For Print in: Proceedings ICCHP 2000 - International Conference on Computers Helping with Special Needs. Karlsruhe 17-21 July 2000.

TYPOGRAPHIC SEMANTICS OF WEBPAGES ACCESSIBLE FOR VISUALLY IMPAIRED USERS

Mapping Layout and Interaction Objects to an Auditory Interaction Space

Peter Gorny[1]

Abstract

Binaural sound perception allows humans to localize sound sources in space. Modern spatialization soundcards already offer the possibility to place virtual sounds in a virtual auditory environment. We use this to augment the acoustical environments of blind users with virtual auditory interaction objects („hearcons") representing the interaction objects of a webbrowser GUI. The cursor is also represented as hearcon, e.g., by a humble bee, and allows to manipulate hearcons like normal icons, with touchpad, trackball, or mouse. Supplemented with text-to-speech- or braille-output the visually impaired user can receive the textual information of a webpage including its semantics coded in its typography, topography, and topology of its components.

Keywords

Binaural perception, spatialized sounds, auditory interaction objects, barrierfree GUI, accessability, visually impaired users, blind users, accessible internet.

1. Introduction

Imagine a blind user sits at her computer equipped with a touch pad to control the cursor. She is investigating internet pages, e.g., in search of interpretations or transcriptions of Shakespeare's Taming of the Shrew into modern English. All buttons and links on the web pages are mapped into an auditory interaction space produced by a system of four loudspeakers and a standard spatialization soundcard. She can handle the interactive elements in the web browser window with the same ease as a sighted user - by just moving the cursor, represented by the sound of a humble bee, to one of the buttons or hypertext links on the screen, which are also represented by spatialized sounds („hearcons"). All hearcons seem to hang on a (virtual) sound wall behind the computer as an augmentation of her normal acoustical environment *(Fig. 1)*. With the first „click" the interaction object is selected and a spatialized voice reads its name, with the second click, the object is activated. If it is, for example, a text document, the text is read by a voice „from the off" (beyond the sound wall area) via a text-to-speech transformer, and/or presented on a braille output line.

In the preliminary project SPUI-B (Sterophonic User Interfaces for Blind Users) [3], [4] we have investigated the feasibility of the interaction concept and qualitatively evaluated it with the help of 18 blind users.

For the implementation of the described concept in a given application (a web browser) we have to overcome several difficulties, namely:

- the directional discretization of sounds by the human aural perception system is approx. 1/20 of the visual perception;

---

[1] Prof. Dr.-Ing. Peter Gorny, OFFIS / University of Oldenburg, D-26111 Oldenburg
Internet: gorny@uni-oldenburg.de, www-cg-hci.informatik.uni-oldenburg.de

