# The Automated Pool Trainer - a Multi Modal System for Learning the Game of Pool

Lars Bo Larsen, Peter M. Jensen, Kenneth Kammersgaard, Lars Kromann
Center for PersonKommunikation
Aalborg University, Aalborg, Denmark
Email: *lbl@cpk.auc.dk, [pmj, kek, lkr]@mindpass.com*

**Abstract:**

*This paper describes the Automated Pool Trainer (APT), a multi modal pool training system developed at Aalborg University. The philosophy of the system is to automate the learning process, in this case learning the game of pool. The APT system is based upon the Target Pool methodology developed by pool player Kim Davenport. It is a multi modal system, utilizing spoken interaction combined with a graphical output and a computer controlled laser pointer for user communication. The trainee selects a suitable exercise among a number of predefined courses (aided by the system). The system issues instructions on how to place the balls on the table, shows the (optimal) shot, and records and evaluates the performance of the player. The instructions are given orally (speech synthesis) combined with gestures (using a laser as a virtual "pointer" on the pool table). A resume is shown on a projected display, together with e.g. a close up of how to hit the cue ball. The system also includes a speech recognizer for interpreting user utterances. A camera is mounted above the pool table for checking the positions of the ball and recording the shot, which can be played back for the user. The shot is evaluated and scored. The system keeps a user profile for all users.*

**Keywords:**

**HCI, multi modality, image analysis, computer vision, speech recognition, speech synthesis, spoken dialogue systems, pool, usability.**

## 1 Introduction

This paper describes an Automated Pool Trainer (APT), a multi modal pool training system, as shown in Figure 1 below. It is based on a previous project carried out at the Center for PersonKommunikation at Aalborg University, where a number of hardware and software modules were integrated into an open architecture to provide the "IntelliMedia WorkBench [Brøndsted 1998]". The intention of the workbench is to provide a basic setup, in which new applications, such as the present one can be built.

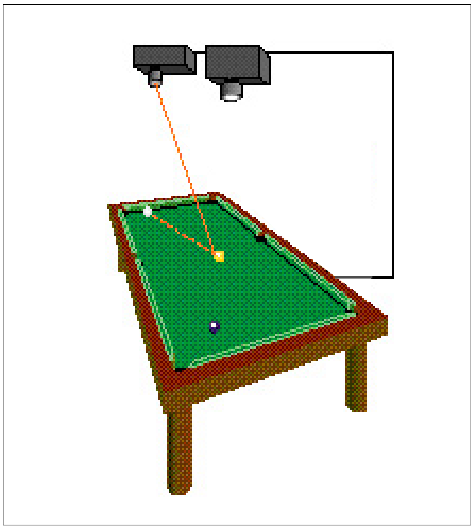

Figure 1. The Automatic Pool Training System uses a camera and a laser to interact directly with the pool table.

The Automatic Pool Training system is based on the widely used Target Pool [Davenport 1992], developed by the professional pool player Kim Davenport. The following sections present the original Target Pool and an overall description of the APT. This is followed by a closer account of the system architecture and the individual components in greater detail. The design approach is discussed and the results of a user test are presented. The system is reviewed with regard to similar systems and technologies.

## 2 Target Pool

The basic idea of Target Pool is simple: To present the trainee with a number of exercises, and record and evaluate his or hers performance after each shot. Based on the performance a new exercise (or the same again) is suggested. In addition to this, a golf-like handicap is

