

From Simulated Dialogues to Interactive Performances

Elisabeth André
University of Augsburg &
German Research Center for Artificial Intelligence

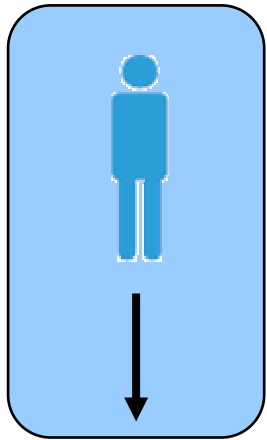


Phenotype / Metaphor

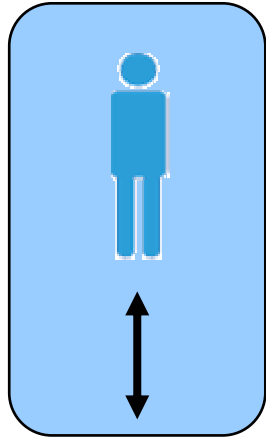


Types of User Interfaces with Conversational Embodied Agents

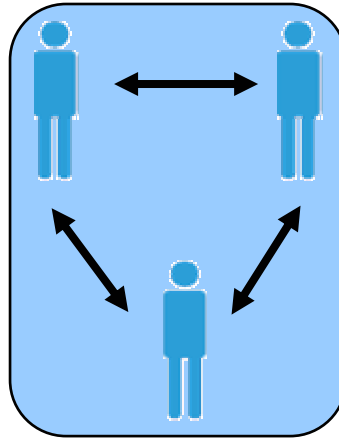
**non-
interactive
characters**



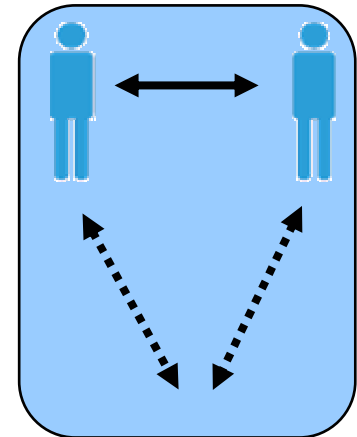
**interactive
characters**



**character
teams**

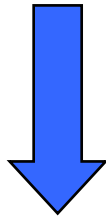


**interactive
performances**



Basic Idea: Speech-Act Theoretic Approach

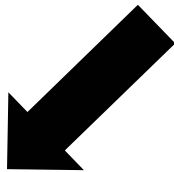
“Not only the generation of spoken language, but also the presentation of multimedia material can be considered as a sequence of communicative acts to achieve certain goals”
(André & Rist 1990)



Exploitation of Planning Methods for Automated Script Generation

Design of Believable Body Language

extraction of rules and equivalence classes of gestures from annotated video by means of qualitative and quantitative analysis



creation of animation clips for each equivalence class



operationalization of rules within behavior engine

ANVIL: Multi-Track Annotation of Video and Language

rheto-
rical
relations

...

heme-rheme

trans-
literated
language

...

postures

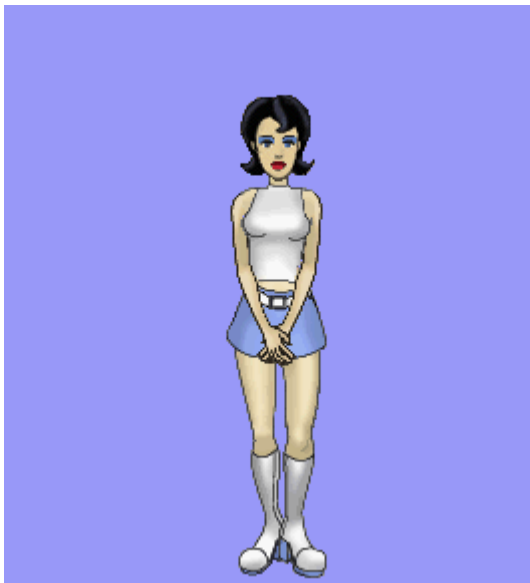
The screenshot displays the ANVIL software interface, which is divided into several windows. The top-left window is a menu bar with 'File', 'Edit', 'View', and 'Debug' options, followed by a list of actions: 'open file lq4-reich.anvil', 'created new annotation object', 'read anvil file', 'create player for: video.quicktime', 'FrameRate = 0.0', 'Movie duration: 195.84', and 'No. of frames: 4895'. Below this, it shows 'Frame number: 179', 'Media time: 00:07:14', and 'Modified: false'. The top-middle window is a video player showing a man in a suit and glasses, with a 'ZDF' logo in the top-left corner. The top-right window is a 'TAGGER' window with fields for 'Track: trl', 'Start: 00:06:63', and 'End: 00:07:22'. It also has 'Attributes' (token: auf, emphasis:, number of syllables:), a 'Comment' field, and buttons for 'start', 'end', 'cut', 'edit', and 'del'. The bottom window is the 'ANVILATION' window, which is a complex multi-track annotation interface. It features a timeline with various colored bars representing different annotation layers. The layers include 'take' (with 'Loff' and 'Löffler' labels), 'trl' (with a long sequence of small colored bars), 'ling' (with 'rst', 'theme-rheme', and 'topic' sub-layers), 'head' (with 'ling' and 'rheme' labels), 'posture' (with 'upper' and 'lower' sub-layers), and 'beat' (with 'left' and 'right' labels). The timeline is marked with a vertical red line at the current media time.

