Intelligent Headsets for Supporting Digital Lecture Halls

Christoph Trompler Erwin Aitenbichler Guido Rößling
Department of Computer Science
Darmstadt University of Technology
64289 Darmstadt, Germany

Abstract

Lecture halls are often the main places for discussion between educators and students. They are therefore ideal starting points for incorporating technology into teaching. To be widely accepted, the technology must be useful without distracting lecturer or audience. We present a lecture hall infrastructure and a headset that provide attractive but unobtrusive support for both educators and students.

1. Introduction

Intelligent classrooms and lecture halls have been a hot research area over the last years. Research has followed such diverse topics as presentation recording, ubiquitous integration of local participants and computerized videography. Location, control and awareness of and through participants in the intelligent rooms are one of many aspects of research in this area. In this paper, we present the integration of a device we call the *Talking Assistant* (TA) into these smart rooms.

This paper is divided in the following sections. We first present our existing intelligent lecture hall and its architecture, which we call 'Digital Lecture Hall' (DLH). We then give a general description of the Talking Assistant. Both Digital Lecture Hall and Talking Assistant are being developed by two research groups at our Computer Science Department. The first two sections give an insight in the cornerstones of both projects and discuss related work in their domain. Shortcomings in our Digital Lecture Halls and usage scenarios concerning the TA will be discussed later. Finally, we report on the implementation status, and conclude the paper with a summary and outlook on future work.

2. Digital Lecture Halls (DLHs)

We believe that many computer supporting bursts in education in the past, such as CBT, ITS, Explorative Learning and Virtual Universities, failed due to the missing digital affinity to both educators and students. A lecture hall is the place where most of the teaching is done. Both educators and students share it as a common workspace. Thus, the lecture hall is the best starting point for architectures that bring computer support to the masses in conventional universities.

Our Digital Lecture Halls (DLH) architecture as presented in [6] defines services founded around computer based presentation: the *Virtual Multi Board (VMB)* application combines application recording for broadcast with layered annotation functions known from electronic whiteboards. The lecturer can use multiple private and public working areas, like private pen-tablet displays and public e-whiteboards simultaneously. The emulation of the 'good old sliding blackboard' through the use of projector-arrays is another important feature. We try to compensate for restricted screen estate of computer presentations and whiteboards, as these lead to short reviewing times of presented facts for the students.

Recordings made by the Virtual Multi Board are stored and streamed live during a lecture to the students *ToGather* tool. ToGather enables students with a wireless-LAN capable notebook to directly add their own annotations and lecture notes to the presentation. ToGather also supports individual groups of students, letting them share a common view on the group members private notes. After a synchronous lecture phase, ToGather can be run offline, letting students review their notes wherever they desire. Students may also synchronize parts of their offline lecture archives amongst each other in an ad-hoc operating mode.

The wireless LAN infrastructure found in an DLH setup, provides a digital communication medium between instructor and auditorium. While with a classroom-sized audience, conventional communication should be no problem, in larger venues like introductory lectures with over 600 participants, capturing someone's opinion is nearly impossible. We use a system called *OCLI* [13] to perform computer-based live evaluation of the lecturers performance, opt-in from the students, and deployment of small knowledge tests, quizzes and votes. Our current version also offers pseudo-anonymous participation. Asking questions can therefore become less frightening for some students, as their real identity is hidden. On the other hand, the educator can still send direct replies to the queries.

We are also researching a tool we call the *Lecture TVRemote*. Like OCLI, it can be used for evaluation during the lecture. It also enables the user to set markers for regions of interest during the lecture. These may be attached with categories or short textual notes. After the lecture, the markers are uploaded to our recording server and can be used to download a customized compilation of a lecture.

Digital Lecture Halls also include automated audio and video recording facilities. Our first Digital Lecture Hall prototype was equipped with four remotely controllable cameras for tracking the lecturer and the auditorium.

External applications can be integrated into the DLH architecture. Examples are Learning (Content) Management Systems or specialized computer-based learning materials such as algorithm animation software[9].

Figure 1: Snapshot taken from a lecture recording in the Digital Lecture Hall prototype, showing the lecturer in front of the multi-projection wall.

2.1. Related Work

Classroom 2000 at Georgia Tech [1] is one of the precursors of computer supported lecture rooms. Certainly, a lot has changed since 1999 concerning available hardware and the general equipment of students with hardware. For example, over 40% of the computer science freshmen at our university have a laptop with wireless LAN access. Supporting mobile users is becoming an evident problem today.