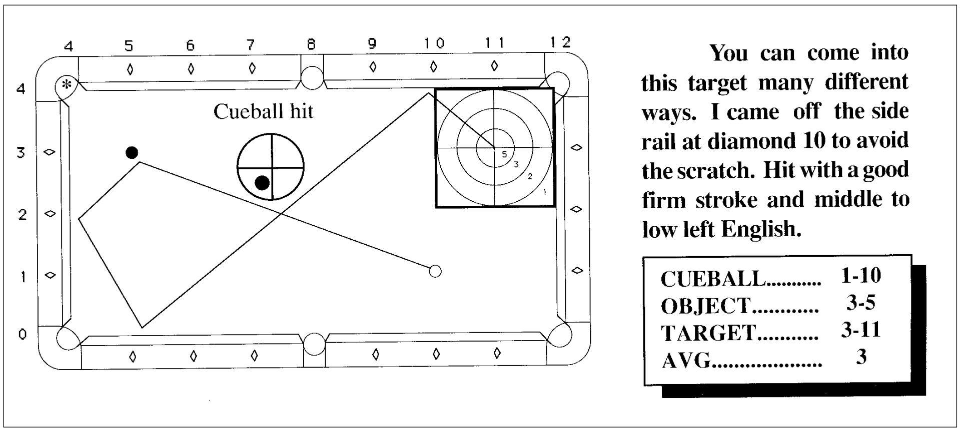


Figure 2. Example of a Target Pool exercise, as shown in [Davenport 92, p.38]

introduced, allowing users at different levels to compete. It can also be used to track users' progress over time. The only equipment needed apart from the pool table and queues, etc., are a booklet describing the exercises, a score-board, and a thin cloth with a printed target, to be placed on the pool table. Usually only the cue ball and one other ball (the object ball) are used.

The objective of Target Pool is to sink the object ball in the specified pocket and cause the cue ball to come to rest on the target. There are several courses to play in Target Pool – from beginner to expert. Each course consists of ten exercises. The cue ball, target and the object ball (if needed for the exercise) are placed on the table at the co-ordinates given in each shot description. Figure 7 above shows an example of an exercise description for Target Pool.

Each shot layout shows the position of the cue ball, object ball and target. There is also a solid line showing the suggested path of the cue ball.

A diagram labelled; "Cue ball hit" is shown for each exercise. This is the cue ball seen from the shooters position. The black spot on the cue ball is where the cue should contact/touch the cue ball. A star in a pocket indicates that this is the pocket where the object ball should drop into.

A short explanation of the shot is shown to the right of the pool table diagram. Below the explanation is a box showing the co-ordinates where the cue ball, object ball and target are to be positioned. Also in the box is the average score (Kim Davenport's average) for each shot. An example of such a shot layout is shown in Figure 2.

## 3 The Automated Pool Trainer

The Automated Pool Trainer is basically an automatic system implementing the Target Pool training scheme. The philosophy of the system is to automate the teaching process, in learning the game of pool. To make the learning process optimal, theories about learning and HCI have been employed during the design. The learning process has two aspects: Learning the game of Pool and learning how to use the system. Both learning aspects have been considered in the development of the system.

The core of the automated system consists of a PC connected to various input and output devices. The input devices are a CCD camera, placed above the pool table, and a microphone used for spoken input. The output-devices are a laser, placed alongside the camera, a loudspeaker and a projector. This configuration is shown in Figure 8 below.

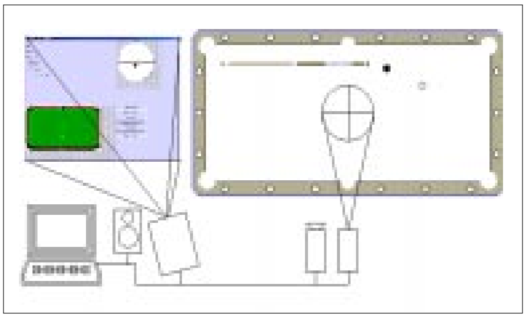

Figure 3. The APT system. The system consists of a pool table, a projector showing the scoreboard, exercises, pool table, etc, a camera, a microphone and a speaker and a laser. [From Jensen, p. 32]

The APT system implements the following general functionality:

- Identification of the positions of the pool table (extent of the baize and the pockets) and the balls currently placed on the table.
- Detection of moving balls.
- A laser beam is used to illustrate the correct layout of the balls and the target for a given exercise. This is achieved by drawing points and lines directly on the surface of the pool table. Furthermore, the laser beam is also used to show the actual path taken by the ball for feedback purposes.
- Instructions are given by synthetic voice, combined with illustrations and a written resume projected onto a large graphical display mounted on the wall.
- Feedback is provided by synthetic voice and graphics (score, replay of shot)
- User commands can be issued by voice or a keyboard + mouse interface.
- An interface agent is used for focusing the user – system communication in one version of the system.

