

A Multi Modal Pool Trainer

Lars Bo Larsen, Tom Brøndsted

Center for PersonKommunikation

Aalborg University,

Aalborg, Denmark

Email: {lbl,tb}@cpk.auc.dk

1 Abstract

This paper presents a multi modal pool training system currently being developed at the Center for PersonKommunikation, Aalborg University.

The aim of the system is to automate the learning process, in this case learning the game of pool. The Automated Pool Trainer (APT) utilizes multi modal user-system communication, to facilitate the user interaction. The paper describes the philosophy on which the system is designed, as well as the system architecture and individual modules. The paper concludes with a presentation of some preliminary test results and a discussion of the suitability of the particular and similar systems.

1.1 Keywords

HCI, multi modal interaction, image analysis, computer vision, speech recognition, speech synthesis, natural language processing, spoken dialogue systems, usability, pool.

2 Introduction

The pool training is based on a previous project carried out at the Center for PersonKommunikation at Aalborg University. In this project, 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 was to provide a basic setup, in which new applications, such as the present one can be built. The APT can be viewed as such an application, and a number of the modules from this have been used in the building of the APT.

The Automatic Pool Training system is based on the widely used Target Pool [Davenport 1992], developed by the professional pool player Kim Davenport. The purpose of Target Pool is to enable a trainee to follow a self-study course consisting of a number of exercises.

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.

3 Target Pool

The basic idea of Target Pool is simple: To present the trainee with a number of exercises (accompanied with detailed instructions), 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 introduced, allowing users at different levels to compete. It can also be used to track users' progress over time.

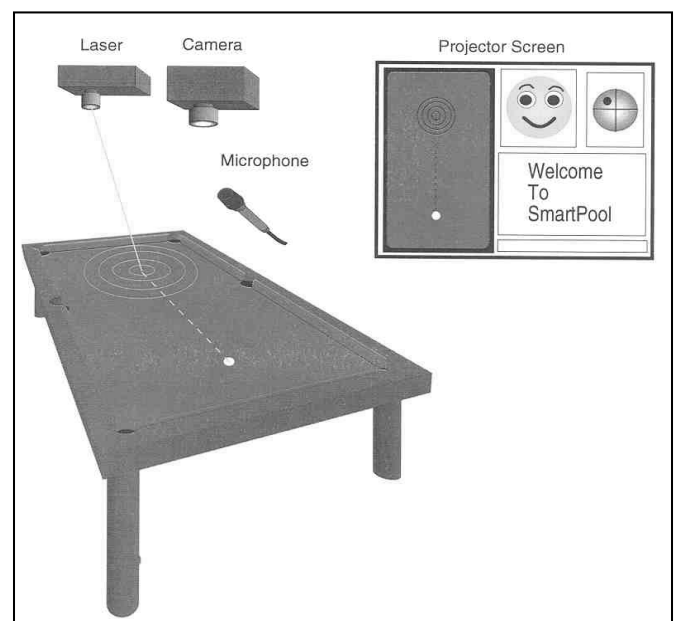


Figure 1 The Automatic Pool Training System uses a camera and a laser to interact directly with the pool table.

The only equipment needed apart from the pool table and queues, etc., are a booklet describing the exercises, a scoreboard, 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.

Figure 2 shows an example of an exercise description for Target Pool.

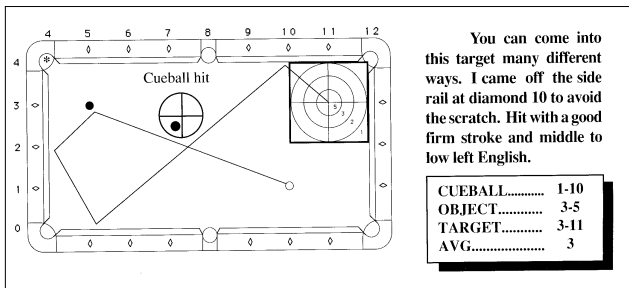


Figure 2. Example of a Target Pool exercise, as shown in [Davenport 92, p.38]

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.