Computer supported videography in the DLH is only in early development. Rui et al. [10] present a videography system based on passive audio and visual tracking of the presenter and a 'virtual director' for switching between input sources. In a later section, we will present methods using the Talking Assistant to help tracking lecturers and students.

Computer-based interaction support in lectures has been presented by several other research groups [5, 12]. Compared to the cited tools, OCLI supports different devices through an XSLT-based XML approach requiring only an HTML or WAP browser engine. We have integrated control over OCLI into VMB, reducing the UI overhead while lecturing.

3. Talking Assistant

The Talking Assistant (TA) is a voice-centric user terminal in the form factor of a headset. The headset connects wirelessly to the network infrastructure via a Bluetooth radio.

A combination of sensors is used to gather context information, such as the location of the mobile user in space and his heading in all three rotational axes. The location is determined by an optical tracking system built by our group. It offers a large field of view, high range and centimeter accuracy. The heading is determined with an electronic compass and an acceleration sensor is used to measure the tilt angles.

In a larger context, the TA realizes the concept of a personalized minimal entity (ME) device in ubiquitous computing. A ME serves as a digital representative of the user, and features minimal ("boot strapping") support for communication, interaction, context-awareness, and security.

Amongst lecture support as being described in the context of this paper, first research applications for the TA target also support for office work and mobile work, and personalized exhibition, tour or museum guides.

From these applications, the following key requirements were derived for the TA concerning it self and its supporting tracking system:

- **eyes-free:** A lecturer has to observe the auditorium. A mobile field technician has to focus attention on the components to maintain. In exhibitions, textual explanations on signs or displays distract the user from examining the exhibits, thus reducing the overall experience. With Talking Assistants (TAs), visitors can explore the environment in a natural way.

- **hands-free:** Following the invisible computing paradigm [7], the technology should act unobtrusively in the background. This requires small hardware that does not encumber the user and is easy to carry.

- **location aware:** For office applications, the positioning system must be accurate enough to detect in front of which terminal the user currently is. For the lecture scenario, we wanted to be able to track the auditorium at least with seat resolution in small or medium-sized (200 seats) lecture halls. In exhibitions, the location system must be accurate enough to determine at which exhibit the user is currently looking. This requires an accuracy in the centimeter range and measuring of the user's heading.

- **personalized:** In guidance applications, personalization (e.g., reflecting the user's age, interests, available time, and tour history) is the single most important advantage over conventional electronic guides. The TA can optionally require the user to authenticate herself once the device is put on. We are currently investigating the use of speaker recognition for this purpose. The device stays personalized as long as it is worn by the user. A distance sensor detects automatically if the user takes the device off. While personalized, the TA can be used as a ubiquitous single-sign-on solution for all TA-aware terminals.

- **networked:** The wireless network interface permits the device to act as a "thin client" that relies on services in the network. The computational power is in the service infrastructure and therefore avoids complexity at the end-device.

3.1. Headset Hardware

Figure 2: Talking Assistant

The following section gives a brief overview of the hardware features. A more comprehensive discussion of the previous TA version can be found in [2]. An Atmel 8-bit RISC microcontroller forms the core of the device. A Bluetooth module is used for wireless data communication with the service infrastructure.

The TA supports audio in-/output and features a hardware MP3-decoder for enhanced playback quality and more efficient use of resources like memory and network bandwidth. Recorded audio is PCM or ADPCM encoded and streamed over the network to a service in the network infrastructure that implements the application logic. The TA also features silence detection to avoid unnecessary transmissions.

3.2. Sensors

The sensors are used for determining the position of the mobile user in space and capturing his heading in all three rotational axes. The combination of sensors is essential for best operation. The ADXL202 is a single-chip 2-axis acceleration sensor with a measurement range of $\pm$2g. It can measure both dynamic acceleration (e.g. vibration) and static acceleration (e.g. gravity). We use it as a dual-axis tilt sensor: the force of gravity is used as an input vector to determine the tilt of the device in space. The dynamic acceleration component can be used to detect head gestures like nodding and head shaking.

The heading is measured with an electronic compass which usually must be operated level to the ground. When there is significant tilt present, the heading value is compensated depending on the pitch and roll values measured by the acceleration sensor. Finally, the infrared location system makes use of the heading to determine the direction in which to emit beacon signals.

