Auditory Interaction Objects for Mobile Applications

Palle Klante

Kuratorium OFFIS e.V., Oldenburg, Germany.

Abstract. In mobile working situations users have to take care of their environment and interact with a mobile device. Besides the double workload an attention switching, because of the two-fold usage of the visual sense takes place. An alternative solution for mobile scenarios is the usage of an auditory environment for an eyes-free interaction. This paper presents such a concept of auditory environments and demonstrates the idea and the design of auditory interaction objects for mobile applications.

Keywords. Hearcons, Mobile, Auditory Interaction Objects, Usability Engineering

1. Introduction

Mobile devices are often used in totally different scenarios, places, and tasks than traditional desktop computers. But when we inspect the user interface of the mobile devices, we find that the same interaction paradigms are used. Due to this fact some problems arise: In mobile scenarios users have to take care of their environment and at the same time interact with the device. Because of this double workload (resulting in an attention switching) a two-fold usage of the visual sense takes place.

The aim of this approach is to use the visual sense only to orientate oneself in the real world. Various signals in our environment are presented and perceived over the visual sensory channel. The usage of an auditory environment with an eyes-free interaction is an alternative for mobile scenarios to transmit information. Information objects and functions will be presented as auditory cues in a three dimensional auditory virtual world around the user. They can be manipulated by the user.

This paper presents a framework to improve the usage of auditory environments and demonstrates the idea of auditory interaction objects (audIO). Each audIO is composed of multiple spatialised, interactive, and permanently played auditory cues, so called Hearcons (Bölke et al., 1995). A Hearcon represents the smallest unit in an auditory environment, and can be characterized by the following attributes:

- a spatial position in the virtual environment
- a specific sound
- a variable volume and
- an interaction area to interact with the object.

The auditory interaction objects are enriched with speech output and input. As the audio objects are located in a three-dimensional environment with the user in the center, one needs an interaction device with three axes to interact and manipulate these objects. The three-dimensional selection of these objects is a difficult task while on the move. Due to this fact we decided to use a small hardware device with simple and robust interaction capabilities. In combination all Hearcons of the audIOs are placed in predefined areas which are easy to target with that digital device.

We developed a Java-based toolkit of auditory interaction objects. As a first prototypical example an auditory MP3-Player was implemented. This application demonstrates various interaction tasks for a typical mobile workplace.

The remainder of this paper first presents a brief survey of the state-of the-art of auditory displays in mobile applications. After that, the general Usability Engineering process for building usable auditory environments is presented (Chapter 3). The result of this process is the definition of the mentioned auditory interaction objects for mobile applications. They are introduced in Chapter 4, followed by an evaluation conducted with the developed MP3-Player application which uses these interaction objects (Chapter 5). The paper closes with a conclusion and presentation of the main achievements.

2. Related Work

First steps on auditory displays were made by the definition of Earcons (Blattner et al., 1989) and Auditory Icons (Gaver, 1989). They were mainly used to reduce usability problems in graphical user interfaces. The idea for augmenting a mobile workplace with a virtual auditory environment on the other hand is quite new. As a consequence there exist only a few guidelines on how to design such auditory information spaces.

In the AudioAura (Mynatt et al., 1998) for example physical actions are used to trigger auditory supported processes. By entering a room a summary of received e-mails (while absent) is presented. By moving in front of a colleague’s door auditory information is presented on how long the colleague will be away. At least a group pulse was presented to indicate the working situation of colleagues working in distance.

In the Nomadic Radio (Sawhney et al., 1999) different presentation forms for the same information units are investigated. The system used a scalable notification technique to present information to the user according to the actual workload in mobile situations. The project focuses on the information presentation techniques, rather than on an interactive selection and manipulation method. As a consequence, the user´s interaction possibilities are quite simple. There was no way to select information.

With AccesSights (Klante et al., 2004) a system is presented which enables blind and visually impaired persons to receive multimedia tourism information on the move. The necessary information to solve the arising tasks are communicated by an auditory environment. Users can actively ask for more information and can select the desired information parts.

Gesture interaction techniques for wearable devices, which guarantee real eyes-free interaction on the move, were successfully investigated by Brewster et al. (Brewster et al., 2003). They use different concepts for information presentation and interaction (hand and head gestures). The usage of gesture techniques in our application domain is problematic, as users normally move their hands and head when walking. The system must decide if the movement is part of an input or not.