4 The Automated Pool Trainer

We found the Target Pool scheme uniquely well suited as a domain for experimenting with multi modal user interaction, for a number of reasons:

- It addresses a learning situation, which for some time has been of high interest as a domain for computer automation, and thus poses some interesting challenges.
- It is basically for training a manual skill, which the majority of automated learning systems do not address at all.
- A traditional desktop WIMP interface would be unsuitable, as it would not “fit in” the training situation (which is characterized by direct interaction with the physical world (the pool table)).
- The exercises as defined in the Target Pool scheme seemed to be natural candidates for a multi modal presentation, as they consist of a number of graphical elements, as well as an oral (textual) description.
- Furthermore, when consulting pool instructors, we were recommended to use Target Pool, as it was widely known, and is successfully used in training programs in Pool halls.

The APT is basically an automatic system implementing the Target Pool training scheme. To optimise the learning process, Dreyfus [Dreyfus 1986] theory about the transformation “from Novice to Expert” have been employed during the design, as it addresses the learning process as a combination of acquiring both theoretical knowledge and physical skills, which is the case here.

4.1 System architecture and modules

The user addresses the APT by:

- Speech (using a wireless microphone)
- Direct interacting with the pool table (by placing the balls and performing the shots. The camera placed above the table picks this up).

- For the present also by keyboard and mouse (for complex/seldom used commands)

The APT addresses the user by:

- Synthetic speech (reducing the need for the user to switch his attention from the pool table to the monitor screen)
- Graphics (illustrating the spoken instructions, by showing a drawing of the pool table layout, where to hit the cue ball, etc.)
- Video (instant replay movie of users shot for evaluation)
- Text (creating a persistent record of the spoken instructions)
- Laser Pointer (interacts directly with the pool table by drawing on the surface to indicate positions of balls, the target, etc., thus alleviating the need for the user to switch his attention from the pool table)

A projector screen mounted on the wall displays the text and graphics.

4.2 System architecture and modules

As mentioned in the introduction, the APT is built upon the IntelliMedia Platform. It consists of a number of software and hardware modules, loosely coupled in a blackboard architecture. Figure 3 below shows the migration of the original IntelliMedia blackboard to an early version of the APT [Bondesen 1999]. In the present version, the blackboard has been omitted, and modalities are integrated directly.

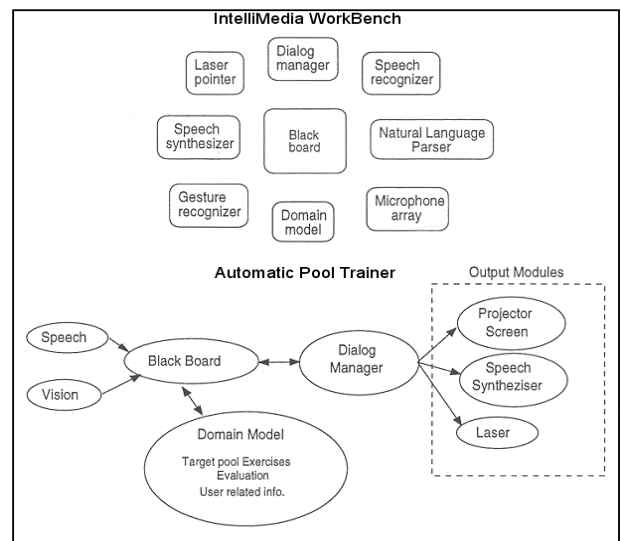


Figure 3 Migration from the original architecture to the APT architecture (from [Bondesen 1999])

4.3 APT Subsystems

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], [Jensen 2000] and [Larsen 2001] for a fuller description.

4.3.1 *The laser subsystem*

A 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. For more details, see [Moeslund 1998], [Brøndsted 1998], [Lausen 1997].

4.3.2 *The Image Analysis Subsystem*

The image analysis subsystem handles the task of detecting the balls on the pool (still and moving). 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 record and evaluate the shot and give the user feedback of how well he or she performed the shot.

All image analysis is performed on difference images. These are obtained by generating a reference image of the empty pool table. 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. The image processing is rather complex, and will not be described in further detail here.

