DESIGN PROCESS FOR AUDITORY INTERFACES

Sylvain Daudé

CLIPS-IMAG Equipe IIHM
BP 53
38041 Grenoble cedex 9, France
`sylvain.daude@imag.fr`

Laurence Nigay

CLIPS-IMAG Equipe IIHM
BP 53
38041 Grenoble cedex 9, France
`laurence.nigay@imag.fr`

ABSTRACT

In this paper we present a unified framework for the design of auditory interfaces. We describe the steps of the sonification process and their parameters. The process is modeled as a sequence of transformation functions from the data to be conveyed to the produced sounds. The usefulness of the framework for classifying existing auditory interfaces and for designing new ones is also discussed.

1. INTRODUCTION

The design of auditory interfaces involves several research fields, including human-computer interaction, psychoacoustics, digital signal processing and information visualization. This interlinking location raises the problem of a unified design process that would integrate results from various fields. In this article we present the abstraction levels and design parameters involved in each step of a unified design framework, with an emphasis on contextual information. We then compare our approach to related studies and underline the contributions of our framework for the classification and the design of auditory interfaces.

2. SONIFICATION PROCESS

Our framework is inspired by Ed Chi's data pipeline for information visualization [1] presented in Figure 1. We also present our sonification process (on the right in Figure 1).

2.1. From Data to Data View : Data Transformation (F1)

F1 aims at extracting useful information from the original data. This step is common to both the visualization and the sonification processes.

F1 depends on both data semantics and users' tasks. For example, the AROMA system [3] jams the sound signal coming from distant users talk (data semantics) while keeping enough information for speaker identification (a users' task). The design of F1 should therefore include an analysis of (i) data, (ii) useful information (data semantics) and (iii) users' tasks. This analysis can rely on existing task and data classifications [4][5]. For example, Barrass's TaDa system [4] provides a systematic method for analyzing these characteristics from usage scenarios.

2.2. From Data View to Abstract Sound Space: Sonification Transformation (F2)

2.2.1. *Abstract sound space*

In the visualization process presented in Figure 1, the data view is transformed into an Euclidean space, then displayed on screen according to visual cues including color, size and shape. In the sonification process, F2 defines, from the data view, an Euclidean space generally centered on the user's position or reduced to a point if the sonification is not spatialized. F2 involves two successive transformations: (1) the representation of each single element by an abstract sound (auditory cues including timbre, intensity varying in time), and (2) the coordination of the corresponding sounds.

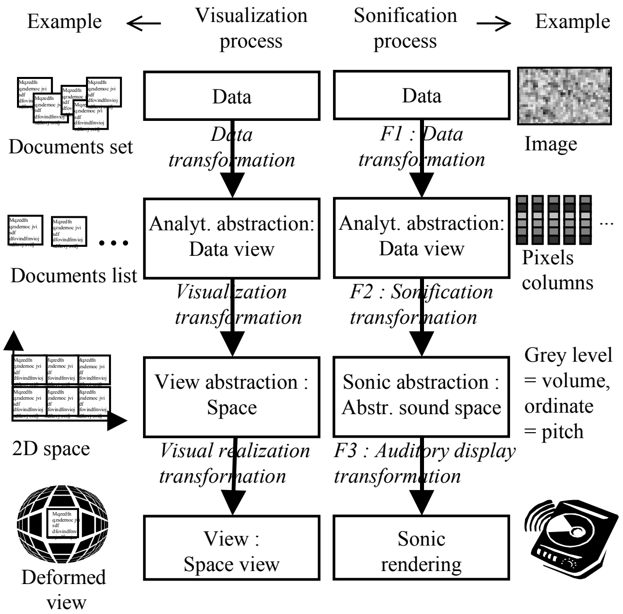

Figure 1. *Chi's data pipeline for information visualization (left) [1] versus our framework (right). Example on the right inspired from [2].*

2.2.2. *Representation of a single element*

F2 first represents each single element of a data set by an abstract sound. To do so, F2 involves three independent mapping functions namely (i) mapping of semantics, (ii) mapping of structures and (iii) mapping of values.

