Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

A GENERIC, SEMANTICALLY BASED DESIGN APPROACH FOR SPATIAL AUDITORY COMPUTER DISPLAYS

Christopher Frauenberger, Robert Höldrich

Institute of Electronic Music and Acoustics
University of Music and dramatic Arts
Inffeldgasse 10/3, 8010 Graz, Austria
frauenberger@iem.at, hoeldrich@iem.at

Alberto de Campo

Academy of Media Arts Cologne
Peter-Welter-Platz 2,
D-50676 Cologne, Germany
adc@khm.de

ABSTRACT

This paper describes a design approach for creating generic computer user interfaces with spatial auditory displays. It proposes a structured depiction process from formulating mode independent descriptions of user interfaces (UIs), to audio rendering methods for virtual environments. As the key step in the process a semantic taxonomy of user interface content is proposed. Finding semantic classifications of UI entities corresponding to properties of auditory objects is the ultimate goal. We beleive that this abstract approach detaches the process from visual paradigms and will reveal valuable insights into the representation of user interfaces in the auditory domain.

Possible ways of accessing operating systems for UI information are discussed along with an overview over common accessibility interfaces. Critical aspects are highlighted for the composition of auditory UI entities in spatial environments and state-of-the-art techniques are presented for the creation of 3D audio. Besides some possible fields of application, relevant utility and usability engineering aspects are discussed.

1. INTRODUCTION

The most common mode for human-computer interaction, the visual mode, is dominating user interfaces, but the importance of exploiting other modes of interaction is increasing. This has various reasons, firstly, the content of interfaces is growing and becoming more complex. Secondly, in more and more applications the visual mode is restricted by form factors, the mobility of the user or simply by being occupied for other tasks. Thirdly, the fact that computers play a more central role in our society nowadays, builds up the awareness that they must be available for all parts of society. People with visual disabilities have major disadvantages in accessing computers because of the lack of efficient non-visual user interfaces.

In comparison to graphical user interfaces (GUIs) little is known of user design methodologies, usability engineering and design principles for efficient auditory user interfaces (AUIs). Research has shown that auditory displays can be utilised in various applications and provide efficient solutions for specific problems [1, 2, 3, 4, 5]. In contrast, this paper is intended to propose a design approach towards a generic, user centred and semantically based auditory display for human-computer interaction. It should initiate a discussion about more generic approaches to human-computer interaction in the auditory domain detached from visual paradigms.

The goal is to find methodologies for interpreting common computer user interfaces in spatial auditory environments. We believe that spatial, virtual environments are capable of presenting user interfaces much more efficiently than common sequential audio techniques (like screenreaders) and can even adequately replace graphical user interfaces. The process of presenting a user interface in the auditory domain is proposed as:

→ User Interface (UI) Description

→ Taxonomy of human-computer interaction semantics

→ Mapping into the auditory domain

→ Audio Rendering

As a starting point the process needs a mode independent and abstract description of the user interface to be presented. Further: *"In a human-computer dialogue, the way in which things are said will be closely related to what is said"*[6]. Following this idea, a semantic classification scheme will be necessary in the depiction process to find the relevant properties of UI tasks for auditory representation. UI entities will be mapped into the auditory domain and composed into a virtual scene which can be rendered in different output formats. UI entities may also provide methods for interaction. The biggest challenge in this depiction process is to find proper mappings between the semantic UI classes and acoustic representation classes, because only with a content aware method of mapping it will be possible to find coherent mappings of UI tasks ensuring their interoperability in the interface as a whole.

According to this depiction process this paper is structured as follows:

- Section 2 provides an overview over accessibility interfaces and discusses how to come to a user interface description.
- Section 3 concentrates on methods of taxonomy and discusses semantically based schemes to classify UI tasks.
- Section 4 discusses the mapping process including the mapping of UI entities, audio scene modelling and audio scene description with MPEG-4.
- Section 5 shows various possibilities for how MPEG-4 streams can be rendered. Different output formats are described along with state-of-the-art sound wave reproduction techniques.
- Section 6 shows possible application fields for such a framework.
- Section 7 discusses aspects of utility and usability engineering effecting AUIs.
- Section 8 concludes the thoughts presented.