<http://www.dfki.de/~kipp/research/anvil.html>

Classification of Persona Gestures

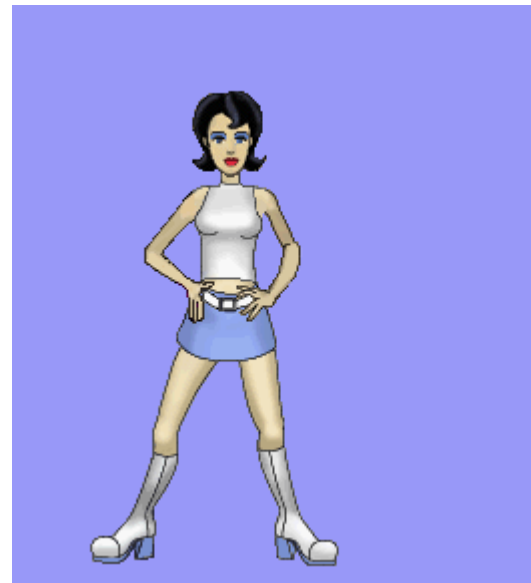
Talking Posture 1

- cautious, hesitant
- appeal for compliance
- avoids body-gestures

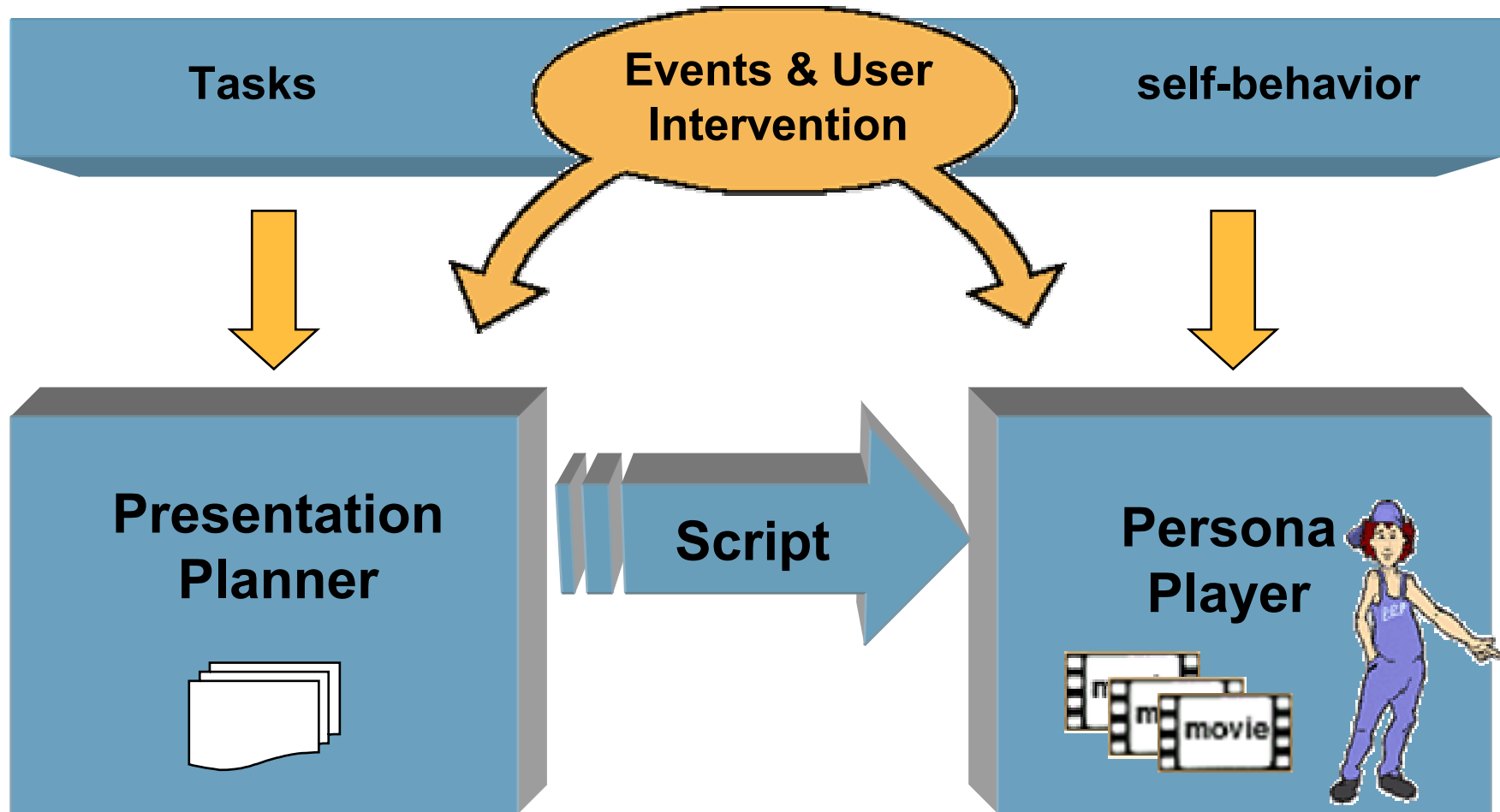


Talking Posture 2

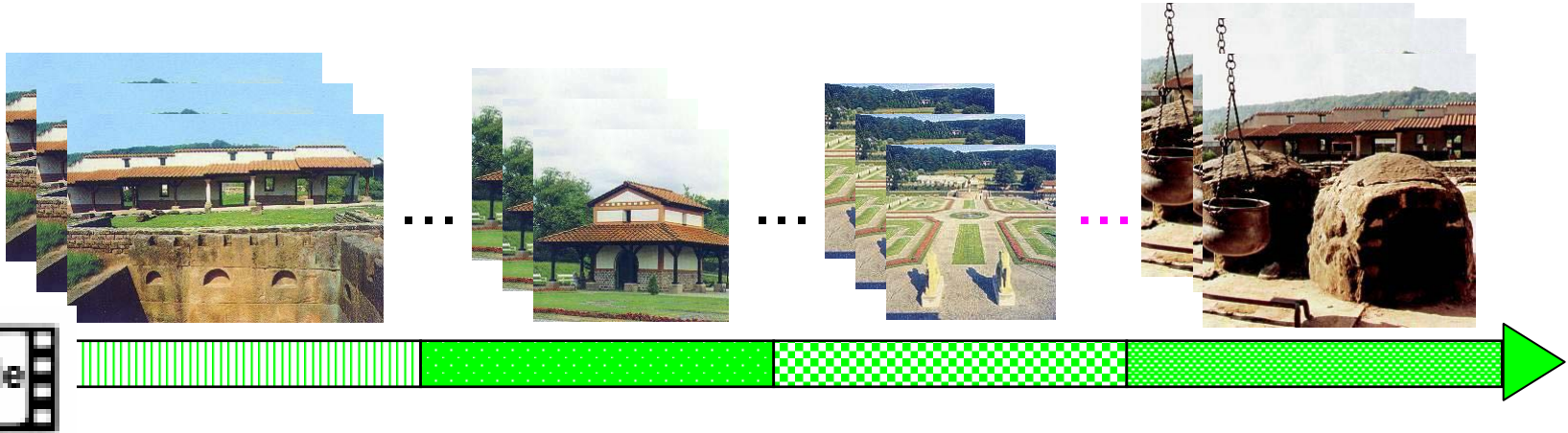
- active, attentive
- self-confident
- uses body-gestures