(i) From the perceived sound, the user must be able to interpret the corresponding information. Therefore we need to study the relation between the information semantics and the sound semantics:

- First, the relation between the two semantics is characterized by their "resemblance", which can be either *symbolic* -the sound does not resemble the data-, *indexical* -the sound is causally connected to the data- or *iconic* -the sound resembles the data- [4]. A physical model of a sound source, in which the information to be represented is a stimulus exciting the sound source [6], is one example of an indexical relation between information and sound.
- Second, the relation between the two semantics can be arbitrary or not. A non-arbitrary mapping relies on an already existing system of meaning. Thus, according to Bernsen [7], linguistic mappings are not arbitrary. An example of linguistic mapping is Blattner's earcons [8]: one element is sonified by a sequence of sounds organized according to a grammar. On the other hand, a direct mapping of data values to acoustical cues [10] is arbitrary.

It is important to note that degree of resemblance and arbitrary nature are two interdependent design parameters characterizing the relation between information semantics and sound semantics. Indexical and iconic relations cannot be arbitrary, whereas a symbolic relation can be either arbitrary or not. An example of a symbolic arbitrary mapping is a system where a mail delivery is associated with a door slam; on the other hand, earcons correspond to a symbolic non-arbitrary relation between the two semantics.

Moreover, F2 associates the contents of the information with the contents of the sound. These contents can be broken up into structure (ii) and value (iii), implying two additional sonification levels.

(ii) An example of structure mapping is presented in Figure 1: F2 consists of presenting a picture by going through its columns. Each pixel of a column is sonified according to its gray level and its position in the column. The gray level and the ordinate of a pixel are two ordered graphical cues that are represented respectively by an intensity and a pitch (two ordered auditory cues). In addition, if the data view contains a referential, such as bounds or an average value, it may be useful to sonify them: this need has been identified for data mining [11]. For example in [12], a progression bar is sonified by an organ sound whose pitch approaches the pitch of a « reference » guitar played at regular intervals.

(iii) Finally, value mapping consists of representing values using auditory cues. An abstract sound is firstly characterized by its duration, which defines a temporal window during which a set of auditory cues including timbre, pitch, volume and position can vary. The design choice of these cues is based on perceptive (a) and cognitive (b) parameters. (a) Considering sound perceptive properties allows the designer to predict the faithfulness of the representation according to the perceived precision, importance and cross-influences of the produced sounds with respect to the initial information. SoundChooser [4] is one example of a tool where sounds are organized according to perception principles. Guidelines for designing a perceptively relevant auditory interface can be found in [4]: for example, alarm and warning sounds should have frequencies between 500 and 3000 Hz. (b) An important sound cognitive parameter is its nature, which can be either everyday, vocal, musical, verbal or synthetic [4]. Everyday and vocal sounds relate to an ecological approach of sonification and correspond to Gaver's auditory icons [13]. When correctly designed, auditory icons provide enjoyable and easily interpretable sounds [13], although learning their meaning is as slow as for earcons [14]. Musical sounds are enjoyable and take advantage of the implicit semantics of music tonality and dynamics [15]. Verbal sounds are precise and direct but limited to small sets of data [5]. Finally, with synthetic sounds, there is no analogy with the real world. For example Sheppard-Risset sounds, which seem to go up or down indefinitely, were used for audio feedback when scrolling and for monitoring the progress of long system operations [6]. A wide range of synthetic sounds can be generated by low cost algorithms [16] or by wave or time-frequency editors.

Sound position is a special cue as its semantics can be independent from the one of the sound itself. In particular, the spatial distribution of sounds can rely on an analogy with a real world spatial phenomenon. This can be a sonic phenomenon, like the acoustic environment of a cathedral. On the other hand, the *Diary in the Sky* system [17] relies on an analogy with a non-sonic spatial phenomenon. It displays a sequence of appointments by means of spatial sounds: each scheduled appointment is linked to a direction (ex. in front of the user for noon, to the right for three o'clock etc). Indeed, the spatial distribution of sounds relies on the analogy of a clock surrounding the user.