2. ACCESSIBILITY INTERFACES

The first step for presenting a user interface in the auditory domain is to know about the content and the functionality (task model) of the user interface. In case of newly designed interfaces any information about structure and content is available by the design anyway. This is not the case for the task of transforming existing user interfaces into the auditory domain. We believe it is crucial for the success of audio interfaces to support the transformation of existing user interfaces. This might make compromises necessary, but ensures the utility in a wide range of applications or platforms and less development effort for "cross-mode" user interfaces.

However, there are a couple of possibilities to access user interface information in a modern computer operating system's architecture. As an example figure 1 shows a UNIX like X Windows architecture with the GTK+ graphical toolkit for GUIs. For representing the application's user interface in the auditory domain it is possible to rewrite only the device dependent layer of the X Windows server (1). This would be a very difficult task as the requests for this layer are already very specific for graphical depiction. The Mercator project [7] aimed to intercept the communication between X Clients and the X Server (2) and to interpret these calls for audio devices. The protocol is, however, very complex and also very specificly defined for graphical output. The most convenient way would be to retrieve user interface information directly from the application (4), but that would mean to rewrite each application for the use with AUIs.

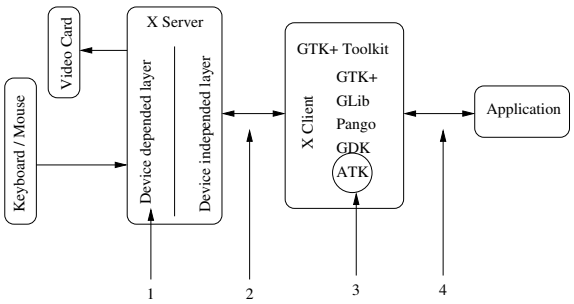

Figure 1: X Windows architecture with GTK+ Toolkit

Modern desktop environments are built upon graphical toolkits like the GTK+ Toolkit to make developments easier and more powerful. GTK+ provides APIs (Application Programming Interfaces) for building widgets, buttons and other graphical elements easily. Within GTK+ the ATK library is responsible for exposing information about the application's face through an accessibility interface (3). It is intended to provide information for assistive technologies such as magnifiers or screen-readers.

Like in this example other desktop environments like Windows or KDE do expose information through their accessibility interfaces. At this level of abstraction of the representation of user interfaces, elements are less specific to their graphical depiction. Therefore, they are good candidates for being the source for the proposed auditory depiction process. Although these accessibility interfaces are not exclusively intended to serve AUIs as information source they provide good information at the right level of abstraction. Vice-versa spatial auditory displays are not exclusively meant as assistive technology, but they definitely may improve the access of the visually impaired and blind to modern information technology.

The subsequent sections shortly describe the two most prominent accessibility interfaces.

2.1. Assistive Technology Service Provider Interface

The Assistive Technology Service Provider Interface (AT-SPI) was developed by the GNOME accessibility project GAP [8]. GAP works on the definition of the AT-SPI standard [9] and the development of applications which use the interface to make the GNOME desktop environment accessible for people with disabilities. The most developed application is Gnopernicus, an integrated screen-magnifier and screen-reader for GNOME.

AT-SPI is intended to become the standard accessibility interface for UNIX desktop environments. It was developed by the GNOME project, but was designed to support other desktop environments as well. The KDE desktop project which is based on the Qt Toolkit announced its support for AT-SPI [10] recently and also Java applications are supported [11]. AT-SPI can be considered as a common standard for most of the GNU/Linux world and some UNIX derivatives on which open source desktop environments can be used.

Through its API it exposes as much information about graphical user interface widgets as possible. Figure 2 shows a typical representation of a GNOME terminal application through AT-SPI using the tool *at-poke*.