First steps in developing auditory interaction objects have been done by enhancing interaction objects (Widgets) in graphical user interfaces in order to make them more usable. The SonicFinder (Gaver, 1989) uses Auditory Icons and later Lumsden et al. (Lumsden et al., 2001) made some investigations on how to reduce the slip-of problems on buttons and in menus by the help of Earcons. A major result of their work was an architecture to integrate auditory support into graphical user interfaces by developing an abstract layer for event handling and media mapping.

A method on how to design graphical interaction objects for virtual environments is described by Beu et al. (Beu et al., 2001). They split up the process into multiple phases and therewith reduced the design complexity. Their approach considered the idea of iteration. Additionally they looked for an approach where in each task various parallel designs are created by competing design teams. They received a large number of alternatives from which they were choosing the best.

The results of the above mentioned projects and the experience of usability engineering methods are combined in a special usability engineering process for auditory environments. This process will be described briefly in the next paragraph, before presenting the development of the mobile auditory interaction objects and their evaluation.

3. Usability engineering of auditory interaction objects

The idea to develop an auditory environment should be based on contextual or task related requirements. It is one of many possible solutions to present information. So before starting the process one has to decide if the actual context, environment, and user characteristics and the users' tasks are suitable for an auditory environment. Nevertheless there exist some arguments in favor of the development of an auditory environment: The users are blind, they need there eyes for other tasks, a background process must be continuously checked, etc.

In order to build auditory environments we defined a complete process to support the members of an interdisciplinary design team. The process is horizontally structured in five phases (see Figure 1).

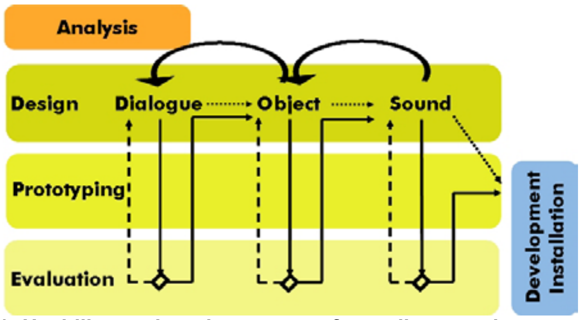

Fig. 1: Usability engineering process for auditory environments

The process starts with an analysis phase, in which the user characteristics, the context, the environment, and a hierarchical task description are analysed and defined. Additionally usability goals must be described, which have to be fulfilled by the final product. Based on this the interdisciplinary design team enters the first horizontal phase, the design phase, which is split up in three vertically organized parts, as shown in Figure 2.

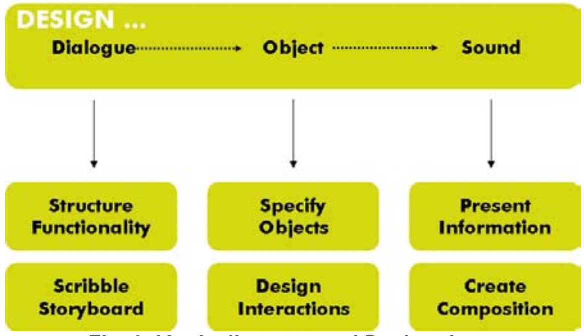

Fig. 2: Vertically separated Designphase

The design phase is dominated by transforming the hierarchical task description of the analysis phase into a dialogue structure. In the first vertical step of the design phase (structuring the dialogue) the task descriptions are therefore transformed into those functionalities the auditory environment should provide. Afterwards the underlying objects from the functionalities must be identified. These steps results in building up a storyboard from the objects. Each object defines a new scene. The objects´ functionalities are forming a group around the object and are integrated in the scene. After connecting the scenes the storyboard describes the complete dialogue structure of the auditory environment.

After this first design phase, another more detailed design phase begins by analysing the identified objects and functionalities. Goal of the analysis is a list of all information an object should communicate to support the user by solving the task. The objects and functionalities are finally handed over to the sound design, to create an acoustic composition containing all presented information.

The next horizontal phase is prototyping. For each of the aforementioned vertical steps a sample prototype can be build, as well as a complete prototype including all vertically design decision. The prototypes are evaluated with regard to the predefined usability goals in the following horizontal phase. The process should be iterated if problems occur. At the end of this process the development and installation of the final application takes place.