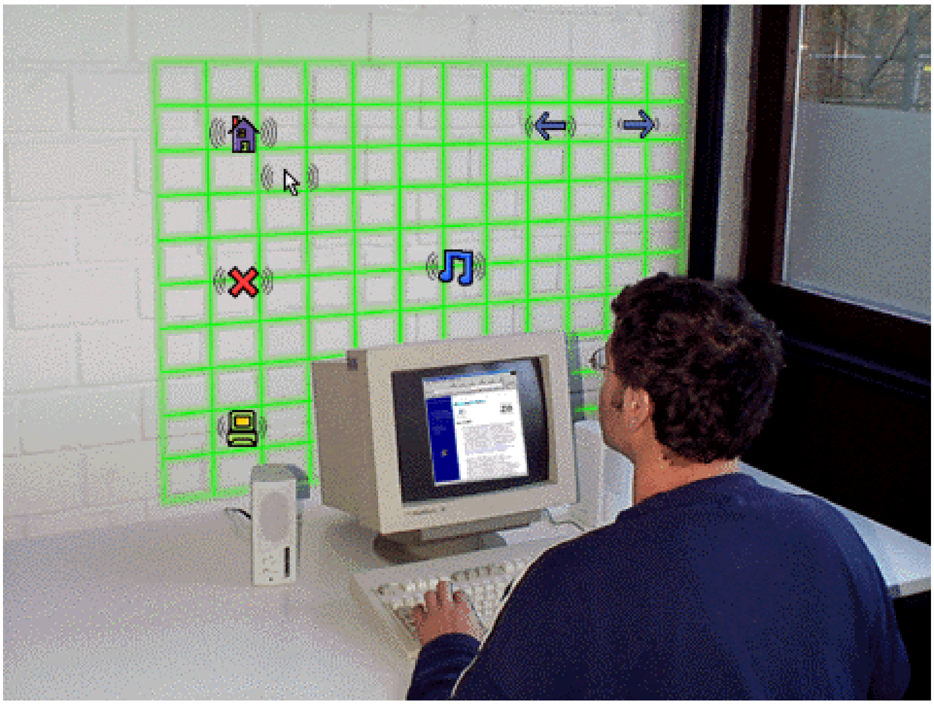

Fig. 1: A virtual sound wall augmenting the real auditory environment of a user

- the directional discretization of two adjacent sounds is high (approx. 3°) in the direction of the nose (the lateral angle α=0°) and very poor (α>30°) at the direction α>75° – therefore humans have learned to turn their face toward the sound sources;
- the localization of a sound in forward direction cannot be distinguished from a sound coming from backward without turning the head slightly.
- more than five simultaneous sounds cannot be distinguished by most humans – they become part of the white noise;
- the use of four or five loudspeakers for the localization of spatialized sounds forces the user to keep the head in the ideal point between the loudspeakers;
- the use of headphones makes it necessary to measure the turning of the head and to rotate the augmented auditory space in the opposite direction in realtime, so the relation between the sounds from the real auditory environment and the augmented reality stays constant (presumed the headphones are „open" to the environmental sounds – a necessary precondition for blind users);
- the use of an "auxel" array (auditory pixels) overcomes most of the disadvantages of headpho-nes or loudspeakers, but this technology is commercially not yet available;
- the various types of icons or links or other interactive UI-objects normally represent something statical: for example, they *offer* a system function. Therefore there are normally no metaphorical sounds available for the interaction objects. The sounds have to be selected for ideographical purposes;
- the selected sounds will partly appear simultaneously – easily leading to cacophony;
- the texts on web pages are very often multilingual, e.g., German and French users will have to cope with English labels on buttons or in URL, or part of the page content is presented in English. Therefore the text-to-speech system has to be constructed as an intelligent assistant in order to choose the appropriate pronounciation.

The solutions we developed so far for these problems for our assistive system will be described in this short report.

2. GUIS and their barriers for visually impaired users

The advantages of GUIs (versus textual interfaces with command and menu-oriented control) have been discussed widely for more than one decade. We refrain from repeating all arguments and reasons and list only the four most important benefits in the context of our subject:

1. The use of metaphores related to the user environment visually represented by icons, which are references to data structures, programs, functions, etc.;
2. Direct manipulation of these visible objects (icons) with rapid visible feedback for all actions;
3. Ease of control of simultaneously available programs and functions;
4. Structuring and organizing visible objects topographically on the screen area by linking semantics to their positions and to layout features (type, size, style and color of the fonts, background color, borders, tables, etc.).

GUIs aim at enhancing the usability of computer systems for non-computer-specialized users for their normal work tasks, but they fail completely for visually impaired users - on the contrary: they cause new obstacles for blind users of computer systems.

Boyd [5] stated the following three problems ("barriers") which cause difficulties for blind users to work with GUIs in the same way as sighted users do:

1. The *pixel barrier* refers to the problem that the screen output is stored as a pixel map in memory and cannot be read by usual text-to-speech systems. Therefore it is necessary to intercept the text output from all software components before it is mapped to the screen buffer.
2. The *mouse barrier* refers to the problem that blind users cannot handle the mouse (or any other pointing device) in an effective way as an input device; this is a psychomotorical problem resulting from the lack of feedback.
3. The *graphics barrier* is located on the semantical level. It refers to the fact, that information is presented in pictorial elements which humans recognize as (previously learned) patterns, and that topography and topology give additional hints about the represented objects and their relations. The transformation of these graphical representations will always be accompanied with a loss of information. Textual descriptions of graphics tend to be long, imprecise and complicated.