## 4 APT Architecture and Subsystems

As described in the preceding section, the system is built around a powerful PC equipped with a number of devices and software modules. In the following sections, each subsystem is described in some detail, although the reader is referred to [Bondesen 1999] and [Jensen 2000] for a full description. Most attention is given to the image analysis subsystem, as the laser and speech components are commercial products, whereas the image analysis subsystem has been developed solely for the APT.

### 4.1 The Image Analysis Subsystem

The image analysis subsystem handles the task of detecting the balls on the pool table. This includes both still balls and moving balls. When the user is to place the balls for an exercise, the system has to detect when the balls are placed correctly. When the user performs the shot, the image analysis must detect the ball movement. This makes it possible for the system to evaluate the shot and give the user feedback of how well he or she performed the shot. Consequently, the three main functions of the image analysis subsystem are:

- Detection of still balls placed on the pool table.
- Detection of the ball movement during a shot.
- Detection of when the cue ball is hit.

A number of steps must be taken to implement this. Firstly, the system must be calibrated, and the images transformed and filtered to eliminate radial distortion in the lens [Gonzalez 1992] and obtain a suitable representation, from which the balls' position and movements can be extracted.

All image analysis is performed on *difference images*. These are obtained by generating a *reference image* of the empty pool table. By averaging over 25 images noise pixels can be eliminated. Subtracting the reference image from any given image will make any changes (as e.g. the balls) stand out clearly (see Figure 4 below).

Furthermore, the difference image is converted into a binary image by applying a threshold.

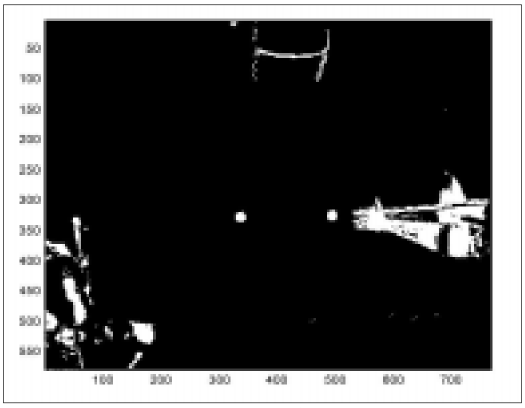


Figure 4. An example of a binary image made by subtracting a reference image from the present image, and applying a threshold to the resulting difference image. The two balls placed on the pool table are easy to identify. (from [Jensen, p.75])

Identification of still balls is rather straight forward, even when two balls are touching. This case is illustrated in Figure 5 below. The edge pixels are removed to check if the object contains two separate balls.

When a ball is moving a characteristic pattern emerges, due to the interlaced line scan of the CCD chip [Gonzales 1992]. This is shown in Figure 6 below.

Note the characteristic pattern created by the CCD line scan. A closer analysis of this pattern can be used to determine the speed and direction of the balls, given that the line-scan frequency and the size of the ball are known.