4 Auditory Interaction Objects

Looking at the above mentioned mobile applications and auditory environments one can see the different possibilities and design alternatives. When analyzing mobile applications we often found similar basic tasks the user has to solve when he is working with user interfaces. These tasks describe basic functionalities each application should provide:

- Structuring information in a hierarchical manner. An information space is typically very large, navigation strategies must be provided in order that the user will not get lost in the presented information space.
- Structuring information in lists. Separators must be provided to support the user´s perception of the content.
- Presenting information about the current location in the information space. Identification Hearcons providing information about active objects. A scroll bar Hearcons gives information about the actual location within a menu.
- Selecting objects from a given set.
- Start a process or use the functionality.
- When handling large data sets like searching in data sets or updating them, the process often does not finish immediately. The ongoing process can be observed and controlled by some kind of progress indicator.
- When problems occur within an application, different sorts of messages must inform the user about the actual problem and the state of the application (confirmation, information, warning and error messages).

With regard to Vanderdonckts categorization of abstract and concrete interaction objects (Vanderdonckt, 1995) and the above mentioned tasks we developed a first toolkit for the usage in mobile scenarios with minimal input possibilities. The toolkit offers the following objects:

- an auditory interaction space
- a menu
- a menu Item
- a separator
- a progress bar
- an edit Box
- a selector 1of N
- a selector M of N

To make these interaction objects available for other developers all of them are documented by the following characteristics:

Short description	***Presenting the main idea and functionality of the interaction object***
Elements	List of re-used elements
Illustration	Graphical presentation of the interaction idea
Behavior	Behavior without user action
Actions	Feedback on user actions
Speech	Used Speech Dialogues
Criteria	Important acoustic and HCI Criteria
Adaptivity	Personalization, Individualization

Tab 1: AudIO characteristics

We choosed to implement our interaction objects for a PDA (e.g. iPAQ) in order to use the widest available hardware and software resources. This brought us to some design limitations concerning the input possibilities. Mobile speech input technology is very poor at the moment and can be used only for basic commands. As a very simple interaction form we integrated the PDAs hardware buttons. The auditory output is realized through open headphones.

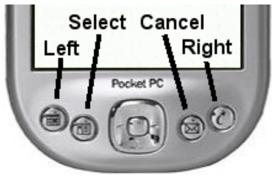

Fig. 3: Four Button Input Device

4.1 Interacting with the audIO

The used auditory environment consists of three main areas and considers the limitations described in Donker et al. (Donker et al., 2002). Figure 4 shows the rooms areas:

- An interaction and information panel in front of the user, which contains the menu items and all selectable elements. Because humans can localize objects best when they are in front of them, most of the objects should be placed there.
- Behind the user in medium height is the identification area to present information which is always active while the application is running (modus and information about active objects).
- Behind the user, at the bottom, a status area is located. In the status or message area short information of ongoing processes is shown.

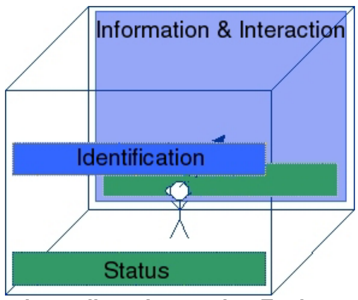

Fig. 4: Areas in the auditory Interaction Environment

In the following, the design and interaction concept of the auditory interaction objects is explained. After starting the application a short speech output gives the application’s name. Figure 5 shows the initial state of the application. The actual auditory environment consists of a menu with three elements (each element is represented by a Hearcon): In front of the user the menu items are located. In front bottom left corner the scroll bar audIO is located. It signals the actual position in the menu (here at the beginning of the menu). Behind the user the identification Hearcon for the actual menu is presented.

Two objects are in front of the user: A focused one in the center and another one on the right side. In order to select the next object the user presses the right key. Now the objects are moving to their new positions. The focused MenuItem reaches its new position on the left and the right MenuItem moves to the middle and a new third MenuItem appears on the right. It is the third menu item in the actual menu. While moving, the MenuItems are changing their volume constantly. The new focused center object is getting more loudly. While moving the menu Hearcons, the scroll bar Hearcon is moving to its new position. This Hearcon reaches its maximum sound level, when all other Hearcons have 50% of their maximum volume. The user can not localize the scrollbar Hearcon exactly. This should only give an overview, how much objects are within the presented menu.