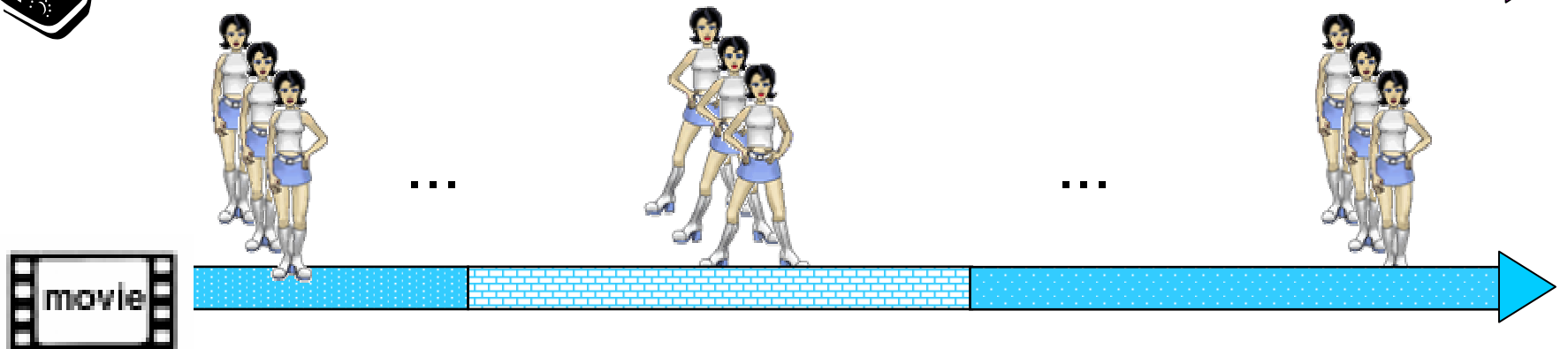
Automated Approach to Character Control



■ Subtask: Temporal Coordination of Media Objects



 *“At archeological finds in Schwarzenacker, ...*



Temporal Constraints to Synchronize Persona Actions with Other Media

(define-plan-operator

:header (A0 (ShowPresentation ?topic))

:constraints

(***and*** (BELP (Describes ?audio ?topic)) (BELP (Audio ?audio))
(BELP (TalkingGesture ?video)) (BELP (RepeatGesture ?video)))

:inferiors

(A1 (PresentPictureSequence (?topic)))
(A2 (SAddSmileCode (?audio)))
(A3 (SAddSmilCode (?video)))

:temporal ((**A2 (d) A1**) (**2 <= begin A2 - begin A1**) (**A2 (e) A3**))

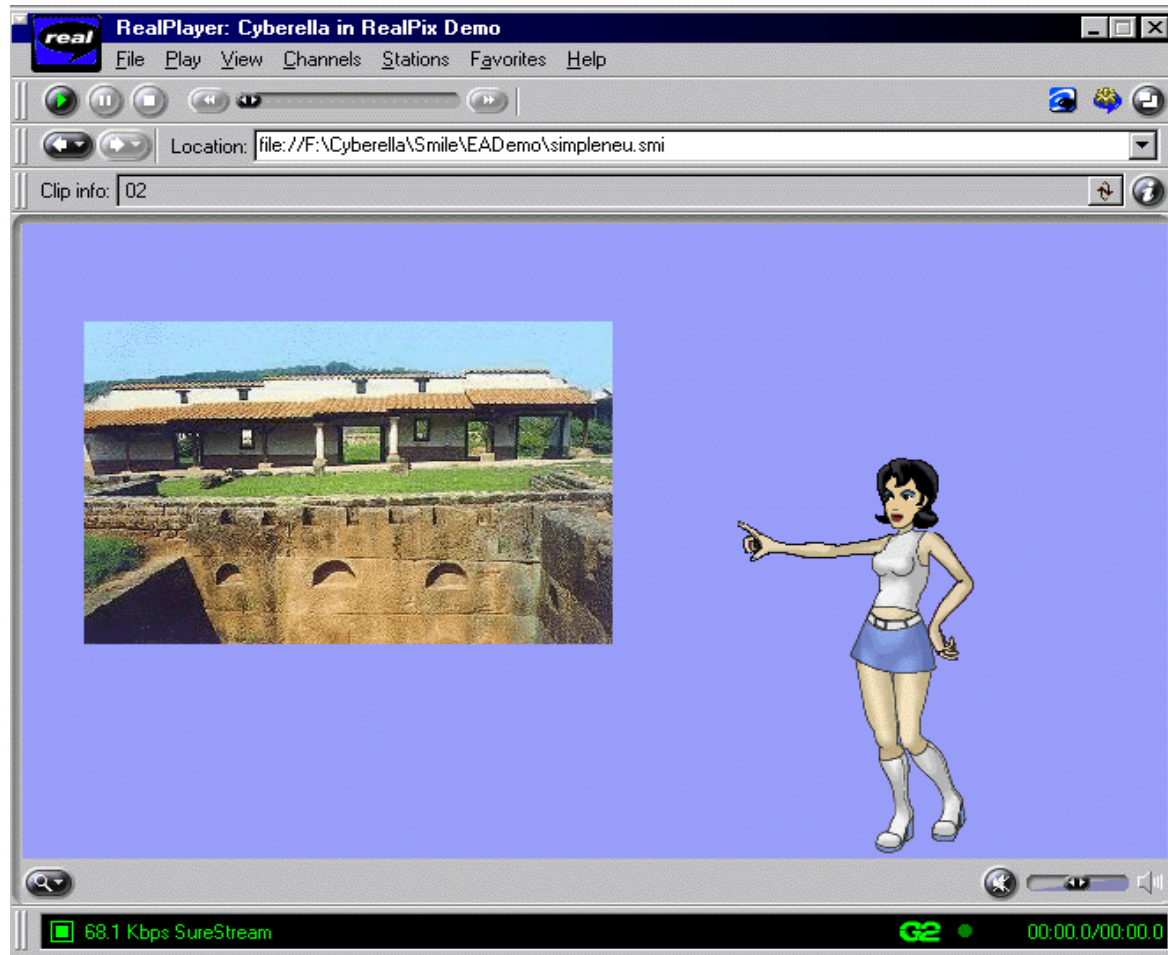
:spatial

((aligntop A1) (alignleft A1)
(1 <= bottom A0 - bottom A3 <= 1)
(1 <= right A0 - right A3 <= 1))