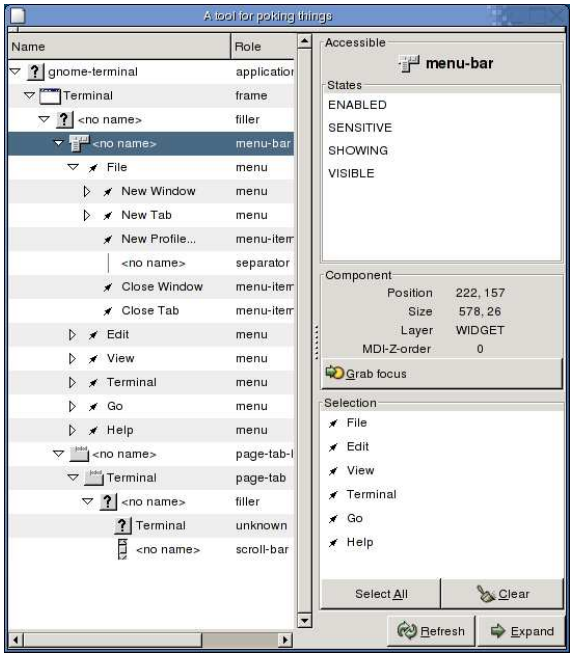

Figure 2: Information about a GNOME terminal through AT-SPI

2.2. Microsoft Active Accessibility

The system architecture of the Microsoft Windows family is different to the client-server architecture found on X Windows based systems. However, Microsoft's effort Active Accessibility[12] resulted in a similar accessibility interface exposing all graphical user interface elements and their properties. The interface description is available from [13].

Because of the dominance of Windows based personal computers the MS AA framework is the most used and best known accessibility interface. Most of the assistive technology software products like the screenreader JAWS[1] are available with MS AA support.

3. INFORMATION MODELLING

After knowing about the structure and the content of the user interface the more complex process of information modelling needs to provide every information for an automated mapping process. This includes to formulate a mode independent representation of the UI and the classification according to a semantic taxonomy methodology. This extra information about user interfaces will allow us to be absolutely detached from visual concepts, but have further control over the meanings and the intention of user interfaces.

In the graphical domain cross platform and cross compiler developments have resulted in abstract description languages like XUL (XML User interface Language)[14]. XUL descriptions are independent representations of user interfaces and can be automatically transformed into programming code for a certain language and platform.

Listing 1: Simple XUL Example

```
<?xml version="1.0"?>
...
<toolbox flex="1">
  <menubar id="sample-menubar">
    <menu id="file-menu" label="File">
      <menupopup id="file-popup">
        <menuitem label="New"/>
        <menuitem label="Open"/>
        <menuitem label="Save"/>
        <menuseparator/>
        <menuitem label="Exit"/>
      </menupopup>
    </menu>
  </menubar>
</toolbox>
```

Listing 1 shows a simple example of a menubar. XUL implements scripting similar to Java Script to describe the functionality of the user interface. Other markup languages for user interfaces include UIML, XIML, XAML or AUIML [15].

We propose to extend such description languages with semantic classification schemes to find an input source for automatic UI mapping. Semantic information about UI tasks are more important for their acoustic representation than they are for their graphical representation. For example, a button in a small widget will be rendered the very same way every time it appears in any kind of dialog. The affordance of pressing the button is given by the button-metaphor, but its intention is only provided by its label or its surroundings. For an acoustic representation of the button it is important whether it is a confirmation of an alert or the update of the view of directory listings. Such information must be made available for the mapping process.

An hierarchical, semantic classification scheme as sketched in figure 3 determines the relevant properties of the interaction tasks. Such a scheme must consider the specific properties of acoustic

[1] registered trademark by Freedom Scientific

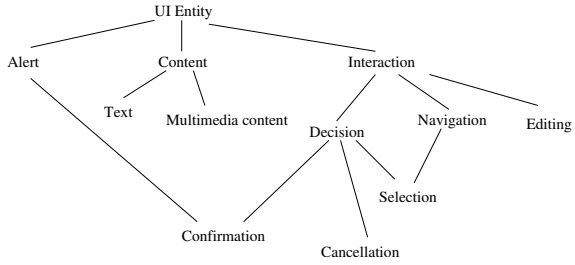

Figure 3: Semantic classification scheme

perception. A design of the scheme needs to answer the following questions:

- Which semantic classes are available from user interfaces?
- Which semantic classes need to be distinguished considering acoustic representation? Which are relevant?
- Can sound properties be found to reflect the classifications?

The given example in figure 3 is already very near to the elements available from the interface to build. To start at the very top level of classification the following taxonomy of human-computer interaction content was proposed in [6]:

Dialogue control information: Not concerned about the task, but handling the interaction itself.