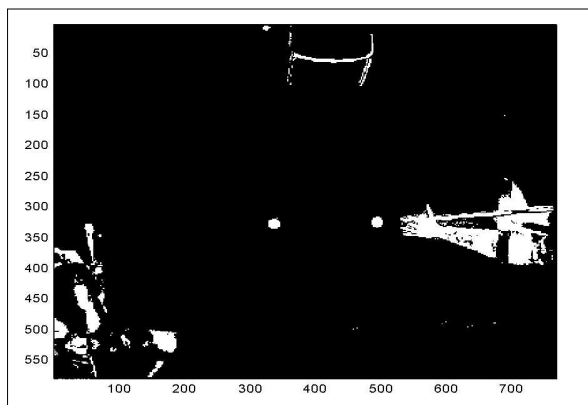


Figure 4. An example of a binary image made by subtracting a reference image from the present image, and applying a threshold to the resulting difference image. The two balls placed on the pool table are easy to identify. (from [Jensen, p.75])

4.3.3 *The speech and NLP subsystem*

Spoken language communication: Speech recognition can be handled by SAPI/JSAPI compliant products like IBM's ViaVoice or by any real-time continuous speech supporting the HTK format for acoustic models. This includes the latest CPK speech recogniser [Christensen 1998], [Olsen

2000] as well as the commercial graphVite recogniser [Power et al. 1997]. Recognition is based on HMMs (Hidden Markov Models) of generalised triphones for acoustic decoding of English or Danish. For training of HMMs, the SpeechDat (II) corpus has been employed [Lindberg 1999]. The language model constraining the recogniser is generated automatically by a grammar converter in the NLP module (see below). For speech synthesis, several commercial products have been employed. Integration of the CPK diphone-based speech synthesiser [Andersen 2000], which is under development for Danish, is being considered.

Natural language processing is based on a compound feature based (so-called unification) grammar formalism for extracting semantics from the one-best utterance text output from the speech recogniser [Brøndsted 1999a], [Brøndsted 1999b]. The parser carries out a syntactic constituent analysis of input and subsequently maps values into semantic frames. The rules used for syntactic parsing are based on a subset of the EUROTRA formalism, i.e. in terms of lexical rules and structure building rules [Bech 1991]. Semantic rules define certain syntactic subtrees and which frames to create if the subtrees are found in the syntactic parse trees. The module is also capable of generating finite state approximations of the unification grammars to be used by a standard grammar constrained speech recogniser like graphVite (HTK Standard Lattice format) or IBM's ViaVoice (JSGF format). A natural language generator has been constructed, however so far generation is conducted by using canned text.

The dialogue is designed to handle both novice and experienced users. This is achieved by implementing a strategy of mixed initiative [Larsen 1997a,b].

Instructions and responses are generated by a combination of synthesized speech, laser gestures and graphics. These are synchronized to provide an integrated response. Because of the non-persistence 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.

Generally speaking, the interaction is designed to be as little intrusive as possible, when the user is actually interacting (playing) at the pool table. This means that the system will be fairly verbose when instructing about an exercise. However, when the user is at the table, the only system output that are really needed are the objects drawn by the laser directly on the pool table. After each trial shot, the system will immediately be ready for the next without any interfering dialog. This ensures the optimal flow of the exercises.

5 User tests

A user test was carried out with nine users, two of whom were pool instructors, skilled in using the original Target

Pool system by Davenport. A number of scenarios were defined and carried out by all users, who were then required to fill out a questionnaire. The following conclusions were drawn from the experiment:

- Most users favoured 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 favourable, and none were confused about the combined display and pool table interaction
- The quality of the voice was too poor.

Figure 5 below shows the user responses to the animated interface agent.

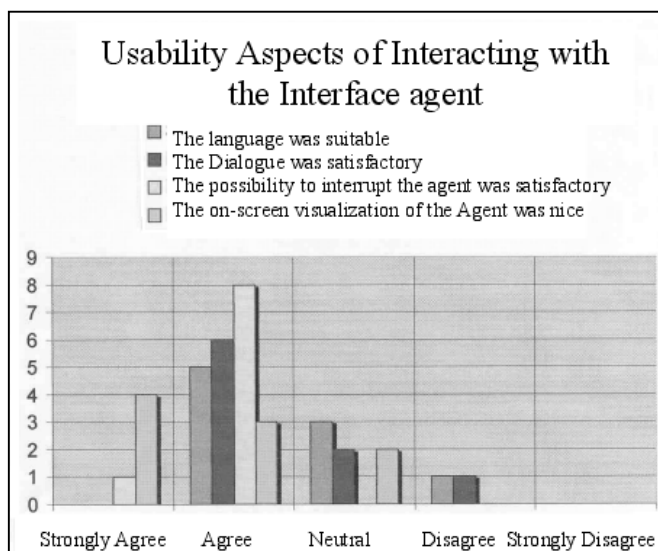


Figure 5. User responses regarding the interface agent

6 Conclusions and Future Work

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 detailed knowledge about the direction and speed of the balls, which puts even higher demands on the image analysis subsystem.