Automatically Generated W3C-SMIL Script for a Presentation with Cyberella

```
<smil>
<head>
  <meta name="title" content="Cyberella in RealPix Demo" />
  <layout>
    <root-layout width="650" height="385" background-color="#9898F8"/>
    <region id="reg471101" top="0" left="12" width="385" height="284"
      background-color="#9898F8"/>
    <region id="reg081591" top="95" left="354" width="295" height="385"
      background-color="#9898F8"/>
  </layout> </head>
<body>
  <par>
    <img region= " reg471101" begin= " 0.0" src= "schwarzen.rp" fill="freeze"/>
    <animation begin= "2.0s" end="15.1" region="reg471101" src= "talk-gst.rp"/>
    <audio begin= "2.0s" end="15.1" src="schwarzen.rm" />
  </par> </body>
</smil>
```

Using a Standard SMIL-Player for Script Execution



Single Presenters: Employed Technology

- Use of a hierarchical planner to automatically generate scripts in XML-based mark-up languages (SMIL, PET ...)
- Event-based and timeline-based synchronization mechanisms
 - » Combination of the planning component with a component for spatial and temporal reasoning to enable the generation of SMIL-documents
- Extension of the presentation planner by hypermedia functions by dynamic node expansion

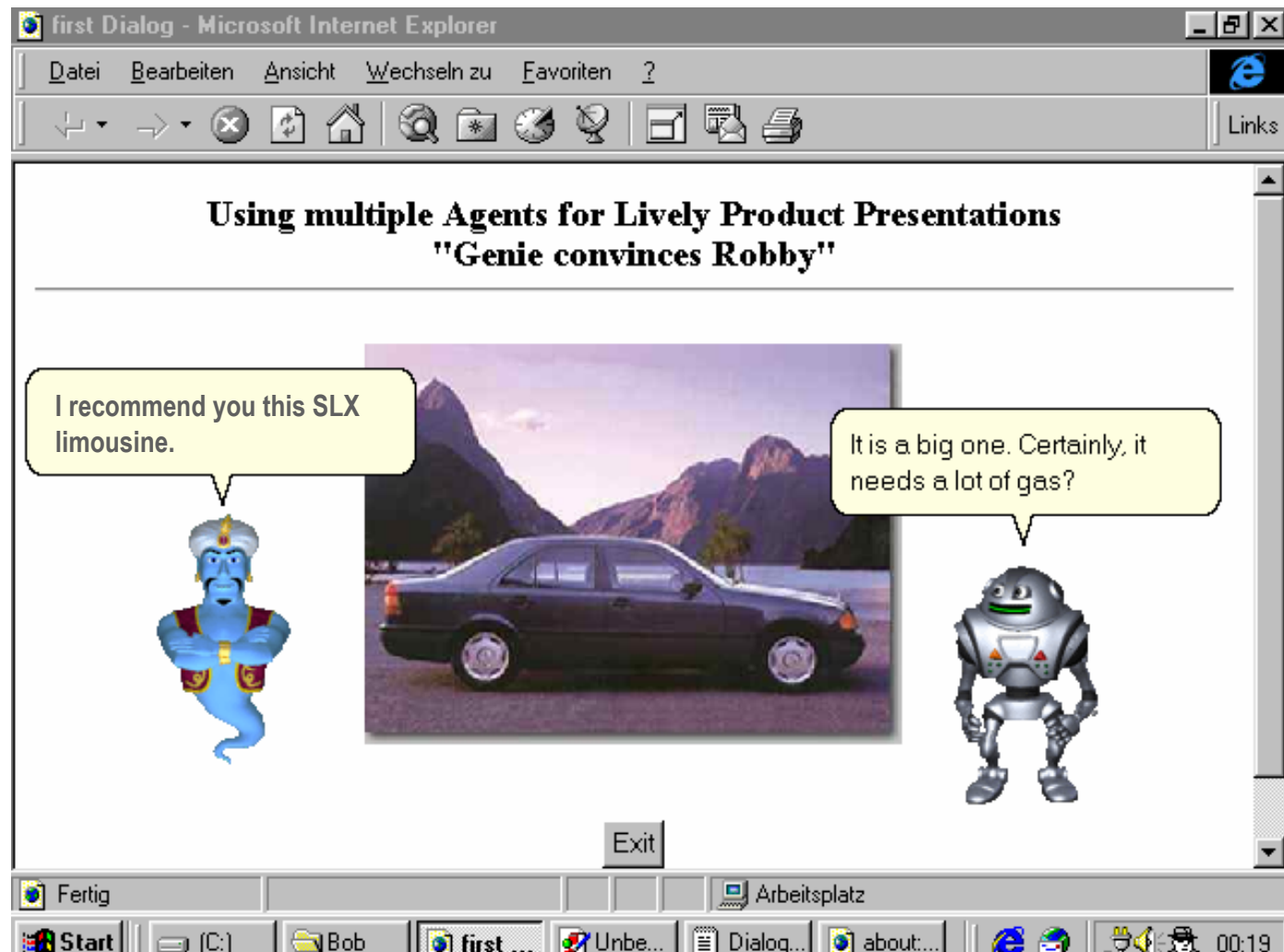
■ From a Single Presenter to Presentation Teams

