

Multimodal Research at CPK, Aalborg

Summary:

The IntelliMedia WorkBench (“Chameleon”)

- Campus Information System
- Multimodal Pool Trainer
 - Displays, Dialogue Walkthru
- Speech Understanding
- Vision Processing

Other (student) projects

New projects: Multimodality in Wireless Networks



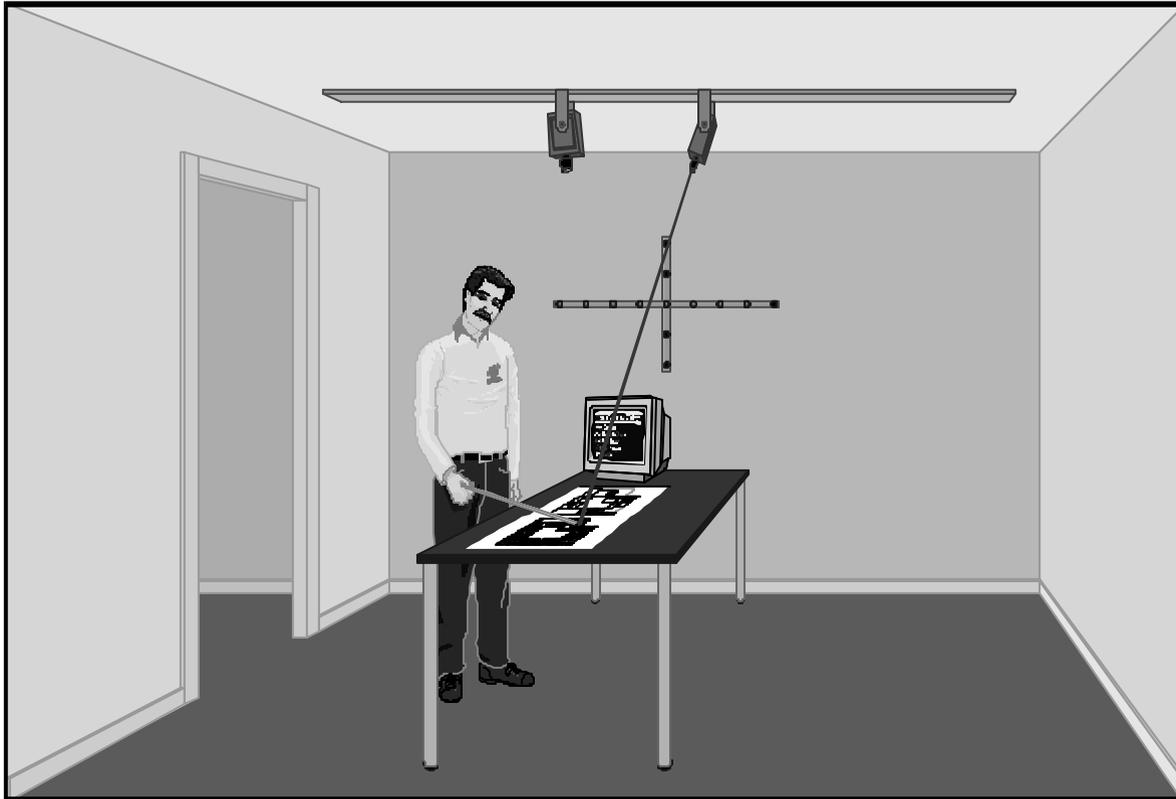
The IntelliMedia Workbench (“Chameleon”)

- A suite of modules for vision and speech processing, dialogue management, laser pointing, blackboard etc.
- Purpose:
 - Cross-disciplinary collaboration at Aalborg University.
 - Exploring cross-media fusing techniques
 - Exploring multimodal human-machine interaction