Task information: The task to solve. Large differences in classification must be expected between application domains.

Domain knowledge: Extended domain-related information.

It is stated that these general classifications seem not be feasible, but constitute a basis for a taxonomy. Unfortunately, little work can be found related to this topic; most is concerned about a taxonomy of graphical representations of user interface entities. This paper is intended to provide the ignition spark for a discussion about mode independent, semantic classifications and their use for auditory displays.

The question of finding auditory classes for semantic classes leads to the term of auditory objects and how they are perceived. It is important to understand how humans create auditory objects from what they perceive to have control over which meanings are transported by certain stimuli. There is strong evidence that the grouping process of perception stimuli to form auditory objects is a non-spatial process, although spatiality remains important to direct the users attention [16]. This emphases the importance of semantics and taxonomy in the process of forming auditory representations for human-computer interaction. The *what* is the most important cue for forming distinguishable objects to which a user may direct its attention to. Finding classes of properties of auditory objects fitting the semantic classes extracted out of user interfaces is the ultimate goal and might be achievable through the approach stated.

4. MAPPING

As the next step in the depiction process the identified and classified UI entities need to be mapped onto sound events. This task can be divided into two sub-tasks: The mapping of the single entities and the composition of the whole user interface as an acoustic scene. Using spatial environments affects the way information

gets mapped into the auditory domain. The subsequent sections provide an overview over mapping techniques and discusses considerations for presenting sound in spatial environments.

4.1. Entity Mapping

Possible sound mappings range from simple background sound to complex speech representations. The aim of this task is to find working acoustic representations of user interface entities with the lowest level of user attention needed and the most flexibility for placing them in a virtual environment possible.

Furthermore, it is essential that this mapping process can be performed automatically and in real-time to be able to use this system with any applications embedded into the accessibility framework. It should not be necessary for interface designers to remodel the interface for auditory displays by hand. The mapping is entirely based on the interface description including structure and semantic information of the entities.

The least attention demanding way of presenting information with sound is non-speech ambient sound. "Audio Aura" [3] showed how such sounds can be used to provide low priority information to the user. Although it does not require a lot of user attention it is a very robust communication because of its omni-presence. It is not a question of the view angle whether one perceives that sound message or not.

More sophisticated sound events include auditory icons and earcons[17]. Auditory icons are small sound pieces out of our every day life which can be intuitively assigned to their meaning. Acoustic metaphors can represent, as their visual counterparts do, very much information by very little because they exploit common knowledge [18]. One of the most popular graphical metaphors, the trash can icon, is easily assigned to a function with which users can delete data with a time-limited recovering option. It is common knowledge how trash cans work and look like. Finding an auditory icon for a trash can is an easy task, but it is uncertain in most cases and subject of target user group investigations what common knowledge includes and which metaphors work.

Earcons are compositions of abstract, synthetisised motives to create auditory messages [17]. Motives are building blocks with the possibility to alter their properties in order to fit together. Hierarchical structures can be designed for earcon composition which are simple versions of iconic languages. To build an "earconic language" it is necessary to design a vocabulary of simple earcons and grammar rules for combining them. Visual iconic languages are widely used (e.g. Windows) and considered to be a powerful design methodology. In contrast to acoustic icons, earcons are not generally intuitive for the user, but must be learned. An "earconic language" will also increase memorability and learnability of earcons and therefore increase usability.

Non-speech sound is often advantageous in comparison to synthesised speech [19]. However, in some cases the use of speech is inevitable. It is the way of communication demanding the most attention from the user.

On basis of the information available about the UI entities, their semantic classification and context it is possible to develop transformations using the palette stated above. To consider cultural differences or simply for personal convenience a visual concept can be used to make the mappings more flexible. *Skins* are used recently in graphical user interfaces to change the appearance of the interface. In analogy to this method acoustic skins can be used to adjust the appearance of acoustic mappings.