2.2.3. *Representation of a set of data*

Coordinating the representation of each element (i.e. an abstract sound) temporally and spatially is the second sonification step (F2). Data (abstract sounds) can either be presented simultaneously or alternately. Concatenating sounds makes perception of the data relationships easier, whereas their temporal separation (silence between two sounds) facilitates the identification of each single element. If F2 involves spatial sounds, data may be displayed either from one position or from several positions. Sound distribution in space allows a better discrimination between elements, a better memorization [17] as well as a greater feeling of immersion. Nevertheless sound distribution in space may create ruptures in the perception of a set of data. A key theory for evaluating and predicting the perceptive grouping of sounds into streams is the Gestalt theory [18]. An approach which implicitly integrates Gestalt theory principles as well as user's analysis experience consists of organizing sounds according to a musical structure [15].

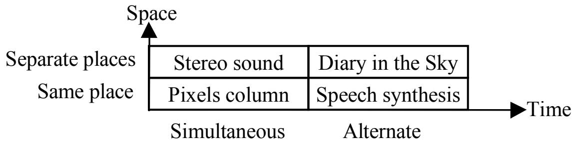

Figure 2. *Classification of four systems according to the temporal and spatial representation of data.*

In Figure 2, four systems are classified according to the temporal and spatial distribution of sounds. In the example of Figure 1, the sonification of all the pixels in a column are played at the same time and from the same position. On the other hand, a stereo sound is made of two sounds played at the same time from different positions. The *Diary in the Sky* system presents appointments one after the other, from separate positions. Finally, speech synthesis, but also auditory icons and Blattner's earcons present data sequentially from a unique position.

2.3. From Abstract Sound Space to Sonic Rendering: Auditory Display Transformation (F3)

Auditory Display Transformation (F3) consists of three steps: (i) computation of device parameters from sound space parameters, (ii) computation by the sound engine of the sound signal according to device parameters and (iii) display of the signal on a physical device [4]. The only step that the designer can control is step (i): computing device parameters from the abstract sound space parameters is costly. Simplified acoustic laws [19] or sound compressions are used to reduce the computation. The sound engine processing power (ii) as well as the physical device properties (iii) are not controllable by the designer but act as constraints during the design process.

We describe a physical device by six characteristics: (1) its sonic precision with relation to the data to be displayed (2) its degree of spatiality (mono, stereo, multi-channel 2D or 3D), (3) the geometrical distribution and properties of its loudspeakers, (4) its degree of sharing i.e. the number of users able to perceive the information at the same time, (5) the liberty of movement it allows to the user and (6) its degree of immersion, indicating if the device isolates the user from the real world. Thus, the Soundbeam Neckset [20] is characterized by a good precision for voice (1) and a stereo sound (2) displayed through two directional loudspeakers located on the user's shoulders (3). The Soundbeam Neckset is used by one user at a time (4) and allows a complete liberty of movement because it is portative (5); finally its immersion degree is low (6). Knowledge of the physical device characteristics can help to optimize the device parameters computation (i). For example, knowing the position of two loudspeakers allows us to reduce their interference by a cross-correlation method and thus to produce an accurate spatialization.

3. INFLUENCE OF CONTEXTUAL INFORMATION

In this section we emphasize the fact that contextual information influences each step of our design process. Our definition of context includes user's preferences, platform resources, physical and social environment of the user.

3.1. Data Transformation (F1)

The usefulness of an information depends on the context in which it is used. The importance of considering the user's tasks for F1 in a static way has already been underlined. However, the importance of a task dynamically varies. For instance, if a task becomes secondary, F1 can be modified so that the data view contains only data changes and not all the data values, in order to free the user's attention. Social and physical context also influences the information to be displayed. For example confidential information should not be displayed when the user is not alone (social context). Likewise in a noisy environment (physical context), F1 should filter the information in order to only present the most important information, such as alarms and warnings.

3.2. Sonification Transformation (F2)

Sonification depends on several context-dependent properties including understandability, noticeability and non-disturbance.

Understandability is directly linked to user's knowledge and preferences. According to Vickers [15], musical sonifications are a solution to individual disparities because music is a universal communication media, since occidental music has spread all over the world.