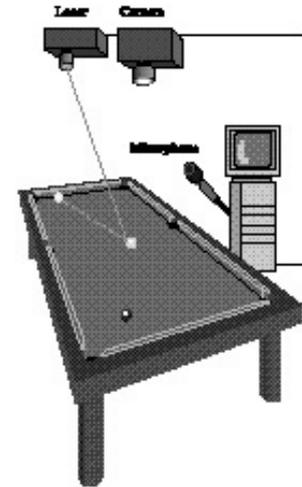
Workbench Application 1

A Campus Information System



Workbench Application 2

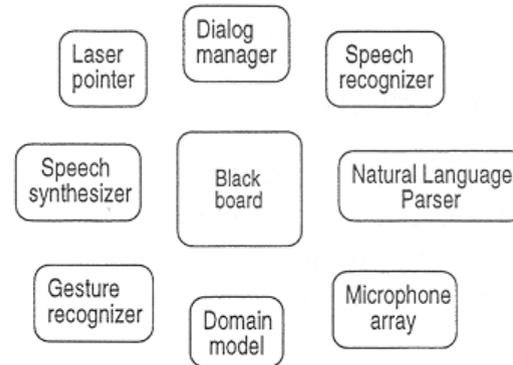
Multimodal Pool Trainer



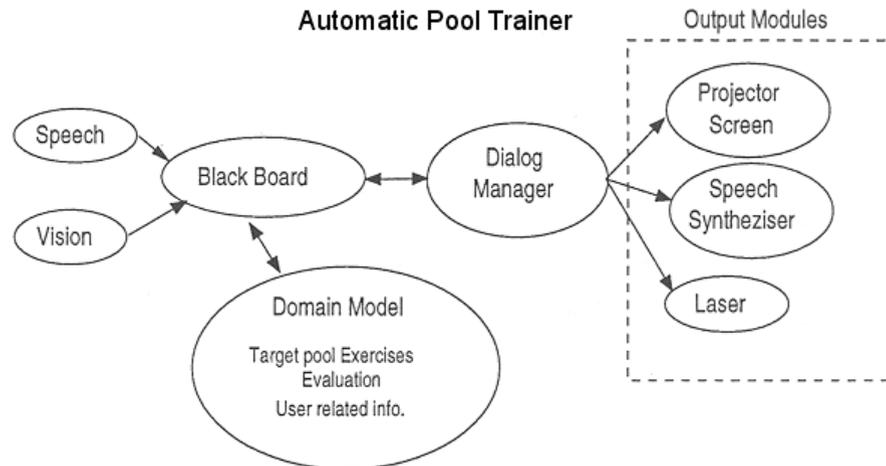
Architecture

- Initially designed WorkBench architecture (as used in The Campus Information system)
- and as used in the Pool Trainer

IntelliMedia WorkBench



Automatic Pool Trainer



The Game of Pool

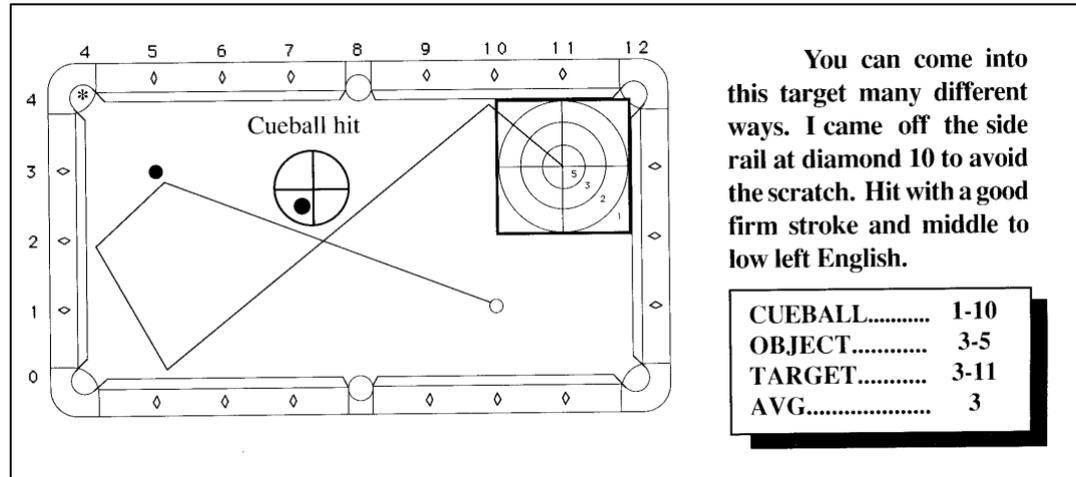
Pool is a game that requires a combination of strategic thinking as well as physical skills. Without one, the other is not of much use.

Basically, the most important requirement for any pool player is the ability to shoot the target balls into the pocket, while ensuring a good position of the cue ball for the next shot.



Target Pool

The automatic Pool Trainer is based on the widely used Target Pool system, developed by the professional pool player Kim Davenport.