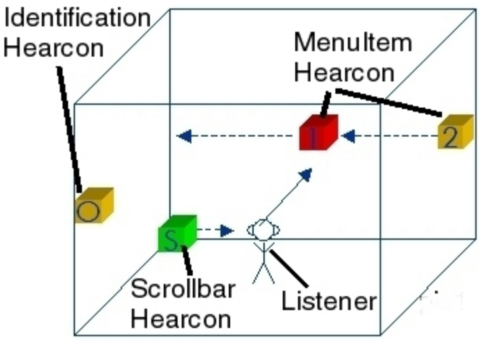

Fig. 5: Interaction example for the auditory menu

4.2 Design characteristics

For all interaction objects specific sounds were generated. We mandated a composer to create a musical arrangement and built a library of instrumental sounds for the auditory interaction objects. For events and messages associative sounds were chosen. It would be easier to interact with the system, if all objects have an associative sound. But this design alternative was out of question, as every day sounds annoy after a short time. Additionally it is very difficult to get a large number of different everyday sounds, which still fit together.

Three dimensional perceptions is a key factor for auditory environments. When perceiving various interaction objects at once it is necessary to support a good differentiation mechanism for the underlying Hearcons, which also brings up the problem of masking. Hearcons can get lost for the user, when they overlap. If the Hearcons are positioned too close together and have similar sounds, e.g. by using the same rhythm or timbre, the user will perceive one big element.

Presenting information with moving objects is very important for recognizing the current system status. It is much easier to remember and identify an object, when one knows where it comes from and how it changed over time. Moving the object and varying the volume generates a history for the user. An example: After the user activates the selected MenuItem in a Menu, the focus MenuItem moves to the user´s back and gets slowly more silent. When reaching the identification area at the back it is only half of its volume. At this moment the scroll bar is rising in. Appearing at the current location (at the beginning of the menu on the left) the volume is reaching its maximum volume for a short while and then falls back to its minimum. After the user knows his current position in the menu, the menu items appear by slowly raising their volume. After a while, again depending on the user´s selected parameters, all the objects are getting more silent, and the focus Hearcon presents its label by synthesized speech output. After the speech output terminates the objects are getting louder again.

Considering the problem of annoyance the system goes down after some settable time of inactivity. A special reactivation routine informs the user about the actual system state after pressing one of the buttons.

5 Evaluation

This paragraph explains the evaluation of the auditory interaction objects. Therefore we developed a sample application which uses the objects. After presenting this scenario the evaluation is described and finally the results are presented.

5.1 MP3-Player Scenario

As a sample application for the evaluation we developed an auditory MP3-Player. Today's mobile music equipment benefits from miniaturization and small form. But there are design limitations: If a MP3-Player has more functions and features other than playing and stopping sound files, a graphical display must be integrated, which makes the device very large. In our scenario the user can select menu items with his favourite music songs or start a predefined play list. Functions like "edit play list", "edit directories" are positioned in a menu hierarchy in front of the user. Status information like "battery charge state" and "percentage of play list" are located as background information behind the user. The stored songs are presented in a hierarchical organized menu; maybe with album information or organized by other attributes, like interpret or genre. Even with this large functionality the system still can be managed with only a few keyboard buttons.

5.2 Evaluation structure and used Methods

As the toolkit implementation was at an early development stage we decided to conduct a development attendant evaluation with a formative character. We did not conduct the evaluation in a mobile environment, because of hardware restriction.

Eight participants took part in the evaluation. They were all students and did not work in the field of auditory environments. They had no experience in using auditory environments before the evaluation, but they were mostly familiar with the functionality of software MP3-Players. Their age ranged between 22 and 30 years. They all had no hearing impairment.

Before the evaluation a short introduction to the system was given. Afterwards the participants "played" with the system for 15 minutes and then received a description about the tasks they had to solve within the test scenario. They had to solve eight tasks in 30 minutes which concerned different aspects of the used auditory interaction objects. The participants had to move horizontally within the menu and navigate up and down in the hierarchy. Three tasks concerned the ignition of functions, like playing an MP3-file, getting more information about a given MP3-File, or receiving help from the system. By virtually downloading an MP3-File they had to use the progress bar. Saving a play list was only possible by using the selection objects.