4.2. Scene Modelling

Three dimensional, surrounding interfaces have the advantage of a much bigger display area than two dimensional screens or even one dimensional screenreaders. But they need to overcome the problem of concurring information and its interference. With both human senses, hearing and seeing, we have learned to navigate in our everyday three dimensional environment. However, visual techniques to present human-computer interfaces with concurrent visual information are more common than acoustic techniques. The two main reasons for this are: 1) the creation of two or three dimensional sensation is technically harder to realise in the audio domain. 2) Source segregation of concurrently presented information sources is a more complex process with the audio mode because of less focus capability. However, with high definition audio rendering and more knowledge about acoustic perception spatial audio displays proved to be usable and performing well [5, 20].

The segregation of sound sources in a virtual 3D environment is a necessary prerequisite of presenting information parallel to the user without confusion. Research has shown that sound source discrimination is a complex process. From the physiological point of view segregation depends on localisation cues like monaural spectrum changes or binaural sound wave differences [21]. But as stated in [22] sound source segregation is not only a question of localising the sources, it is heavily dependent on psychoacoustic effects. The content of the presented audio streams is crucial for the user to be able to segregate them. This has been shown for non-speech sounds [23], earcons [24] and speech sources [22]. Other psychoacoustic effects like informational masking [25] can additionally influence the presentation.

The task to model a scene out of the transformed user interface entities can be summarised by:

- Which UI entities can be presented concurrently?
- Which acoustic properties of UI entities can be altered to make them work concurrently?
- How can the UI entities be placed in the environment?
- What is the maximum load of the auditory display?
- How are the functional bindings between the UI entities depictable?

This paper proposes the development of algorithms to solve this depiction process in analogy to the problem of colouring countries on a map. UI entities ware mapped according to an acoustic skin to provide a common "hear & feel". However, they need to provide properties to alter for the scene modelling in order to be discriminable without decreasing their quality of mapping. After determining the placement of the entities an algorithm can solve the problem of which properties need to be altered to make neighbour regions discriminable (to colour the countries).

The placement of the user interface entities can be determined from the task model of the interface. Their structure and functional coherences can be used to choose from template arrangements like menus, desktops or content areas.

The question of the maximum load for an auditory display is very hard one to answer. It is hard to find robust metrics for the "amount" of information and furthermore it is individual and subject of training. A good candidate might be the level user attention needed for a UI entity (ranging from background sound to speech, see above). If limits are reached and no more entities can be depicted on the display, techniques like the acoustic lens or hiding entities in a group entity may be used.

4.3. Scene Description

From this point on a formal description of the audio scene is needed as the input for the audio rendering engine. Many standards include scene descriptions and source positioning. One of the most promising to become a widely accepted standard is MPEG-4. It is a multimedia streaming format including audio and video compression techniques as well as virtual scene descriptions [26].

A major drawback of MPEG-4 is its complexity. The MPEG committee therefore created MPEG-4 profiles which are subsets of the standard. For this system all audio related profiles (Speech, Synthesis, High Quality, Low Latency etc.) and the AudioBIFS (Audio Binary Format for Scene Description) are relevant. An overview over the capabilities of MPEG-4 for audio is given in [27].

An important feature of MPEG-4 for auditory displays is the definition of interaction techniques in the standard. BIFS distinguishes client side and server side interaction. Client side interaction involves user interaction which is handled entirely by the client player. All information needed is available at the client (e.g. rotation of the listener according to a head-tracking device). Server side interaction manipulates the content of the screen so that additional information from the server is needed (e.g. opening the door to another room). This method allows efficient coding of interaction needed for auditory displays.

MPEG-4 is the format of choice for modularising the proposed system. It is important to use a widely accepted standard to be able to provide a common interface to various audio rendering engines. The subsequent section deals with some rendering techniques which are capable of creating scenes described by MPEG-4 AudioBIFS.

5. RENDERING

The quality of the created virtual environment is an important factor for the performance of a spatial auditory display. As stated above, sound source segregation is made possible by the capability to localise sources on different positions. Recent developments in signal-processing and the increasing computational power available led to sophisticated sound field reproduction algorithms which can now be implemented in real-time. These techniques include (binaural) Higher Order Ambisonics (HOA) [28], Wave Field Synthesis (WFS)[29] and Vector Base Amplitude Panning(VBAP) [30].

The decision over which technique to use depends on the requirements on the output format. Aspects are:

- Is the form factor of the output device restricted?
- Is the possible impact on the surrounding acceptable?
- Does the user have to communicate with others while using the system?
- What are the minimal quality requirements for the application?