Example of a typical Target Pool exercise



The computer Vision subsystem

The main functions of the image analysis subsystem are:

- Calibration and detection of the positions of the empty pool table, i.e. the rails, diamonds and pockets.
- Detection of still and moving balls placed on the pool table.
- Detection of when the cue ball is hit.
- Recording of the shot



The computer Vision subsystem

All image analysis is carried out on binary *difference images*. This greatly reduces the time and space requirements for the image processing

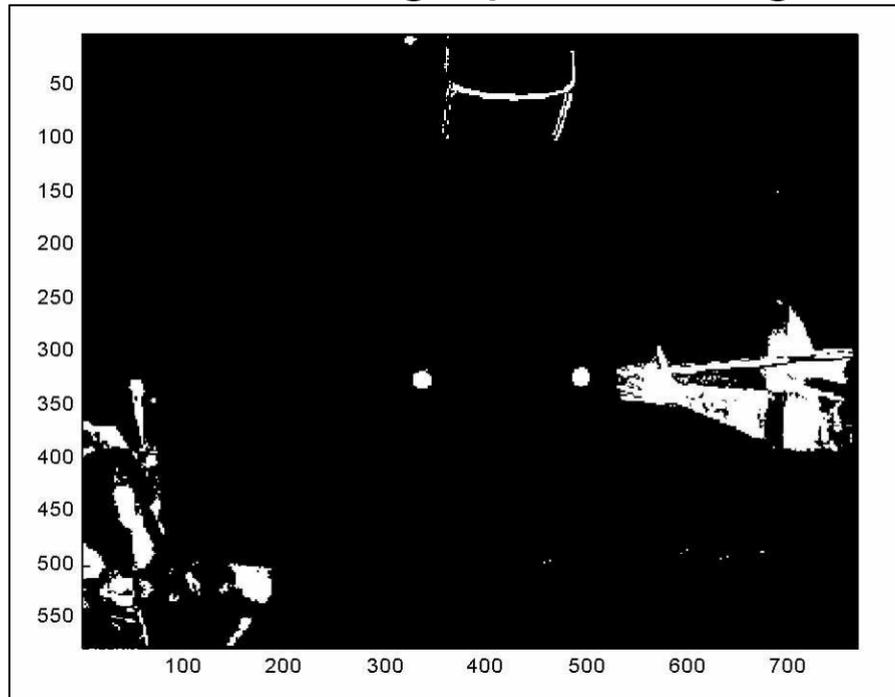
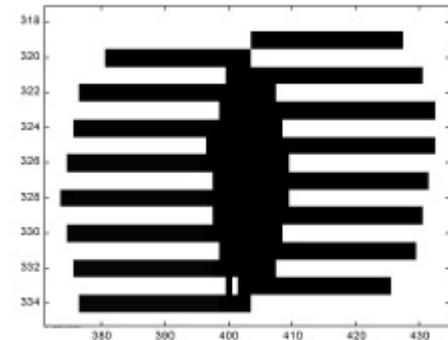
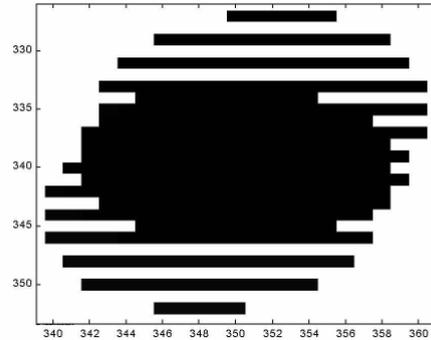
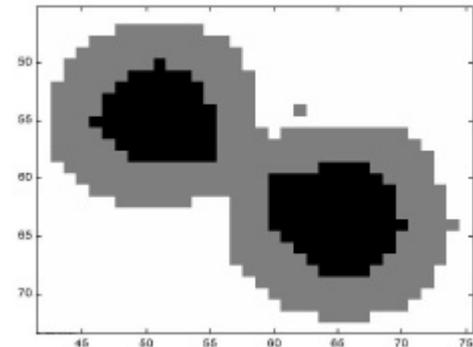
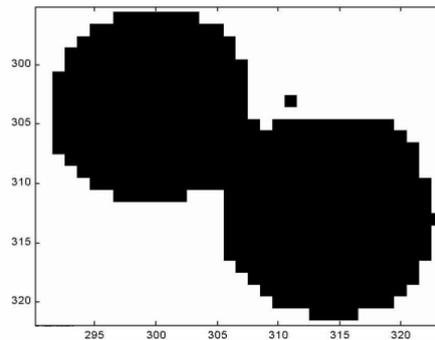


Image Processing

Detection of still and moving balls benefits from the distinctive patterns created by the CCD chip line scan effect.