3.3. Positioning System

Our system called IRIS-LPS (InfraRed Indoor Scout) is an optical infrared local positioning system [3]. The tracked objects carry active tags that emit infrared signals which are received by a stationary mounted stereo-camera. For tracking persons, an active tag can be a badge or a Talking Assistant device. The system is based on cheap off-the-shelf components, is easy to deploy, and features a large range of coverage. It is capable of tracking a large number of tags without significant performance impact, since the sampling rate remains constant with an increasing number of tags.

Eight IR emitter diodes are mounted on top of the headset, each covering an angle of 45 degrees. We use this large number of emitters with a relatively narrow angle to achieve a greater range of the system. Only the emitter that is oriented towards the receiving stereo camera is used to send out a beacon signal at any time. The direction of the stereo camera is given by the heading information, the current position and the world model. Because the emitters are worn on the head, they are visible most of the time to a ceiling-mounted camera. In this case, the direct line of sight requirement is hardly a restriction.

3.4. Related work

Compared to other Bluetooth headsets that are commercially, off-the-shelf available, the TA offers an open hardware and software platform which enables us to add sensors/actors on demand and to run custom software directly on the device.

Concerning location systems, the solutions available today can be classified by the underlying transmission medium as follows:

- **Radio Frequency (RF):** RF-based systems currently offer an accuracy of about 1-3 meters [4]. They do not require a direct line of sight. However, their accuracy is significantly degraded by multiple path and fading effects. Our experience shows that the measurement result can even be influenced by varying the number of people standing close to a tag. RF-based systems are suitable for certain industrial applications. In office-like environments, they often cannot distinguish between different rooms. The state-of-the-art of RF-based technology and its drawbacks are described in [8].

- **Ultrasound:** In the Active Bat system, tags emit an ultrasonic pulse to a grid of ceiling-mounted receivers. The system can locate tags to within 9cm of their true position. Like RF-based approaches, ultrasound-based systems suffer in their accuracy from reflections and obstacles between senders and receivers. Deploying the Active Bat system requires laying out a grid of sensors on the ceiling. This task is rather complex and requires a precise placement of sensors.

- **Magnetic:** Magnetic tracking is commonly used in VR and motion capture applications. It offers a high resolution but is limited to a small and precisely controlled environment. The system performance is typically severely degraded by the presence of metal in floors, walls or carried by persons.

- **Infrared:** A major disadvantage of IR optical systems is that a direct line of sight between sender and receiver is required. However, if the emitters can be mounted at exposed places, like on the Talking Assistant Headset, this restriction does not severely limit their use. Commercially available IR optical tracking systems such as Optotrak and Firefly mainly target medical and motion capture applications. These systems are precise to the millimeter range, but the operation volume is limited to a few cubic meters. Xync offers systems with higher range for studio applications but requires big active tags that are typically mounted on studio cameras.

Compared to those systems, our solution supports a considerably larger field of view and higher range at the cost of reduced accuracy. Furthermore, our system is based on cheap off-the-shelf available components.

4. Integration of the TA into the DLH

In this section, we describe some of the problem areas in the existing DLH system and how these issues can be resolved by integrating TAs.

4.1. Camera tracking

Our first prototypical DLH (Figure 3) as well as further portable DLH setups showed that adding technology to existing lecture halls tends to be problematic. Although the lecturer usually only moves within a bounded area, adapted lecture halls tend to be a hostile environment for tracking people. Especially critical are camera and microphone placement used for recording as well as for passive tracking. Restrictions from existing building structures and fixed furniture arrangements often allow only suboptimal camera and microphone placement. In one case, cameras had to operate at full zoom level and were prone to vibrations in the auditorium. People moved into the camera's light path and further caused the cameras auto-focus to fail in an unpredictable manner. In a smaller auditorium, acoustic echoes from the lecture hall walls, high background noise levels from cooling fans of computer equipment and data projectors as well as feedback from the amplifying system for the speaker made exact instructor and student tracking based on microphone readings imprecise and unusable.

Some pan-tilt-zoom cameras have built-in tracing capabilities, the video capturing cameras we are currently using, come with such a feature. However this functionality is useless since the cameras simply lock, on moving bright areas like the white shirt or tie of the lecturer and not on the face. Since such systems are not able to distinguish between different persons, a student walking through the picture can completely misguide the cameras.

In fact passive tracking without recognition and identification is inappropriate when more than a single speaker is to be tracked on the podium. This is often the case when students are presenting results in an exercise course, where the lecturer and students are standing on the podium.