Based on the answers to these questions the devices may be chosen out of headphones, small loudspeaker arrays or mid to large scale venues.

Ambisonics is an encoding-decoding scheme decomposing a sound wave into its spherical harmonics and reproducing it in the listening point by the interference of loudspeaker signals. The three most important advantages of Ambisonics for auditory displays are: 1) The Ambisonics scheme can be decoded to binaural output as well as for large loudspeaker arrays [31, 32]. 2) For binaural output the number of sound sources has only minor effects on the needed processing power [31] and 3) Ambisonics features efficient rotation methods important for virtual environments with headtracking (an important cue for directional hearing [33]). This predestines Ambisonics for office solutions using headphones or the sonification of complex data in mid-scale venues.

Wave Field Synthesis reproduces plain sound waves by arrays of spherical emitters. Although the approaches are different, Ambisonics and WFS can be considered as equivalent [34]. Typical applications of WFS include large cinema halls, but also small devices for computer screens have been developed [35].

Vector Base Amplitude Panning is a vector based reformulation of simple amplitude panning methods. It allows the positioning of sound sources within a loudspeaker array placed around the listening point. A typical application was shown in the DIVA system, an interactive, audio-video virtual environment [36].

All those methods are capable of creating the sensation of directional sound and may be used to render information from a MPEG-4 stream. However, at the time being there is no MPEG-4 audio player available supporting such sophisticated sound reproduction techniques.

6. TARGET APPLICATIONS

Target applications are all computer interfaces where graphical output is less efficient, not available or not possible. This includes assistive technology for the visually impaired and the blind, mobile devices or sonification of complex data.

Visually impaired and blind people currently are at a significant disadvantage in accessing modern information technology. More computer interfaces are entirely based on and designed for graphical output. Currently available assistive technology for the people concerned are mainly Braille-lines and screen readers. Both sequential techniques with disadvantages in performance and comfort. A new effort towards spatial audio interfaces could lead to new ways of integrating the visually impaired and blind into our information society better by providing them equal access to computers [37] avoiding a "digital divide". First attempts into the proposed direction are made and have been proved as promising [38].

"Less is more when it comes to mobiles" [39]. Devices for mobile computing get smaller, small devices like mobile phones get smarter. Presenting a rich user interface with small display sizes is an increasing design problem. Graphical displays have high resolutions, but their form factors are very small and low usability is the result. Spatial audio interfaces are not bound to form factors. Their display size is independent from the size of the device. The trend to miniaturisation of mobile devices augmenting our every day life with information is at least playing in favour for auditory displays.

7. UTILITY AND USABILITY ASPECTS

For the system to be accepted by the user it is important to define what system acceptability means for human-computer interfaces. Figure 4 illustrates coherences.

For auditory displays to be considered as useful it must be proved that they are capable of representing user interfaces (utility) and that they are easy to use (usability). The usability metrics shown in figure 4 determine the aspects to consider when designing auditory displays:

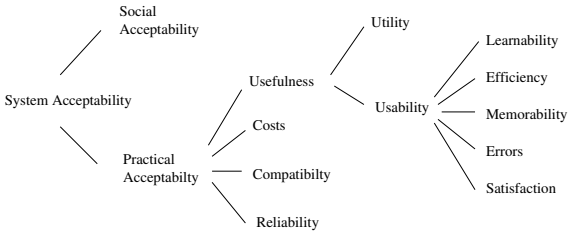

Figure 4: A model for system acceptability from [40]

Learnability is an issue for auditory displays for two reasons: 1) acoustic metaphors are less manifested and common knowledge is less widely spread as in the graphical domain. 2) Users are less used to the auditory mode as human-computer interaction. **Efficiency** is inevitable for the success of auditory displays. The common sequential techniques like screenreaders do have significant disadvantages in terms of performance and are therefore only used in assistive technology. It has shown that the mental overload is a potential problem in spatial auditory environments. Therefore, easy **Memorability** of the interface is important in the design to reduce the mental load. Content must be easily recognisable to avoid confusion. **Errors** definitely reduce the usability of a system. Ambiguous localisation of sound or unexpected event rendering can make the system hard to use. To achieve user **Satisfaction** it is important to make the user feel comfortable in the virtual environment. This addresses mainly the sound design. To avoid user annoyance it is proposed to implement techniques like acoustic skins so that users may have the chance to adjust the output to their convenience.