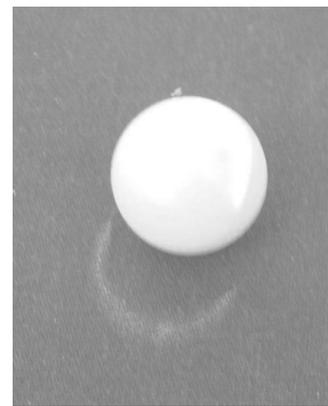
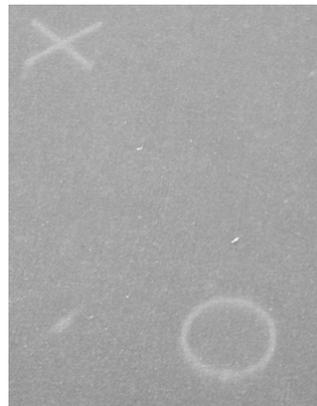
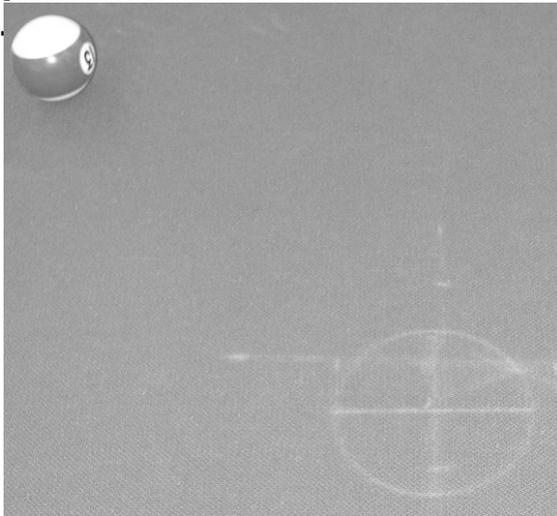
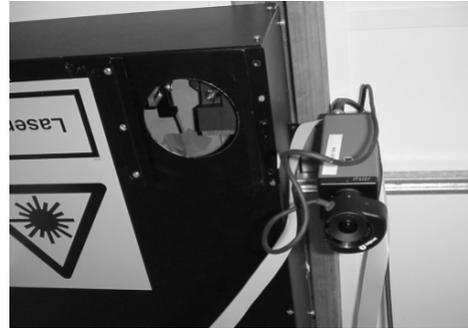


Close-lying balls are detected by removing edge pixels



The Laser Sub-system

The laser is placed above the pool table and is used to draw the target and optimal paths of the cue- and



Mark the positions where the user must place the balls



The Speech Sub System

A number of speech recognition engines have been used in the development of the system.

SR is presently carried out by the IBM ViaVoice recogniser. Previously, Entropics GraphVite/HAPI recognition engine have been used.

We are currently extending the interface (JSAPI) to include the public domain hvite recognition engine from Cambridge University. This will in turn allow us to support a larger number of languages, e.g. through the COST 249 Task Force reference recogniser initiative.



The Speech Sub-system

The CPK Natural language Processing Suite is presently being integrated into the trainer.

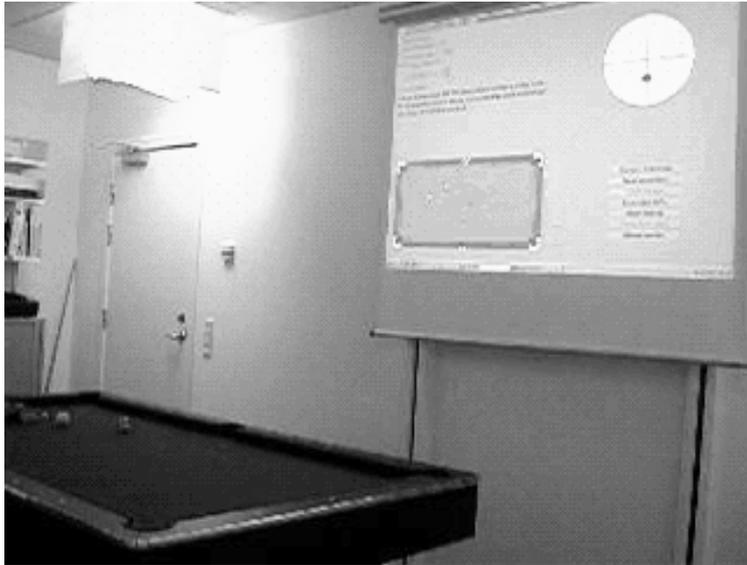
- Apart from enabling a compound feature-based language model, the suite supports a number of popular SR grammar formats, such as htk and jsgf.