As already stated the problem of the graphics barrier is a semantical one and it is questionable whether it will ever be sufficiently solved: Graphics are used to represent an object in an obvious way. For example many icons are simplified pictures of objects the sighted users *see* in their everyday life such as the trashcan or the folder. If such icons are transformed into a set of tactile pixels they cannot be interpreted with the same ease, because a blind person knows the objects only from touching them in their original size. The miniaturized tactile form may be not only ambiguous, but it looses completely the metaphorical link to the real environment. In contrary to pixel graphics, though, some progress can be observed to represent vector graphics (lines) as tactile maps [16].

In a few projects attempts were made to transform graphics into auditory output [20] or to describe pictures for the blind [11], but there still exist no practical solutions to represent any pixel graphics via auditory display.

So far there are only few known solutions for the *layout barrier* caused by the use of layout features for semantic purposes. A word printed in different font types may have completely disjunct meanings. One of the most extreme examples is the ALGOL68 Revised Report [21], which is literally ununderstandable if presented in only one type and style of font, because four different font types and styles are used to connotate the syntactical and semantical level. Here some text-to-speech systems, though, offer the option, that changes of text layout are explained, while the text is read.

3. Existing Adaptations of GUIS for blind users and other related work

Most of the known adaptations of GUIs for blind users do not solve all three barrier problems. For example the GUIB project (Textual and Graphical User Interfaces for Blind People) [11] transforms a screen contents of *MS-Windows* on the one hand into a tactile form, keeping the topographical arrangement of the objects. On the other hand the tactile output is supported by auditory output with spatialized sound. A sound is attached to every object of the GUI and also to the cursor. The sound sources seem to be located in the same position as the visual objects on the screen. Thus the user gets the information about the absolute cursor position on the screen. Other objects only sound when they are touched by the cursor. No direct selection of an object is possible, because no feed-back is given about the relative position of the cursor to a target object when moving the mouse. This situation is comparable with a black screen where only the cursor is visible and a sighted user has to find a certain object on the screen which will light up when touched – an unsatisfying situation.

The SonicFinder [12], [13], [14], an adaptation to make the Finder interface of the Macintosh accessible to visually impaired users, acts in the same way. The objects are implemented as auditorily passive objects which only sound when a certain event has happened. In a later version of the SonicFinder *soundholders* were introduced, which can be attached to an object and which sound permanently to provide a cue of its distance to the cursor. This distance is represented by the sound volume reciprocally to the distance from the cursor.

The Soundtrack project [9], [10] implemented the absolute locations of objects by different tones (organ-keyboard-metaphore), but as tests proved, only a blind musician was able to retrieve the positional cue from the tonal coding, the other visually impaired test persons counted the objects when passing by with the cursor.

In the Mercator project [17], [18], [19] XWindows under UNIX is adapted. The authors solve the navigation problem by replacing the graphical display with a tree-like hierarchical auditory interface. To select a certain object the tree has to be searched by using the cursor keys until the target object is found ("down" to move to a lower hierarchy, "right" to select a branch, "up" and "left" for the reverse operations). The objects again only sound when a certain event has happened.

As a conclusion we may state that all known adaptations of GUIs for blind users do not solve all three problems mentioned above. Blind users are still unable to access GUI-based application systems in the same way as the sighted do. In general practice the GUI is reduced to a textual menu/command-oriented interface using cursor keys for navigation. Neither the semantics of topographical arrangements of UI objects nor their semantical relations hidden in their topology can be recognized by the blind users. Direct manipulation of objects is not possible, since no sufficient feedback is provided to use the pointing device adequately.

4. Special WWW-barriers in an virtual auditory space

Since we have chosen to implement our auditory interaction space for a web browser we have to cope with a set of problems resulting from the deficiencies of the html-standard, when also used for layout purposes, e.g., using dummy gifs for text spacing, tables, and frames, etc.