The tracking system used for the TA is based on active tracking of the IR beacon signals emitted from each TA. Our tracking system allows to simultaneously track and identify all people carring a TA. This allows us to differentiate between different speakers on the podium and unrelated passing by students.

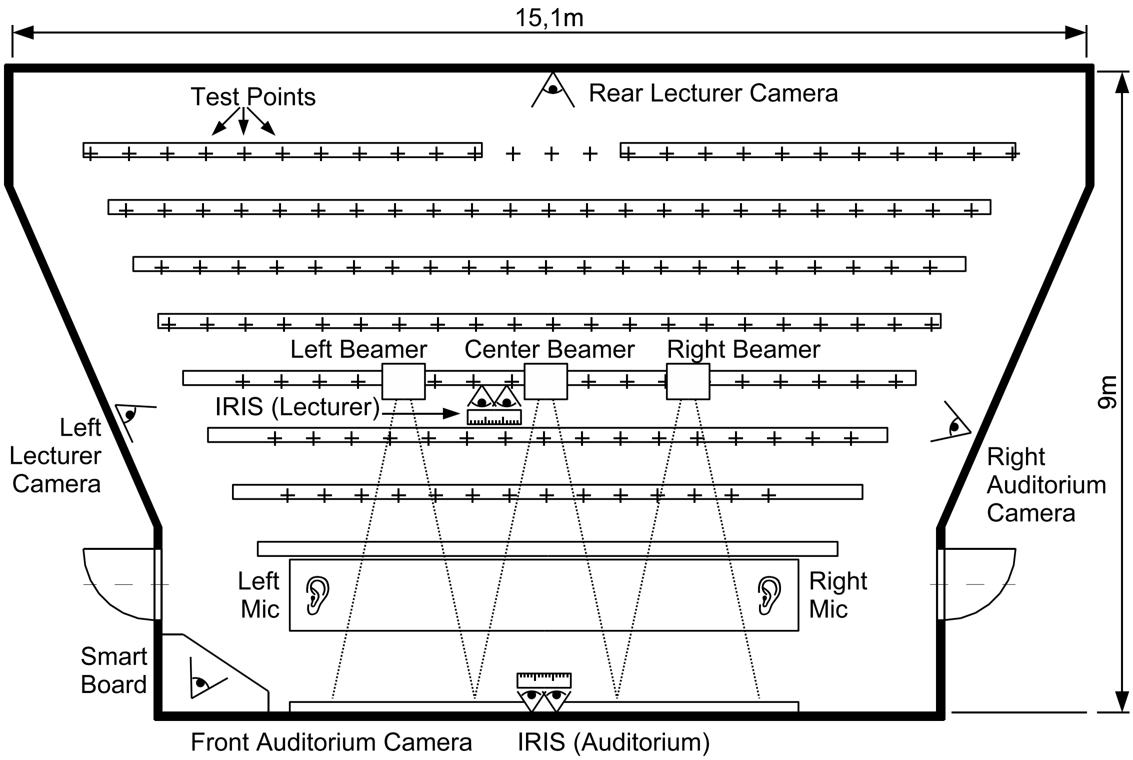

Figure 3: Digital Lecture Hall floor plan

4.2. Student Questions

Distributing TAs among students offers several usage scenarios, especially in larger venues. The TA can be used as hand rising and feedback tool, replacing web-based applications used via wireless appliances and notebooks currently deployed in our DLH environment.

Auditorium microphone systems are usually built on multiple hi-fi wireless microphone transmitter/receiver combos and a multi-channel mixing system in the background. They often requiring an extra operator. The TA can replace these systems using its audio-streaming features. With accompanying services the lecturer can coordinate speakers in the auditorium on his own. Audio mixing, the control over a local amplifying system as well as control of the audio input to the lecture recording systems is fully supported. The location awareness functionalities of the TA, as already presented, allow orienting cameras towards the detected location of a given student. Even if only a few shared TAs are available to the students in the audience, the TA still offers a convenient audience audio recording solution.

4.3. Metadata

Continuously recording interactions of the lecturer with the VMB can generate useful recordings for later review by the student. However, this strongly depends on the working style of the lecturer. Some lecturers enthusiastically employ digital ink on their slides or presented software, using the pen-tablet and the e-whiteboard. Students often dislike this, because of the loss of overview. Other lecturers may simply stand gazing at their projected slides and talk. Even with a video recording showing both lecturer and slide, it is hard to distinguish what exactly the lecturer is referring to.