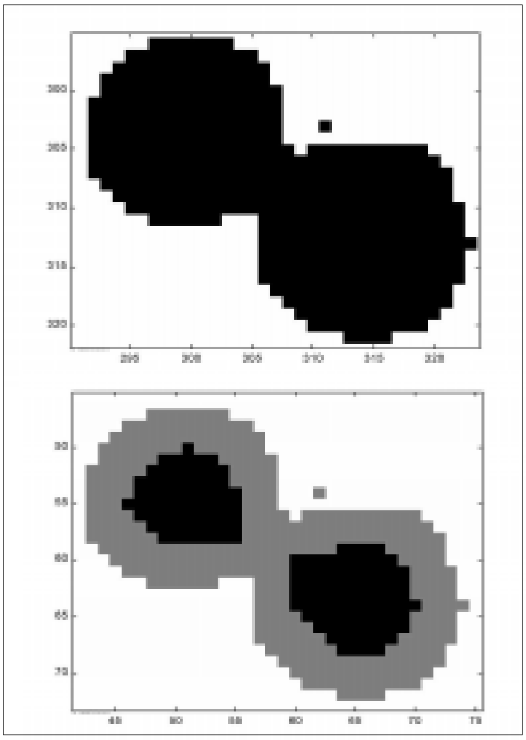

Figure 5. A) Illustrations of a situation where two balls are placed close together, resulting in a detection of one object. B) Resulting object after removal of the edge pixels after 3 iterations. The removed pixels are shown as grey. From [Jensen 2000]

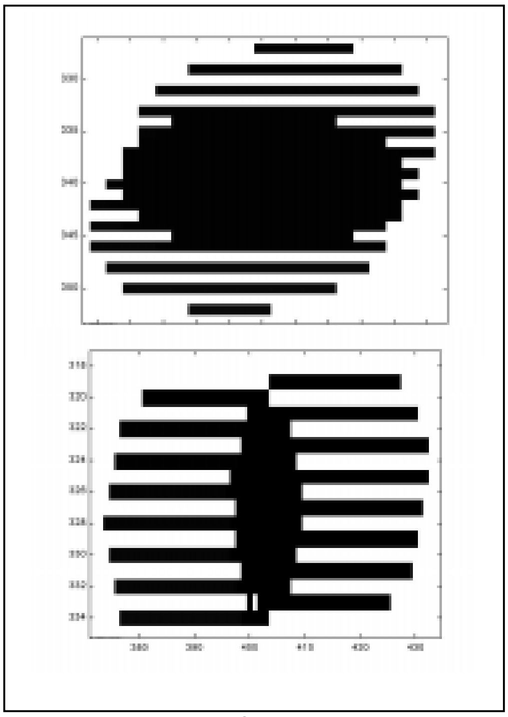

Figure 6. Images of a ball moving. A) Horizontally, and B) vertically.

### 4.2 The laser subsystem

The laser light is generated by a red laser 15mW diode. The laser beam is moved via an X-Y scanner, controlled by a separate PC. It is capable of redrawing a path through 600 points at a rate of 50 Hz. The result is a near flicker-free drawing on the surface of the Pool Table. However, the longer the path, the dimmer the laser path is.

One problem is that a red laser on the green baize of the Pool Table actually is the worst possible combination. This is caused by the green surface absorbing the red light. The problem could be solved using a blue-green laser, but the price will be a factor of ten higher. [Brøndsted 1998], [Lausen 1997].

### 4.3 The speech subsystem

Two speech systems have been used. The present is based on the IBM ViaVoice speech recognition and –synthesis Engine. It is mainly intended for dictation, but can also be used with user-defined grammars [IBM 2000]. It can easily be integrated into applications using the JSAPI interface [SUN 2001]. However, another more elaborate spoken dialogue sub system has been used until recently, and is described in more detail here.

The speech recognizer used is the HVite, from Entropic [Odell 1997]. The Infovox speech synthesizer [INFOVOX 1994] provides spoken output. See [Bondesen 1999] for a detailed account.

The spoken dialogue design closely reflects the task at hand, the intended end-users, and the context of the pool table and the graphical display.

The vocabulary and grammar must be limited to those typical of the domain in order to reduce the complexity. In the present case these are technical terms used in the pool game, and common words used for e.g. commands, questions asking, etc. An important source for the design of the vocabulary was recommendations from pool instructors obtained from a number of interviews. In total, the vocabulary comprises 100 words [Bondesen 1999, p.73-74].

Although the vocabulary seems quite limited, the user is allowed considerable freedom of expression. The recognizer will spot for keywords and –phrases, and ignore in-between speech carrying no semantic information.