The presently known auditory adaptations of web browsers use text-to-speech systems. One of the most advanced systems is the IBM Homepage Reader [1], which offers straight forward sensible solutions for normal hypertext links, for interactive gifs (links behind graphics elements), for tables, and for frames. But this browser is not offering sufficient support for a blind user in a collaborative situation within the joint environment with sighted persons.

No solutions for giving access to the GUIs of Java applets have become known to us. Java-script based interaction elements can normally be deactivated without blocking access to the links behind the element, so this will normally not create serious obstacles.

It is also foreseeable that the fast development progress of the html-standard, will cause difficulties to maintain our adaptation of a web browser.

5. Direct manipulation of auditory objects in a virtual auditory space

Bölke [3] and Yang et al. [22] describe the results of our preliminary project SPUI-B (see also [4]), which can be summarized in the following points:

- Interaction with spatialized interactive auditory user interface objects (hearcons) is feasible for blind users, when not more than 5 hearcons (besides the cursor hearcon) sound simultaneously and with full volume: an implementation of the metaphor "walking by night with a torch through a forest" (distant trees are hardly visible).
- To reduce the number of simultaneously audible hearcons there should be also an easy way to group hearcons.
- The default sounds representing hearcons must be easily adaptable to personal preferences of the users. A large set of alternative sounds must be provided within the system.
- The user must be able to improve localization of a hearcon by moving and turning his head – with a stereophonic loudspeaker system, with an auxel array, or with "open" headphones.
- The mouse is not well suited for blind users; more usable pointing devices are touch pad, or trackball.

In the current project the evaluation of our design decisions with respect to interaction techniques, applicability of sounds, and the usability of the user interface for the adapted application system (webbrowser) will be one of the central topics in order to develop some general recommendations on auditory adaptations of GUI-based applications.

6. Long-term goals of the project ZIB

The short-term goal of the project ZIB is the development or adaptation of a webbrowser with a normal graphical user interface extended with an augmented auditory reality, in which the interactive objects of the user interface are represented by spatialized sounds (hearcons).

Our long-term goal can be best expressed by the report of a blind author, who describes his experience of a garden in rain: the sound of the rain drops on the leaves, the grass, the pavement gave him a direct perception of their topographical arrangement and of their relation to each other [15]. We want to create an intuitive access to the semantics of the topographical environment of a visually impaired person, including access to the virtual objects of computer-mediated information, which have to be integrated into the auditory environment. We are developing a webbrowser which is an assistive adaptation of a given system, and not a specialized technology that would only increase the isolation of a blind person in collaborative situations.

Acknowledgements

ZIB is a subproject within the large project consortium *INVITE – Intuitive User Interfaces for the Information Society of the Future* [23], supported by the German Federal Ministry of Education, Science, Research, and Technology for the period 1999 - 2003. It is organized by the author with the support of OFFIS (Oldenburg R&D-Institute for Informatics Tools and Systems) and in cooperation with Prof. Dr. V. Mellert, ITAP (Institute for Technical and Applied Physics), and Prof. Dr. A. Schick, Dept. of Psychology, University of Oldenburg.

Partners in the INVITE-Consortium will provide us with solutions for video-based headtracking, for gestural input (instead of pointing devices), and for coping with the problem of mixed languages on web pages, which leads to difficulties for text-to-speech systems. Our product will be used by the other partners to enhance the access to the internet for visually impaired persons, who use it for work or learning, information retrieval, e-commerce, entertainment, and leasure.

