picture;
- the design of techniques so that the computer system can learn the defined ontology from examples (enabling it to associate speed measures to subjective qualities of slowness of rapidity);

Hierarchical Real-time Analysis of Gestures Using 2D Video

The goal is to define a human body model and to design of the corresponding algorithm that must enable to extract the parameters of the model from the 2D video picture. In order to have usable results, this model must allow :

- a real-time treatment, which implies a multi-scale model and tools to handle errors
- the detection of mobile , potentially sequential, phenomena, whose dynamicity may convey a sign to the system
- the production of a specific description (dance gestures) for the interpretation subsystem

Real-time qualitative analysis of movement

The goal of this first phase is to position:

- the analysis environment
- the video data treatment kernel
- the mechanism to transfer the information to the decision sub-system

The extracted information will correspond to a set of primitive descriptors about the dancer localization and movement

- bounding box
- direction of movement (up/down, left/right, front/back)
- dynamicity (acceleration/slowing down)

\+ information about quality

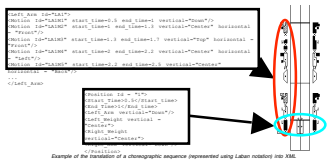
Descriptive analysis of movement

The goal of this second phase is to produce a description of the dancer postures, which requires:

- a description of the position of the different parts of the body
- a semantic interpretation of the succession of the choreographic movements

To achieve this the formalism will be inspired by:

- bio-mechanical models developed for avatars synthesis
  - human body decomposed as segments and links
  - hierarchical decomposition
  - XML-like notation
  - ex: SNHC (MPEG), VRML (W3C)
- formalisms developed by the dancing community
  - description of movement in a prescriptive or analytical way
  - movement of a segment defined relatively to a reference position by an angle and a duration
  - e.g.: Laban notation



Example of the translation of a choreographic sequence (represented using Laban notation) into XML

© IRIT – Thomas Foures

Extraction of the position of the five segments (2 arms,2 legs, upper body + head) from the analysis of a static image extracted from a low quality 2D video recording

Distributed Multi-Sensor Perception

The distributed system to which the dancer will be confronted is distributed by nature, potentially composed of a lot of communicating objects that are characterized by:

- some computing power
- communicating capabilities
- perception and/or action capabilities

Several issues are associated to this natively distributed architecture

- the communicating objects must have the ability to be dynamically interconnected with one another; research will focus on:
  - mechanisms to give IP addresses to each communicating object
  - service directories so that the objects can dynamically find other objects that satisfy their needs;
- the designer and/or director may want to specify some of the relationships between objects. To this end, we may:
  - take inspiration from UML-RT notations to specify the architecture of the system
  - develop distribution mechanisms so that the corresponding architecture automatically sets up
- collaboration mechanisms must be developed so that the objects collectively act in a coherent way on top of individual and local behaviors. In particular, they must:
  - build a global representation of the performance of the dancer
  - collectively decide the reaction to have
  - collectively perform the reaction in a multimodal way

Reconfiguration of the environmental ambiance

Symmetrically to the perception of the performance of the dancer, the system is required to react by adapting the environment accordingly, especially the visual and musical ambience.

Visual synthesis

The work on the distributed creation of a visual ambience will take inspiration from previous works made on the automatic generation of graphical ambience in reaction to the meteorological conditions of a distant place (*The Garden of Chances*, *Mutations*). These works are based on:

- a distributed composition relying on graphical agents characterized by their shape and color
- knowledge given to agents about how to translate some given data into color and/or shape modifications
- capabilities given to agents to interact with one another so as to locally adapt depending on the neighborhood

In addition, agents will have to learn how to translate qualities into visual ambiences depending on the taste of the director

Musical Synthesis

The musical synthesis part is composed of two aspects:

- a modular software synthesizer will be in charge of the sounds synthesis. It deals with:
  - reading of pre-recorded sonorous samplings
  - sounds synthesis
  - analysis and transformation of sounds recorded on stage
  - spatialization effects
- a musical composition environment will provide the tools to temporally organize the commands for the synthesizer
  - the programming will be a mix of functional and visual programming techniques
  - the programming model will correspond to what is called "homogeneous programming" : data structures are enriched with abstraction and application capabilities of λ-calculus

Ontology learning

In order to have a semantic description of the performance of the dancer, it is necessary to have the system know what a fast or slow movement is, an opened or closed movement and so on. What is learned is the ontology defined in terms of qualities and how it can be related to the primitives measured.

This task presents several specificities:

- few examples are available: this will lead to the use of a simplified, abstract representation
- the examples correspond to sonorous, visual, and temporal information: we will use a multi-instances representation
  - more expressive than attribute-value representations
  - less computationally expensive than 1st order logic representations
- the definitions to be learned by the system must be understandable by the director: to this end, we will use a system that works on decision lists