The dialogue is designed to handle both novice and experienced users. This is achieved by implementing a strategy of *mixed initiative* [Larsen 1997a,b]. Mixed-Initiative dialogues have the advantage that, for novice users, the system can retain the control of the interaction throughout the interaction, thus providing guidance. For experienced users, the system allows the user to take control of the interaction, e.g. by issuing a command instead of answering a system question. This technique provides a much more flexible and user-friendly dialogue, but at the possible cost of a higher complexity

Instructions and responses are generated by a combination of synthesized speech, laser gestures and graphics. These are carefully synchronized to provide an integrated response. Because of the non-persistency of speech, a resume of the latest spoken instructions is shown on the graphical display. The result of the speech recognition is also shown for feedback purposes. This reduces the risk of a misunderstanding, and allows the user to quickly identify and recover from errors. To allow the user freedom of movement, a wireless microphone headset is used.

### 4.4 The graphical display subsystem

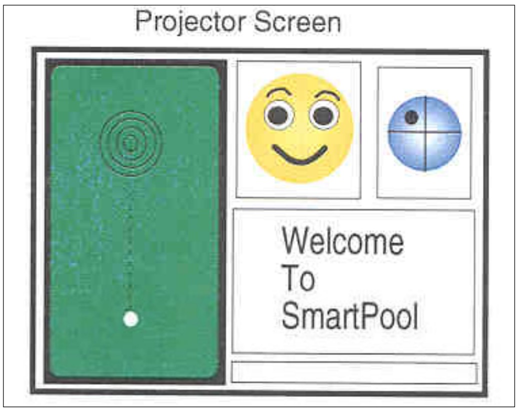


Figure 7. Display layout with interface agent

Figure 7 above and Figure 8 below show two different layouts of the graphical display. The display shown in Figure 7 utilizes an interface agent. This serves several purposes:

- As a focal point for the spoken interaction with the user, by e.g. animation of lip movements;
- As feedback, using facial expressions (happy, sad, etc.) to indicate the evaluation of a user's shot;
- As a means to focus the users' attention by letting the agents' eyes look in the desired direction.

On Figure 8 an alternative layout is shown. Like the previous layout, it displays the pool table and a close-up of the queue-ball, with an indication of where to aim the stroke. The interface agent has been replaced with a more traditional menu interface. This can be accessed either by voice commands or with a mouse. Furthermore, this display provides the ability to show a playback of the users' shots.

Figure 8. Alternative Display layout

## 5 Evaluation and Discussion

The system has been evaluated in a usability test with nine participants, including novice and experienced pool players, and two pool instructors [Bondesen 1999]. Each participant carried out three scenarios, where the first was a trial run to familiarize the user with the operation of the system. The test was carried out with a simulated speech recognizer due to technical problems. However, the users were not told of this, and were equipped with a microphone anyway. After the test, the users were debriefed and each filled out a Likert questionnaire.

A number of observations were made:

- Most favored the basic idea of pre-defined exercises;
- All found that the evaluation/feedback in the form of a score were good, but they were in disagreement of whether further evaluation was necessary;
- All found the combination of the audio and visual modalities favorable, and none were confused about the combined display and pool table interaction;
- The quality of the voice was too poor.

Figure 9 below shows the users' evaluation of the interface agent, which is generally positive.

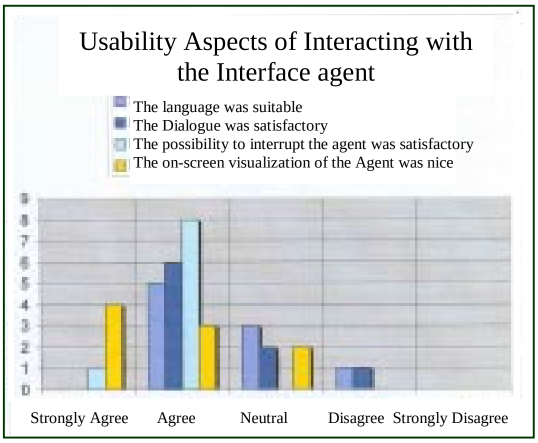


Figure 9. User reactions to the Interface agent [Bondesen 1999]

A number of observations were made during the test:

- The users' missed a demonstration Video
- Delays were irritating
- The test users looked at the interface agent when answering and issuing commands
- In general, the users had a positive attitude towards the system.