A necessary prerequisite to satisfy these usability aspects is to conduct target user research. Knowing the user is inevitable to design good interfaces. Especially when designing for minor groups like the visually impaired and the blind it is important to consider their special needs and abilities. The proposed depiction process presented here is intended to be formalised and modularised. Different user requirements can therefore be considered by reformulating transformation rules or replacing rendering tools.

There are a lot of usability engineering methods available for the design and evaluation of graphical user interfaces. Further investigations will show which of them might be applicable and useful for auditory displays.

8. CONCLUSION

This paper proposed a generic depiction process for computer user interfaces into the auditory domain using spatial audio. The intention is not to focus on specific problem solutions, but to look at the problem in a bigger scope. It might be valuable for the community to take a step back and develop a generic framework in which all the knowledge about the auditory perception and human-computer interaction can be integrated.

A research project with this scope is currently being organised the Institute of Electronic Music and Acoustics[2].

[2] `http://iem.at`

9. REFERENCES

[1] C. Schmandt M. Kobayashi, "Dynamic soundscape: Mapping time to space for audio browsing," in *ACM/SIGCHI 97 Proceedings*, Los Angeles, CA, March 22–27 1997, ACM Conference on Human Factors in Computing Systems, pp. 194–201.

[2] C. Schmandt, "Audio hallway: A virtual acoustic environment for browsing," in *Proc. ACM Conference of Computer Human Interactions*. April 18–23 1998, pp. 163–170, ACM Press, Los Angeles, California, USA.

[3] R. Want E. D. Mynatt, M. Back, "Designing audio aura," in *Proc. ACM Conference of Computer Human Interactions*. 1998, pp. 566–573, ACM Press, Los Angeles, California, USA.

[4] A. Härmä et.al, "Techniques and applications of wearable augmented reality audio," in *Convention Papers*, Amsterdam, Netherlands, March 22–25 2003, Audio Engineering Society, 114th Convention.

[5] S.A. McGookin A. Walker, S. Brewster, "Diary in the sky: A spatial audio display for a mobile calendar," in *IHM-HCI Proceedings*, Lille, France, September 10–14 2001, Interaction Homme-Machine – Human Computer Interaction, IHM-HCI, pp. 531–540.

[6] H. J. P. Timmermans M. K. D. Coomans, "Towards a taxonomy of virtual reality user interfaces," in *Proceedings of the International Conference on Information Visualisation*, London, 27–29 August 1997.

[7] E. Mynatt, *Transforming Graphical Interfaces into Auditory Interfaces*, Ph.D. thesis, Georgia Institute of Technology, August 17 1995.

[8] "The gnome accessibility project," `http://developer.gnome.org/projects/gap/`, 2004.

[9] GNOME `http://www.gnome.org`, *Assistive Technology Service Provider Interface (AT-SPI) C Bindings API Reference*, 2004.

[10] "The kde accessibility project," `http://accessibility.kde.org/`, 2004.

[11] "Sun microsystems accessibility program," `http://www.sun.com/access/`, 2004.

[12] "Microsoft accessibility," `http://www.microsoft.com/enable/`, 2004.

[13] Microsoft, *MS Active Accessibility User Interface Element Reference*, 2004, `http://msdn.microsoft.com/library/default.asp?url=/library/en-us/msaa/m%saapndx_2a05.asp`.

[14] N. Deakin, "Xul tutorial," `http://www.xulplanet.com/tutorials/xultu/`, Dec. 2003.

[15] "Xml markup languages for user interface definition," Tech. Rep., OASIS, Organization for the Advancement of Structured Information Standards, 2004, `http://xml.coverpages.org/userInterfaceXML.html`.

[16] D. van Valkenburg M. Kubovy, "Auditory and visual objects," *Cognition*, vol. 80, no. 1-2, pp. 97–126, June 2001.

[17] R. M. Greenberg M. M. Blattner, D. A. Sumikawa, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[18] W. K. Edwards E. D. Mynatt, *Extraordinary Human-Computer Interaction*, chapter Metaphors for Nonvisual Computing, Cambridge University Press, 1995.