Synthetic speech output is used to achieve the high degree of flexibility needed in the spoken output

- IBMs ViaVoice and the Infovox speech synthesisers have been used, but any SAPI compliant synthesiser is supported
- Speech output is synchronized with the laser, graphics and text output to form an integrated output to the user



Examples



An example of a user interacting with the system

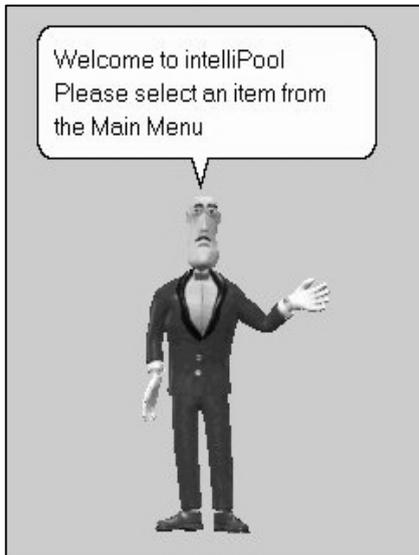


An example as seen by the system's camera



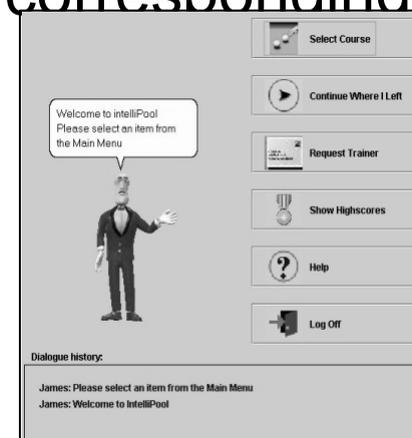
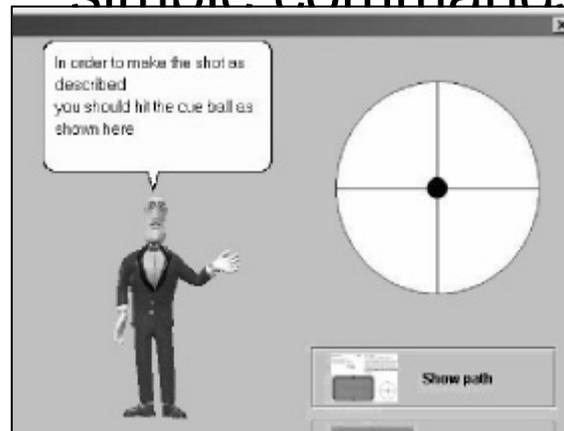


The Display Sub-system



- To issue commands and receive instructions, the user communicates by speech via the interface agent James
- James is animated and understands simple commands corresponding to

He instructs the user by speaking, pointing and moving around on the screen.



Example of the interaction during an exercise

The system is activated; Q takes the initiative:

Q Welcome to SMARTPOOL. Tell me your name.

Svend Svend.

[SMARTPOOL looks up Svend and checks if he is known. Svend is known].

Q Hi Svend.

Q Do you want to continue where you left last time?

Svend Yes.

Q That was Course 2, Exercise 3.

Screen [The exercise is shown on the projector screen. It consists of the layout of the pool table (positions and route of balls), a closeup of the location to hit the que ball, and a verbal instruction]

Q (reads the verbal description aloud)



Example of the interaction during an exercise

Laser [The position of the que ball is indicated with a circle on the table and the target ball with a cross]

Screen The same is shown on the table drawn on the screen.

Svend [Places the balls on the table, but is not careful and does not place it right]

Screen [SMARTPOOL is checking the position of the ball when no more activity can be detected on the table. A ball in the wrong position is shown as red. When a pool ball is placed correctly, it turns from red to white/yellow on the PS].