In Particular, the pool instructors noted that:

- Found the system a bit primitive at first, but changed to a more positive attitude during the test, and ended up by accepting the idea of an automatic trainer;
- The system will never replace a human instructor, but;
- Could be a good as a supplementary tool

## 6 Conclusions and Future Work

As the preceding sections have shown, the design and implementation of the Automatic Pool Training system has been successful. Experiments with users have been carried out with different versions of the system, and have proven the ideas valid. However, a number of problems have been identified. Most notably are:

- The speech modules did not work sufficiently well, and need further improvements;
- 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.

As mentioned in the Introduction, the APT is built utilizing modules and ideas from the Chameleon Project, described in [Brøndsted 1998]. It employed a blackboard architecture for the integration of the different modalities, which might also be advantageous in the present case.

The Stochasticks project [Jebara 1997] at the MIT Media lab also addresses an automatic pool training system. Here, the user is required to wear goggles, where a stream of live images (taken by head-mounted cameras) is superimposed with shots suggested by the system. Although very advanced, the system was clearly designed to demonstrate wearable computers and augmented reality, rather than creating an educational system.

## 7 Acknowledgements

The authors wish to thank Pernille Bondesen, Søren Poulsen and Morten Lykkegaard for their initial work on the Smart Pool trainer.

## 8 References

[Bondesen 1999] Bondesen, P., Poulsen, P., Lykkegaard M., "SmartPool – A multi modal pool training system", Aalborg University, June 1999.

[Brøndsted 1998] Brøndsted, T., Dalsgaard, P., Larsen, L.B., Manthey, M., Mc Kevitt, P., Moeslund, T., Olesen, K. "A platform for developing Intelligent MultiMedia Applications", Technical Report R-98-1004, May, 1998, CPK, Aalborg University

[Davenport 1992]: Davenport, Kim: "Target Pool", Target Pool Productions, P.O Box 219, Marysville, Michigan, 48040, 1992

[Jebara 1997] Jebara, T., Eyster, C., Weaver, J., Starner, T., Pentland A., "Stochasticks: Augmenting the Billiards Experience with Probabilistic Vision and Wearable Computers". Proc. of the International. Symposium on Wearable Computers, Cambridge, MA, Oct. 1997.

[Jensen 2000]: Jensen, P. M., Kammersgaard, K., Kromann, L.: "Intellipool", Aalborg University, June 2000.

[IBM 2000] http://www-4.ibm.com/software/speech/desktop/w8-psl.html

[INFOVOX 1994] "Text-to-speech converter user's manual (ver 3.4)" Technical Report, Telia Promotor Infovox, Sweden.

[Larsen 1997a] Larsen, L.B., "A Strategy for Mixed-initiative Dialogue Control ", in Proceedings of Eurospeech '97,Sep 1997

[Larsen 1997b] Larsen, L.B., "Investigating a Mixed-Initiative Dialogue Management Strategy", in Proceedings of ASRU 1997, Santa Barbara, Dec, 1997

[Lausen 1997] Lausen, H., "LaserXI (2.2), documentation". Laser Interface, Center for Advanced Technology (CAT), Roskilde, Denmark

[Gonzales 1992] Gonzalez, R., Woods, R., "Digital Image Processing", Addison-Wesley 1992.

[Odell 1997] Odell, J., "The Hapi Book (version 1.2)", Entropic Cambridge Laboratory, U.K.

[SUN 2001] http://java.sun.com/products/java-media/speech/