Five evaluation methods were used for extracting different user data:

- User monitoring to record the user´s behaviour.
- Recording interaction steps in a Log-File.
- Thinking loudly while navigating and working with the test scenario.
- Interview to receive more information about the observed user behaviour.
- Questionnaire to receive subjective information about how the users liked the system.

5.3. Results

Different navigation strategies were observed to explore the auditory environment. The strategies depended on the user's mental model they built by using the system and on their learning abilities. Four types of strategies could be identified (see table 2).

Rough Navigation	The user already entered the correct menu hierarchy level, but is still far away from the desired object.
Fine Navigation	The user is already close to the desired object
Wayfinding Down	Move from the top of the hierarchy to the bottom
Wayfinding Up	Move from the bottom of the hierarchy to the root element

Tab 2: Observed navigation strategies

In a more rough navigation, when the users thought they were still far away from the desired target, they explicitly used the Hearcons. Experts who already knew the meaning of the presented sound used the Hearcons when reaching the target area. Novice users did explicitly not use the Hearcons, but waited on the speech output, when they reached the target area. Because in the first prototype the speech output was started too fast, the users did not have the possibility to learn the objects sounds. This made the interaction very slow, as the parallel perception of sound was eliminated.

All participants had difficulties in finding functions located deeply in the hierarchy, which is a problem of building good mental models. When users get lost, they started to walk back all the way up the hierarchy. Therefore it is necessary to provide a function "walk back" to a well known landmark.

The moving objects were helpfully, when navigating through the hierarchy levels. Especially the identification Hearcon which moves from the middle of the interaction area to the identification area, when navigating through the hierarchy downwards provides useful hints about the current location. On the other hand the way from the hierarchy bottom to the top level, was not well designed in the first evaluation. There was missing a hint in which level the user is actually located. This was derogatory for the users, because they had to remember the whole application structure.

The fixed areas of information and interaction, identification and status were useful to categorise the presented information. The users did not have to search their information. This effect was also identified when using the menu, with its fixed positions for the menu items. The users could concentrate themselves on the sounds to identify the objects and did not have to try to select them. But this "slot-mechanism" limits the number of parallel presented items. Only three elements in the information area in front of the user are useful, as interaction is getting more difficult with a larger number.

The users could easily recognise the objects, remember their location and were able to use the known history of the objects to perceive the actual system status. The scroll bar information in the menu was very well accepted and recognized. The pause between moving the menu items, showing the scroll bar and then rising again the menu items, was easy to perceive. The users get an overview, where in the menu they are, and how many elements the menu contains, before the end is reached. It helps the users a lot when looking for the desired area in the rough navigation.

The dynamically changing volumes of the objects were implemented to reduce the number of parallel presented information units. This effect is especially noticed when we were updating the menu structure and the focus Hearcon flew from the front to the back. After reaching the new position, the new focus Hearcon is slowly moving in, and the user could easily concentrate on this new object.

Our evaluation has shown that the majority of participants had much fun to work with our application and could easily use the offered functionality of the auditory interaction objects. For the MP3-Player as an entertainment product the criteria innovative, funny, exciting etc. were mentioned. In contrast to using a traditional MP3-Player, we see a lot of benefits in our system.

In this evaluation we learned once again, that the sound design is one of the major problems in designing auditory environments. But we think that with a robust interaction design, it is possible to have imperfect sounds and a still usable system.

6 Conclusion

There are nevertheless some limitations when designing auditory interaction objects for auditory environments. The number of features must be smaller than for typical graphical displays and the hierarchy depth must be very low. Most of the participants had difficulties to finish the tasks of the test scenario, where it was necessary to move too deep (level > 3) in the hierarchy. Additionally there is a problem with the composition of objects. A modification of one object has a direct impact on other objects. A scroll bar is used in a menu, a menu is used in a menu and also in a selection, etc. It could also be that a progress bar is starting while an important system message is arriving. As a special necessity in auditory environments you have to provide a mechanism to remove not used interaction objects in each step. In graphical User Interfaces it is possible that objects overlap one another. The user will not have difficulties to recognise the correct and active object at the moment. But in auditory environments that is impossible. The parallel output of Hearcons must be designed very carefully. Interferences will change the presentation of objects and could result in making more mistakes.