[19] S. A. Brewster, *Providing a structured method for integrating non-speech audio into human-computer interfaces*, Ph.D. thesis, University of York, UK, 1994, http://www.dcs.gla.ac.uk/~stephen/papers/Brewster_thesis.pdf.

[20] T. Lokki et.al., "Creating interactive virtual auditory environments," *IEEE Computer Graphics and Applications, special issue "Virtual Worlds, Real Sounds*, vol. 22, no. 4, pp. 49–57, July/August 2002, Electronic publication http://www.computer.org/cga/.

[21] S. Charlile V. Best, A. van Schaik, "Two point discrimination in auditory displays," in *ICAD Proceedings*, Boston, MA, USA, July 6–9 2003, International Community for Auditory Display, pp. 17–20.

[22] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35–50, 1992.

[23] J.Hiipakka G. Lorho, J Marila, "Feasability of multiple non-speech sounds presentation using headphones," in *ICAD Proceedings*. ICAD: International Conference on Auditory Display, July–August 2001, Espoo, Finland.

[24] S. Brewster D. K. McGookin, "An investigation into the identification of concurrently presented earcons," in *ICAD Proceedings*, Boston, MA, USA, July 6–9 2003, International Community for Auditory Display, pp. 42–46.

[25] R. A. Lufti E. L. Oh, "Informational masking by everyday sounds," *Journal of the Acoustical Society of America*, vol. 6, no. 106, pp. 3521–3528, 1999.

[26] R. Koenen, "Mpeg-4 overview," Tech. Rep., International Orgranisation for Standardisation, May 2002, http://www.m4if.org/.

[27] J. Huopaniemi E. D. Scheirer, R. Vaananen, "Audiobifs: Describing audio scenes with the mpeg-4 multimedia standard," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 237–250, 1999.

[28] J. S. Bamford, "An analysis of ambisonic sound systems of first and second order," M.S. thesis, University of Waterloo, http://audiolab.uwaterloo.ca/~jeffb/thesis/thesis.html, 1995.

[29] E. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, TU Delft, 1998.

[30] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

[31] M. Noisternig et al., "3d binaural sound reproduction using a virtual ambisonic approach," in *VECIMS Proceedings*, Lugano, Switzerland, 27–29 July 2003, International Symposium on Virtual Environments, Human-Computer Interfaces, and Measurement Systems, IEEE.

[32] J. M. Zmölnig et. al., "The iem-cube," in *ICAD Proceedings*, Boston, USA, July 6–9 2003, Internation Conference on Auditory Display.

[33] P. Minnaar et.al, "The importance of head movements for binaural room synthesis," in *ICAD Proceedings*. ICAD: International Conference on Auditory Display, July–August 2001, Espoo, Finland.

[34] M. A. Poletti, "A unified theory of horizontal holographic sound systems," *Journal of the Audio Engineering Society*, vol. 48, no. 12, pp. 1155–1182, December 2000.

[35] M. Strauss et. al., "A spatial audio interface for desktop applications," in *AES Proceedings, International Conference on Multichannel Audio*, Banff, Canada, June 26–28 2003, AES: Audio Engineering Society.

[36] T. Lokki V. Pulkki, "Creating auditory displays with multiple loudspeakers using vbap: A case study with diva project," in *ICAD Proceedings*, Glasgow, UK, November 1–4 1998, International Community for Auditory Display.

[37] A. I. Tew D. T. Murphy, M. C. Kelly, "3d audio in the 21^{st} century," in *Proc. 8^{th} International Conference, Computers Helping People with Special Needs, ICCHP 2002*, July 2002, pp. 562–564, Linz, Austria.

[38] C. Frauenberger, "Three-dimensional audio interfaces for the blind," M.S. thesis, Graz University of Technology, Department of Communications and Wave Propagation, 2003, http://iem.at/Members/frauenberger/Publications/thesis/html/.

[39] S. Brewster A. Walker, "Spatial audio in small screen device displays," *Personal Technologies*, vol. 4, no. 2, pp. 144–154, 2000.

[40] J. Nielsen, *Usability Engineering*, Academic Press, London, 1993, ISBN 0125184050.