A solution for this problem would be to track where the lecturer is gazing at, respectively where he has turned his head to. We are investigating if we can use the sensor information gathered from the TA's head-movement tracker to extract useful context information.

4.4. Command and Control Interface

Since the Talking Assistant requires authentication by the user once he carries the device, it can be used as a single-sign-on solution. In the DLH the TA can log on to the Virtual Multi Board software. VMB can then automatically load the user's profile and customize itself.

A subset of the VMB functionality can be directly accessed by means of voice commands. For example, the commands “Next Slide” and “Previous Slide” are used to move around in PowerPoint based Presentations and “Make Overview” shifts the current slide to the rightmost display position where it remains visible during the whole presentation or until the presenter explicitly switches to another overview slide. Since almost all lectures are held in German, we use commands in English as trivial trick to distinguish between lecture and command mode.

4.5. Auditory cues

Without prior screening by the lecturer, our interaction support system does not display the students’ incoming questions, comments and evaluation on display areas visible to the students. Showing the students input unfiltered is no wise idea, since almost certainly students would be making fun of this feature. To give students the feeling that they are not using the system alone, we only display a global contribution count.

So only the private display of the lecturer displays notifications of arriving questions and results from continuous live evaluation. Since the private display is usually a pentablet (Figure 2), which the lecturer also uses to annotate and control his presentation, he will regularly receive the students’ feedback. This is in theory. Practically, visual notifications were often overseen by the lecturer.

A resort are auditory cues [11]. We use the audio playback features of the TA. Coupled with the silence detection feature of the recording unit, the notifications can be played back as soon as the lecturer makes a short break. Notifications sound like the well known “academic knock” for a question, yawning if the lecture becomes too boring or moaning if it becomes too tough. In case of an raise hands request as described before hand, the text-to-speech engine can instruct the lecturer where in the auditorium the student has virtually requested to speak.

4.6. Benefits for students

As the TA is primarily an appliance to be used by the lecturer, benefits for the student are mostly of indirect nature: They can expect lecture recordings of better visual quality because of improved camera positioning. The lecturer is able to react sooner to comments and substantial questions of the students because of auditory cues. Especially in large venues, the TA enables students to individually participate in interactive sections, such as open discussions or feedback gathering. In several scenarios, this type of interactivity may in fact only be possible with a TA or similar appliances.

4.7. Benefits for the Lecturer

The TA offers several advantages for lecturers. First of all, it offers hands- and eyes-free operation, thus allowing the lecturer to interact with arbitrary other devices and focus on the auditorium. Lecturers are typically busy with presenting material and do not monitor displays that list incoming questions by students. Auditory cues remind the presenter of new messages in an unobtrusive way. Auditory cues can also be used to provide acoustic feedback about the students’ evaluation of the current lecture. By acting on this information, lecturers can also improve their presentation style over time. Lecturers also benefit from the support for interactivity, as discussed in the previous section.

5 Current results

The floor plan of the DLH is shown in Figure 3. Two video recording cameras are oriented towards the lecturer and two cameras are oriented into the auditorium. Two ceiling-mounted room microphones can be used to record audio in case the lecturer does not use a TA or radio microphone. Three beamers provide a large and seamless projection area on the front wall. The lecturer can use a Wacom tablet on the front desk or the Smartboard to make annotations in the presented material.

A test installation of our IRIS positioning system has been evaluated in the DLH. The camera was mounted in front of the blackboard at a height of about 3 meters. Because the camera is fitted with wide angle lenses, it can cover nearly the full room which measures 15.1 x 9 meters. The lecture hall is in the basement, has no windows and is illuminated by fluorescent light. All lights were turned on during the test. The infrared emitter consisted of one infrared LED with a narrow angle of 20 degrees. One such LED easily ranges 10 meters and more.

The infrared emitter was then placed at one test point after the other. At each point, the position calculated by the positioning system was compared with the expected position and the error distance was calculated. We determined the expected positions by hand using a measuring stick. Thus, the accuracy of the reference system is limited. The result is displayed in Figure 4. The Root Mean Square (RMS) error calculated from all 138 test points is 16.67 cm. The graph shows that the measurement accuracy decreases with increasing angle from the camera axis and with distance.

Evaluations of our undergraduate lectures last winter indicated problems with lecture video recording, as the lecturer often left the area covered by the cameras. Combining the location-aware TA with automatic orientation and zooming of the recording cameras shall result in improved video quality.