The Stochastic 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, this system was clearly designed to demonstrate wearable computers and augmented reality, rather than creating a training system.

7 Acknowledgments

The authors wish to thank Pernille Bondesen, Søren Poulsen, Morten Lykkegaard, Peter Jensen, Kenneth Kammersgaard and Lars Kromann for their initial work on the Automated Pool Trainer.

REFERENCES

- [Andersen 2000] Andersen, O. C. Hoequist, C. Nielsen: Danish Research Ministry's Initiative on Text-to-Speech Synthesis. In Proceedings of Nordic Signal Processing Symposium, Kolmården, Sweden 2000
- [Bech 1991] Bech, A.: Description of the EUROTRA framework. In The Eurotra Formal Specifications, Studies in Machine Translation and Natural Language Processing. C. Copeland, J. Durand, S. Krauer, and B. Maegaard (Eds), Vol. 2, 7-40 Luxembourg: Office for Official Publications of the Commission of the European Community 1991
- [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
- [Brøndsted 1999a] Brøndsted, T.: The CPK NLP Suite for Spoken Language Understanding. In: Eurospeech, 6th European Conference on Speech Communication and Technology, Budapest 1999
- [Brøndsted 1999b] Brøndsted, T.: The Natural Language Processing Modules in REWARD and IntelliMedia 2000+. In LAMBDA 25, S. Kirchmeier-Andersen, H. Erdman Thomsen (eds.). Copenhagen Business School, Dep. of Computational Linguistics 1999
- [Christensen 1998] Christensen, H., Lindberg, B., Steingrimsson, P. : Functional specification of the CPK Spoken LANGUAGE recognition research system (SLANG). Center for PersonKommunikation, Aalborg University, Denmark, March 1998
- [Davenport 1992]: Davenport, Kim: "Target Pool", Target Pool Productions, P.O Box 219, Marysville, Michigan, 48040, 1992

9. [Dreyfus 1986] Dreyfus H., Dreyfus S.: "Mind over Machines" Oxford 1986
10. [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.
11. [Jensen 2000]: Jensen, P. M., Kammersgaard, K., Kromann, L.: "Intellipool", Aalborg University, June 2000.
12. [IBM 2000] <http://www-4.ibm.com/software/speech/desktop/w8-psl.html>
13. [INFOVOX 1994] "Text-to-speech converter user's manual (ver 3.4)" Technical Report, Telia Promotor Infobox, Sweden.
14. [Larsen 1997a] Larsen, L.B., "A Strategy for Mixed-initiative Dialogue Control ", in Proceedings of Eurospeech '97, Sep 1997
15. [Larsen 1997b] Larsen, L.B., "Investigating a Mixed-Initiative Dialogue Management Strategy", in Proceedings of ASRU 1997, Santa Barbara, Dec, 1997
16. [Larsen 2001] Larsen, L. B., P. M. Jensen, K. Kammersgaard, L. Kromann, "The Automated Pool Trainer - A multi modal system for learning the game of Pool", Proceedings of ICIMADE'01 - Intl. Conf. on Intelligent Multimedia and Distance Education, June, 2001
17. [Lausen 1997] Lausen, H., "LaserXI (2.2), documentation". Laser Interface, Center for Advanced Technology (CAT), Roskilde, Denmark
18. [Lindberg 1999] Lindberg, B.: The Danish SpeechDat(II) Corpus - a Spoken Language Resource. In Datalogvistisk Forenings Årsmøde 1999 i København.. Proceedings. CST Working Papers. Report No. 3, B. Maegaard, C. Povlsen & J. Wedekind (Ed.) 1999
19. [Moeslund 1998] Moeslund T.B., M. Blidegn, L. Bakman "Controlling a Movable Laser from a PC" Aalborg University 1998, Technical Report R-98-1002, ISSN 0908-1224
20. [Odell 1997] Odell, J., "The Hapi Book (version 1.2)", Entropic Cambridge Laboratory, U.K.
21. [Olsen 2000] Olsen, J.: The SLANG Platform: Design and Philosophy, v. 1. Technical Report, Center for PersonKommunikation, Aalborg University, September 2000
22. [Power 1997] Power, K., Matheson, C., Ollason, D., Morton, R. (1997) The graphVite book (version 1.0), Cambridge, England: Entropic Cambridge Research Laboratory Ltd. 1997
23. [SUN 2001] <http://java.sun.com/products/java-media/speech/>