- New Agent-User Relationship
- User is no longer addressed directly, instead information is conveyed by means of performances to be observed by the user

Why to use multiple presenters?

- Presentation teams convey certain **rhetorical relationships** in a more canonical way
 - » Provide pros and cons
- The single presenters can serve as **indices** which help the user to classify information.
 - » Provide information from different points of view, e.g. businessman versus tourist
- Presentation teams can serve as rhetorical devices that allow for a continuous **reinforcement of beliefs**
 - » involve pseudo-experts to increase evidence

Multiple Presentation Teams



Multiple Presentation Teams: Automatically Generated Variants

Select the agents and their personality:



Genie



Merlin



Peedy



Robby

SELLER		BUYER1		BUYER2	
Genie ▼		Peedy ▼		Merlin ▼	
Agreeableness	Extraversion	Agreeableness	Extraversion	Agreeableness	Extraversion
<input checked="" type="radio"/> agreeable	<input type="radio"/> extravert	<input type="radio"/> agreeable	<input type="radio"/> extravert	<input checked="" type="radio"/> agreeable	<input checked="" type="radio"/> extravert
<input type="radio"/> neutral	<input type="radio"/> neutral	<input type="radio"/> neutral	<input type="radio"/> neutral	<input type="radio"/> neutral	<input type="radio"/> neutral
<input type="radio"/> disagreeable	<input checked="" type="radio"/> introvert	<input checked="" type="radio"/> disagreeable	<input checked="" type="radio"/> introvert	<input type="radio"/> disagreeable	<input type="radio"/> introvert

Multiple Presentation Teams: Automatically Generated Variants

- **Buyer:**

- » Genie: positive, extrovert
- » Robby: positive, introvert

- **Seller:**

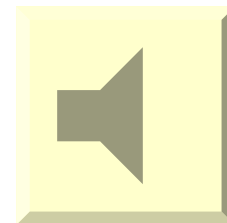
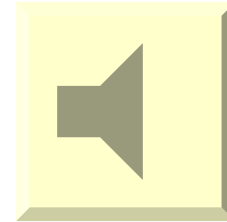
- » Merlin: positive, extrovert

- **Buyer:**

- » Genie: negative, extrovert
- » Robby: negative, extrovert

- **Seller:**

- » Merlin: negative, extrovert



Basic Idea

- **System as a screen writer** who determines all dialogue contributions of the involved agents based on a given presentation goal
- **Presentation goal:**
GOALS :
 PERFORM create_script;
- **Model authoring knowledge** by plan operators
 - » **Complex subgoals:** Multimodal dialogue acts
 PERFORM InformValue \$agent \$object
 \$attribute;
 PERFORM play_animation \$agent \$gesture;
 - » **Primitive subgoals:** Generation of HTML-Code
 EXECUTE JamFileWriter.includeHTML \$text;
- **Result of the planning process:** HTML-Code which contains instructions for the MS agents

Agent Modell

- **Personality:** Extroversion, Agreeableness, Neurotism
FACT agreeableness "Merlin" "agreeable";
- **Emotions:** Type and Intensity
FACT emotion "Merlin" happy high
- **Status**
FACT status "Merlin" "hi_status";
- **Role:** Buyer, Seller
FACT seller "Merlin";
- **Interest/Expertise:** Environment, Sport, Technology ...
FACT interest "sportiness" "Merlin";
- **Attitude:** Negative, Neutral, Positive
FACT attitude "Merlin" "negative"

Example of a Dialogue Strategy

Question:

How much gas does it consume?