Screen [When all balls are in place, the path of the cue ball, the pocket for the target ball, and the target are drawn on the table shown on the screen]

Laser [The target is drawn on the table]

Svend [Shoots the target ball in the pocket and manages to get the que ball fairly close to the target drawn on the table]



Example of the interaction during an exercise

Q Nice Svend, you got 2 points.

Screen[The score is shown on the screen. The status automatically returns to the setup of the exercise]

Laser [The laser switches back from showing the target to the balls initial position]

Svend [Pauses]

Q Do you want to see your stroke?

Svend Yes please.

Screen[The path of the shot together with the original path are shown in different colours.]

Q Do you want to see a replay of your stroke?

Svend Yes please.

Screen[A movie is compiled from the images captured by the camera and is shown on the screen.]

Q Would you like to repeat the exercise or go on to a new?

Svend No thank you.



Comments to the Dialogue

The spoken dialogue can be carried out using the touch screen instead

Dialogue is most intensive during setup and evaluation of the exercise.

Although the example does not illustrate this, the user can take the initiative at almost any point.

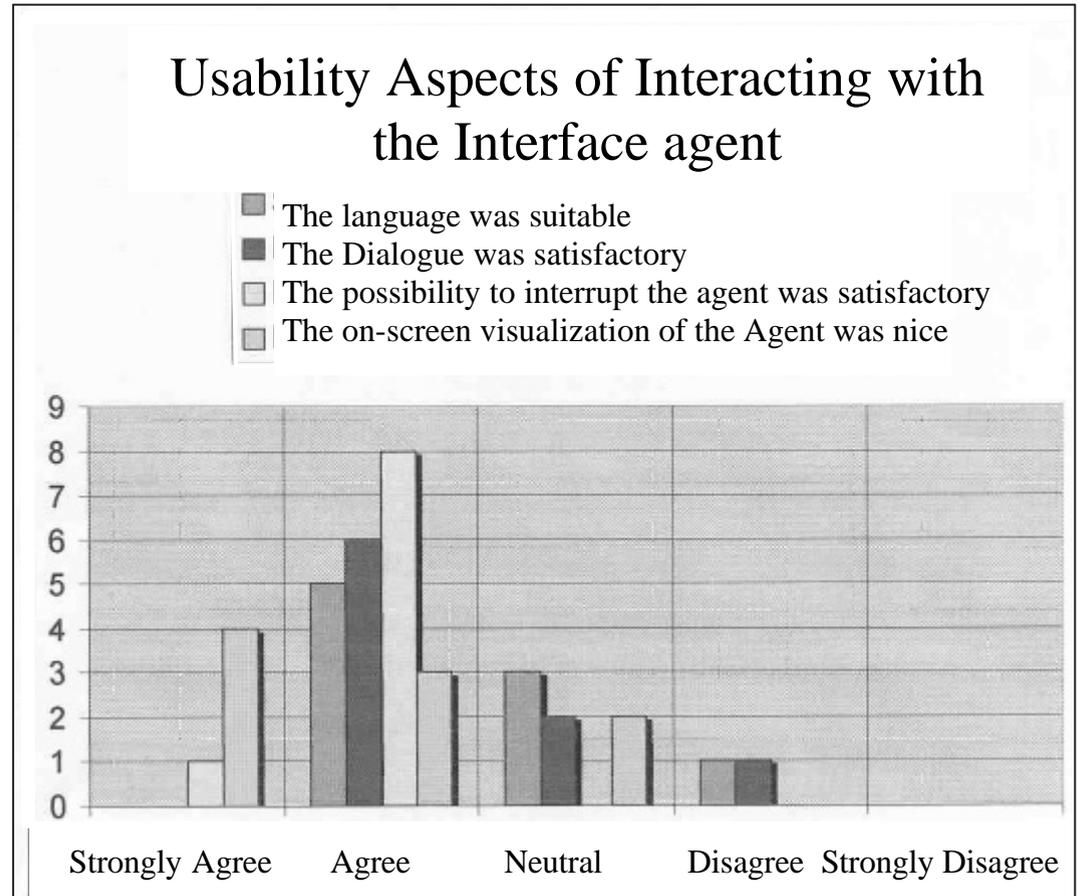
An extensive help function (both about playing pool, the exercises and the system) are available

During the exercise the interaction is almost exclusively non-verbal, via physical interaction with the pool table and display on the wall-screen