All major software components, like room- and camera-control, VMB, ToGather, OCLI and the Talking Assistant Services provide XML/SOAP-based Web Service-Interfaces. Using the VMB as controlling instance, all components are easily integrated with each other.

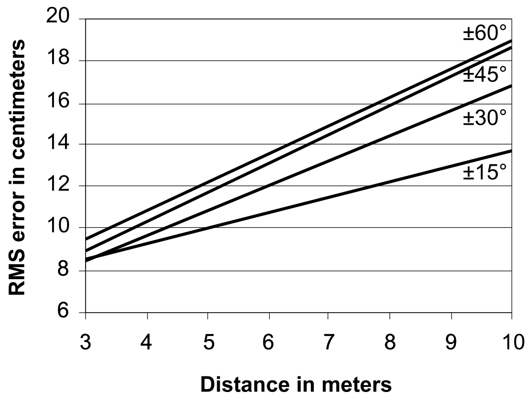

Figure 4: Root Mean Square accuracy of the IRIS tracking system as a function of distance and opening angle.

6. Summary and Conclusions

We have explored the use of context-aware ears-and-mouth devices in a digital lecture hall setup. Equipping the lecturer with a Talking Assistant has several advantages.

- The headset permits a clear audio recording of the speaker, since the distance between mouth and microphone remains constant and is not influenced by head movements. If multiple speakers are wearing TAs, the person who is currently talking can be determined virtually with 100% accuracy.
- Our infrared positioning system provides a robust solution for orienting the recording cameras. Compared to passive image-based tracking systems, the image processing involves less computational resources. Furthermore, the system offers a higher accuracy and reliability.
- The lecturer has a private audio back channel to receive auditory cues.

In our future work, we will investigate the use of context information acquired by the Talking Assistant, like speaker identity, position and heading to derive further presentation meta data. The TA version presented in this paper is based on Bluetooth. The next version currently under development will be based on WLAN to support a higher number of clients, for improved bandwidth and to better match the existing WLAN infrastructure in our lecture halls.

References

[1] G. D. Abowd. Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment. *IBM Systems Journal*, 38(4):508–530, 1999.

[2] Erwin Aitenbichler and Max Mühlhäuser. The Talking Assistant headset: A Novel Terminal for Ubiquitous Computing. Technical report, TK-02/02, Telecooperation Group, Department of Computer Science, Darmstadt University of Technology, 2002.

[3] Erwin Aitenbichler and Max Mühlhäuser. An IR Local Positioning System for Smart Items and Devices. In *Proceedings of the* 23^{rd} *IEEE International Conference on Distributed Computing Systems Workshops (IWSAWC03)*, pages 334–339, 2003.

[4] Jeffrey Hightower and Gaetano Borriello. Location Systems for Ubiquitous Computing. *Computer, Issue on Location Aware Computing*, Volume 34 Number 8:57–66, 2001.

[5] Martin Mauve, Nicolai Scheele, and Werner Geyer. Enhancing Synchronous Distance Education with Pervasive Devices. In *Informatik 2001* , volume 157, pages 1117–1122, 2001.

[6] Max Mühlhäuser and Christoph Trompler. Learning in the Digital Learning Age: Paving a Smooth Path with Digital Lecture Halls. In 35^{th} *Annual Hawai'i International Conference on System Sciences*, volume 1, pages 31–41, 2002.

[7] Don A. Norman. *The Invisible Computer*. MIT Press, 1998.

[8] K. Pahlavan, X. Li, and J. P. Mäkelä. Indoor Geolocation Science and Technology. *IEEE Communications*, pages 112–118, 2002.

[9] Guido Rößling and Bernd Freisleben. ANIMAL: A System for Supporting Multiple Roles in Algorithm Animation. *Journal of Visual Languages and Computing*, 13(3):341–354, 2002.

[10] Yong Rui, Liwei He, Anoop Gupta, and Qiong Liu. Building an Intelligent Camera Management System. In *Proceedings of the Ninth ACM International Conference on Multimedia*, pages 2–11, 2001.

[11] Nitin Sawhney and Chris Schmandt. Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 96–103, 1999.

[12] Paul G. Shotsberger and Ron Vetter. Teaching and Learning in the Wireless Classroom. *Computer*, 34(3):110–111, 2001.

[13] Christoph Trompler, Max Mühlhäuser, and Witold Wegner. Open Client Lecture Interaction: An Approach to Wireless Learners-in-the-Loop. In *4th International Conference on New Educational Environments*, pages 43–46, 2002.