The toolkit is realised through a model-view-controller concept. So it is possible to integrate other scenarios with different tasks and different interaction objects. The open architecture supports this idea. But it is a problem to transfer a given GUI to the audIO, because of the menu hierarchy complexity and the large number of parallelly presented objects. The user interface transformation from a mobile device is instead much easier, because similar problems exist and the hierarchy depths and the number of menu items is already reduced. The future aim is to implement this linking and offer a new adaptive set of interaction objects for any application.

7. References

Beu, A.; Hassenzahl, M.; Quaet-Faslem, P.; Burmester, M. (2001), *Benutzerzentrierte Gestaltung eines mobilen Service- und Wartungssystems unter Verwendung von AR-Technologie*. in MMI-interaktiv, Nr.5, Juli/2001

Blattner, M. M.; Sumikawa, D. A.; Greenberg, R.M. (1989), *Earcons and Icons: Their Structure and Common Design Principles*. In Buxton, W. (ed.) Journal of Human-Computer Interaction (1989), Special Issue on Nonspeech Audio; Volume 4; Number 1. Lawrence Erlbaum Associates, Inc, Page 11-44

Bölke, L. und Gorny, P. (1995), *Direkte Manipulation akustischer Objekte*. In D. Böcker (Hg.): Proceedings Software-Ergonomie '95. Fachtagung des German Chapter ACM und der GI. Darmstadt Feb. 1995. Stuttgart, 1995: Teubner, Page 93-106

Bowman, D.A.; Kruijff, E.; LaViola Jr., J.J.; Pouyrev, I. (2001). An Introduction to 3-D User Interface Design. in Presence, Volume 10, Number 1, February 2001, Massachusetts Institute of Technology, Page 96-108

Brewster, S.; Lumsden, J.; Bell, M.; Hall, M.; Tasker, S.; (2003). *Multimodal 'Eyes-free' Interaction Techniques for Wearable Devices*. In Proceedings of CHI2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA. ACM Press, Page TODO-TODO

Donker, H.; Klante, P.; Gorny, P. (2002), *The Design of Auditory User Interfaces for Blind Users*. in Proceedings of the Second Nordic Conference on Human-Computer Interaction NordiCHI, Tradition and Transcendence, Edited by Olav W. Bertelsen, Susanne Bodker and Kari Kuutti, October 19-23, 2002, Aarhus Denmark, The Asscociation for Computing Machinery, Inc. New York 2002, Page 149-155

Gaver, W.W.; (1989), *The SonicFinder: An Interface That Uses Auditory Icons*. in William Buxton (ed.) Journal of Human-Computer Interaction (1989), Special Issue on Nonspeech Audio; Volume 4; Number 1. Lawrence Erlbaum Associates, Inc, Page 67-94

Klante, P. (2003). *Visually supported design of auditory user interfaces*, in Human-Computer Interaction: Theory and Practice (Part II) Volume 2 of the Proceedings of the 10th HCI International 2003, 22 - 27 June 2003, Crete Greece, Edited by Constantine Stephanidis and Julie Jacko, Lawrence Erlbaum Asscociates, Publishers, Mahweh, New Jersey, London, Page 696-700

Klante, P.; Krösche, J.; Boll, S.; (2004); *AccesSights - A Multimodal Location-Aware Mobile Tourist Information System*, accepted for the 9th International Conference on Computers Helping People with Special Needs, ICCHP2004, Université Pierre et Marie Curie, Paris, France, July 7-9, 2004

Lumsden, J. and Brewster, S. (2001), *Guidelines For Using The Toolkit of Sonically-Enhanced Widgets*. Department of Computing Science Technical Report TR-2001-100, October, Glasgow University; http://www.dcs.gla.ac.uk/research/audio_toolkit/tkpublications.shtml

Mynatt, E. D.; Back, M.; Want, R.; Baer, M.; Ellis, J. B. (1998), *Designing Audio Aura*. In Proceedings of the CHI98 18-23 April Los Angeles, Page 566-573

Sawhney, N.; Schmandt, C. (1999), *Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging*. In Proceedings of CHI 99 Pittsburgh USA 15-20May 1999, Seite 96-103

Vanderdonckt, J. (1995), Tutorial 12: *Tools for Working with Guidelines*. HCI International ´95, 6th International Conference on Human-Computer Interaction, Yokohama, Japan