Answer:

It consumes 8l per 100 km.

Negative Response:

I'm worrying about the running costs.

Dampening Counter:

Forget about the costs.

Think of the prestige!

Header:

(dampening_counter ?agent ?prop
?dim)

Constraints:

(*and*
(positive ?agent)
(pol ?prop ?other_dim positive))

Inferiors:

(Speak ?agent
("Forget about the " ?dim "!"))
(Speak ?agent
("Think of the " ?other_dim "!"))

Further Application of Presentation Teams

**Matze
&
Gerd**

Commentary of



Soccer Games



Summary: Non-Interactive Version

- Approach supports the generation of coherent dialogues
- Adding new agents is combined with some effort, i.e. new plan operators need to be defined
- No interaction possibilities at runtime

Interactive Presentation Teams



Characteristics of the Interactive Presentation Scenario II

- **Open Architecture**
 - » New agents can join at any time.
- **Open-ended Scenario**
 - » There is no pre-defined end.
- **Auto-Progression**
 - » Story unfolds no matter whether the user actively participates or not.

Characteristics of the Interactive Presentation Scenario II

- **Handling of Barge-ins**
 - » Agents may interrupt each other at any time.
- **Computer-Moderated Dialogue**
 - » Meta-agent makes sure that all agents follow an agreed-upon interaction protocol.

Exploitation of an improvisational framework

- An improvisational frame can resolve the conflict between predestination and freedom of interaction on an operational level.
- View an improvisational frame as a collection of contextual constraints for:
 - » character behavior and
 - » structuring the interaction between autonomous characters

Improvisational Framework for a Buyer

```
FACT language "peedy" "en";
```

```
FACT personality "Peedy" "extraversion" "extravert";
```

```
FACT personality "Peedy" "agreeableness" "disagreeable";
```

```
FACT personality "Peedy" "neurotism" "balanced";
```

```
FACT status "Peedy" "hi_status";
```

```
FACT attitude "peedy" "positive";
```

```
FACT role "peedy" "buyer";
```

```
FACT interest "sportiness" "Merlin";
```

New: Each Agent has its own goals

```
PERFORM initialize_agent "peedy";
```

```
MAINTAIN goal "good_behaviour" "greet_back" "peedy";
```

```
ACHIEVE dialogue "buy_car" "peedy";
```

Characteristics of the Interactive Presentation Scenario

- **Character-Centered Approach**

- » Story is not defined by a script, but by the character's role, personality, status, attitude etc.

The screenshot displays the MIAU Menu application interface, which is divided into two main windows: "MIAU Menu" and "Menu".

MIAU Menu Window:

- Agents:** A text field contains the name "james".
- Role:** Three radio buttons are present: ☒ Buyer, ☐ Seller, and ☐ MetaAgent.
- Attitude:** Three radio buttons are present: ☐ Positive, ☐ Neutral, and ☒ Negative.
- Initial Status:** Three radio buttons are present: ☒ High, ☐ Middle, and ☐ Low.
- Personality:** A table with three columns: Positive, Neutral, and Negative. The rows are Extraversion, Agreeableness, and Neurotism.