References

1. C. Asakawa, C. Laws: Home Page Reader: IBM's Talking Web Browser. URL: http://www.rs6000.ibm.com/sns/hprctg.htm
2. M. M. Blattner, D. A. Sumikawa, R. M. Greenberg (1989): Earcons and Icons: Their Structure and Common Design Principles, Human Computer Interaction 1989, Vol.4
3. L. Bölke (1997): Ein akustischer Interaktionsraum für blinde Rechnerbenutzer. Dissertation Universität Oldenburg 1997. Berichte aus dem Fachbereich Informatik 1/97.
4. L. Bölke und P. Gorny (1995): Direkte Manipulation von akustischen Objekten durch blinde Rechnerbenutzer. In: H.-D. Böcker (Hrsg.): Software-Ergonomie 95 – Fachtagung der German Chapter of the ACM und der Gesellschaft für Informatik vom 20. bis 23.2.1995, in Darmstadt Stuttgart 1995: B.G. Teubner. 93-105.
5. L. H. Boyd, W. L. Boyd, G. C. Vanderheiden (1991): Graphics-Based Computers and the Blind: Riding the Tides of Change, in: Proceedings of the 6th Annual Conference „Technology and Persons with Disabilities“ Los Angeles, 20-23.3.1991
6. A. S. Bregman (1990): Auditory Scene Analysis - The Perceptual Organization of Sound, MIT Press, Cambridge 1990
7. M. Cohen, L. Ludwig (1991): Multidimensional audio window management , International Journal of Man-Machine Studies, 1991, Vol 34, pp.319-336
8. M. Cohen (1993): Throwing, pitching and catching sound: audio windowing models and modes, International Journal on Man-Machine-Studies, 1993, Vol 39, pp 269-304
9. A. D. N. Edwards (1988): Design of Auditory Interfaces for Visually Disabled Users, in Human Factors in Computing Systems Proceedings of the CHI '88, ACM SIGCHI
10. A. D. N. Edwards (1989): Soundtrack: An Auditory Interface for Blind Users, Human Computer Interaction 1989, Vol.4
11. P. L. Emiliani (Project Manager) (1993): Publications from the TIDE project GUIB until to September 1993, GUIB Consortium, 1993
12. W. W. Gaver (1989): The SonicFinder: An Interface that uses Auditory Icons, Human Computer Interaction 1989, Vol.4
13. W. W. Gaver, R. B. Smith (1990): Auditory Icons in Large-Scale-Collaborative Environments, in H. Diaper et al. (Eds.): INTER-ACT '90, Proceedings of the 13 3rd Int. Conference on Human-Computer Interaction, Cambridge, U.K.. 27-31 August, 1990
14. W. W. Gaver, R. B. Smith, T. O´Shea (1991): Effective Sounds in complex Systems: The Arkola Simulation, in ´Reaching through technology´, CHI ´91 Proceedings, 1991
15. J. Hull (1992): Im Dunkeln sehen – Erfahrungen eines Blinden. 2. Aufl. (transl from English), Verlag C.H. Beck, München 1992.
16. R. Michel: Interaktiver Layoutentwurf für individuelle taktile Karten. Dissertation Universität Magdeburg, Fakultät für Informatik. März 1999. (In print)
17. E. D. Mynatt, W. K. Edwards (1992): Mapping GUIs to Auditory Interfaces, in Proceedings of the UIST´92, Nov.15-18, 1992, Monterey, CA
18. E. D. Mynatt (1994): Designing with Auditory Icons: How Well Do We Identify Auditory Cues, in: Proceedings of the CHI'94 Conference, Boston, M., April 24-26, 1994
19. E. D. Mynatt (1997): Designing Auditory Cues for Mercator. In: S. Howard et al (eds.): Human-Computer Interaction Interact '97. Proceedings IFIP Conf., Sydney July 1997. 236-243.
20. S. Smith, R. D. Bergeron, G. G. Grinstein (1990): Stereophonic and surface sound generation for exploratory Data Analysis. In CHI ´90 Proceedings April 1990
21. A. van Wijngaarden et al. (eds.): Revised Report on the Algorithmic Language Algol 68. Springer-Verlag New York etc. 1977.
22. J. Yang, P. Gorny, V. Mellert: Sound Pressure Control Compensation for Head Movement of Listeners. In: A. Schick, M. Klatte (eds.): Contributions to Psychological Acoustics. Results of the Seventh Symposium on Psychological Acoustics. Oldenburg 1997: BIS Universität Oldenburg. 281-292.
23. J. Ziegler: Benutzerschnittstellen – mehr als nur Oberfläche. Information Management & Consulting. 14 (1999), 4. 63-68. http://www.invite.de