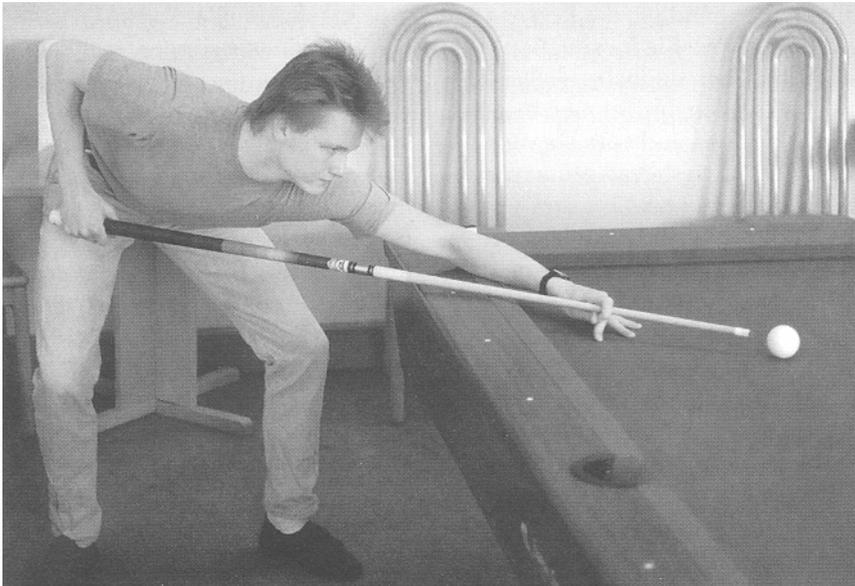
Users Tests

All users were asked to fill out a questionnaire after performing the test

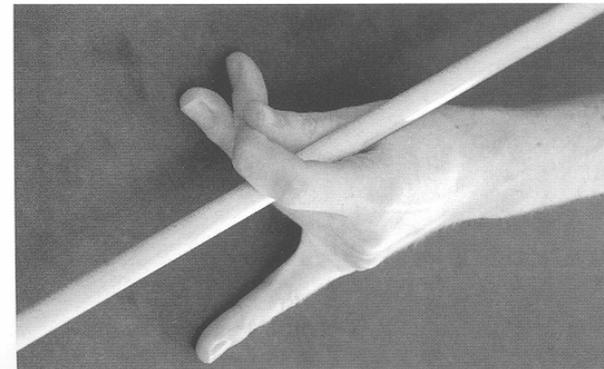
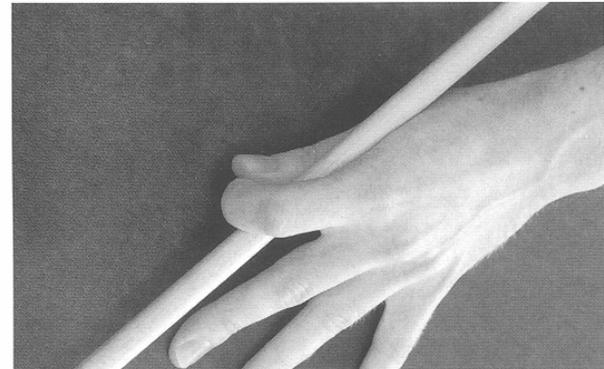


Users Tests

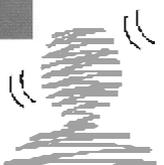
However, the participating pool instructors pointed out a number of issues not addressed by the system, e.g:



Stance



Bridges



Discussion

Overall, the Pool Trainer has been successful. However, some improvements are needed:

- The image analysis subsystem, although performing fast and accurate needs to be made more robust against changes in e.g. the lighting conditions, if the system were to be placed in a non-controlled environment
- If a detailed feedback of the user errors is needed, it will require knowledge about the direction and speed of the balls



Other (student) projects

Affective computing, classification of emotional speech

Recognition of hummed tunes

Enhancing Lego Mind Storm with vision

GPS-systems using touristic (non-true scale) maps

White Board application using gesture recognition



Multimodality in Wireless Networks

- Handheld client - Remote server
- distribution: what is executing where, what is transmitted?
- Selection of modality
 - based on information type (e.g. speech is temporal, don't use it for time tables!)
 - based on situation (e.g. speech enables “eyes-free”/“hands-free” operation)
 - based on network conditions
 - is your modality (what you transmit) sensitive to package loss?
 - Is your modality sensitive sensitive delays
 - does your modality require a bandwidth