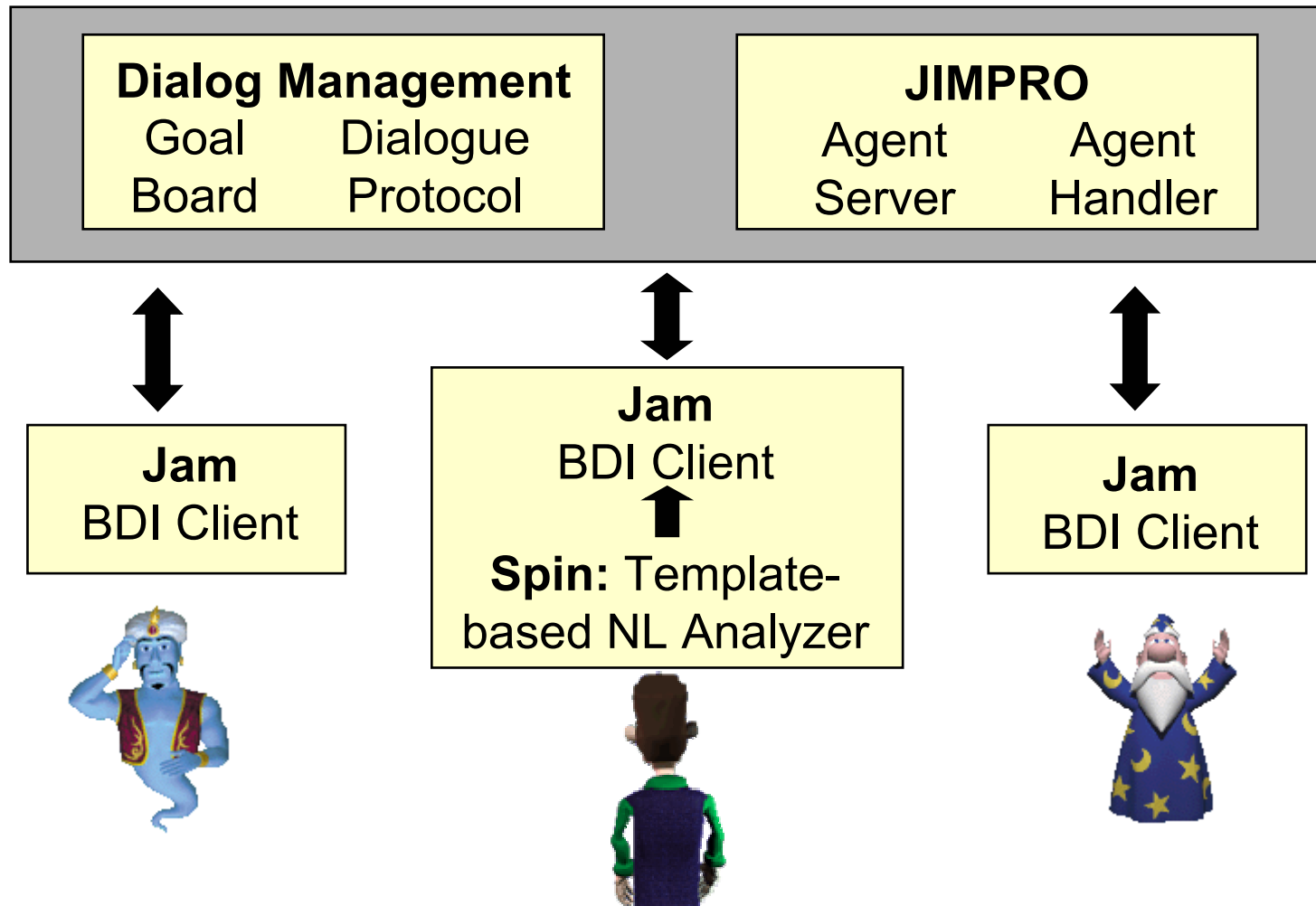
	Positive	Neutral	Negative
Extraversion	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Agreeableness	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Neurotism	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
- Buttons:** "Start", "Advanced", and "Close" are located below the personality table. At the bottom of the window are two large buttons: "Stop Agent Server" (green) and "Start User Shell" (red).

Menu Window:

- Table:** A table with three columns: a category, "Yes", and "No". The categories listed are sportiness, prestige, comfort, environment, security, costs, running costs, and family.

	Yes	No
sportiness	<input type="radio"/>	<input checked="" type="radio"/>
prestige	<input checked="" type="radio"/>	<input type="radio"/>
comfort	<input checked="" type="radio"/>	<input type="radio"/>
environment	<input type="radio"/>	<input checked="" type="radio"/>
security	<input type="radio"/>	<input checked="" type="radio"/>
costs	<input type="radio"/>	<input checked="" type="radio"/>
running costs	<input type="radio"/>	<input checked="" type="radio"/>
family	<input type="radio"/>	<input checked="" type="radio"/>
- Buttons:** A "Back" button is located below the table. At the bottom of the window are two large buttons: "Start Agent Server" (red) and "Start User Shell" (red).

System Architecture for Miau Multi-Party Dialogue Scenario



■ Synchronizing Dialogue Contributions in the Distributed System

- **Who gets the turn?**
 - » Competence
 - » Personality, Emotion
 - » Status, Role
 - » Context
 - Who was addressed?
 - Topic shift ok?
- **Explicitly Addressing Dialogue Partners by:**
 - » Name
 - » Viewing angle
 - » Gestures

Multi-Agent Dialogue Control



Summary: Interactive Presentation Teams

- Approach offers **high flexibility**
 - » Agents may join and leave at any time
 - » User may show up in different roles
 - » User has the option of actively participating, but is not forced to do so.
- **Problem:** How to get interesting interactions?

From Script-Based Approaches to Interactive Performances

	<i>Miau (1st Prototype)</i>	<i>Miau (2nd Prototype)</i>
Metaphor	scripted theatre	improvisational theatre
Scripting Time	prior to presentation, offline	during presentation online
Structuring Principle	plot-centered	character-centered dramatic elements
Script Producer	separate system component	involved characters and user
Technical Realization	centralized planning component	distributed reactive planners