Taking into account the physical and social context may lead to a more noticeable and less disturbing sonification. For instance, Ronkainen [21] proposes to automatically change the bell of a mobile phone to make it more noticeable and in harmony with the user's sonic environment. In another system notification sounds are mixed with the music the user is currently listening to [22]. This latter solution aims at integrating the sonic environment of the user into the abstract sound space.

3.3. Auditory Display Transformation (F3)

Contextual information can help choose the most appropriate rendering technique and physical device. For example, information about the acoustics of the room (physical context) or about the user (user context) can increase the accuracy of spatialization on headphones by respectively applying:

- BRIR filters simulating that the sounds come from the room around the user,
- individualized HRTF: this requires measuring the listener's HRTF.

Moreover, if the computing resources of the platform (the platform being part of the context) are low, low cost algorithms can be chosen. For example, a real-time scheduling strategy for degrading the rendering of the least important sound sources in a Virtual Environment is presented in [23]. Finally, the physical device can be dynamically chosen according to the location of the user, as suggested in the Audio Aura system [24].

4. RELATED WORK AND CONTRIBUTION

4.1. Classification

Several classification schemes have been proposed to differentiate between sonification systems [9]. In [4] a classification scheme that covers all the steps of our design process is presented. The different approaches identified in [4] are: (i) syntactic, with an emphasis on the temporal organization of sounds, (ii) semantic, concerned with the meaning of the presented information, (iii) pragmatic, focusing on the differences of sound « forms », (iv) perceptual, (v) task-oriented, (vi) connotation, concerned with side effects of sound interpretation and (vii) device, focusing on the variations of perceived sounds across sound devices. Figure 3 shows that the sonification approaches identified in [4] can be classified into our framework and cover the entire sonification design process.

Approach	**Position in sonification process**
syntactic	F2 (temporal coordination)
semantic	F1, F2 (analogy), user context
pragmatic	F2 (perception of data sets)
perceptual	F1 (characterization of useful information), F2
task-oriented	F1 (task analysis), interaction context
connotation	User context
device	F3 (sound engine and display device characteristics), platform context

Figure 3. *Sonification approaches [4] according to the steps of our design process.*

However, most existing classifications refer to the Sonification Transformation F2 only. Thus, based on design parameters of F2 only, the most common classification includes four classes, namely artificial stream mapping, Blattner's earcons [8], Gaver's auditory icons [13] and speech synthesis. Indeed earcons correspond to a sonification transformation (F2) based on a grammar, while the three other ones differ by the nature of the manipulated sounds.

Moreover systems previously classed in one category can be distinguished using our framework. For example, mobile phone bells depending (or not) on caller identity, Blattner's earcons, Brewster's serial and parallel earcons [5] were gathered in a single generic « earcons » category. Phone bells differ from other earcons as they are not organized according to a grammar. Moreover various mobile phone bells differ from each other in the information to be sonified i.e. (« X is calling » for the caller-dependent bell, « Somebody is calling » for the caller-independent one). Blattner's and Brewster's earcons differ in the nature of the sounds (synthetic versus musical). Finally, serial and parallel earcons differ in the coordination of data sets in time (alternatively versus simultaneously).

4.2. Design

Our framework organizes the process into several steps and identifies design parameters. It serves therefore as a guide for the design of auditory interfaces. Our framework structures the design and concentrates force (of reasoning) in the appropriate area; this does not mean that there is no role for the artisan and no element of skill and judgement involved during the design.

In [4], a design approach of auditory interfaces differs from our analytical step by step approach by considering existing solutions in similar systems. Indeed Barrass's solution [4] consists of comparing the requirements of a new system to existing system requirements stored in a database [4]. The solution of an existing system is then adapted to the new system. A database of existing sonification mappings was built and put online [25]. It would be interesting to further investigate the complementarity of the two design approaches. For example our framework may help identify the parameters to be modified in order to adapt an existing solution.

5. DIRECTION FOR FUTURE RESEARCH

We aim at integrating in a unified design framework, results from various fields (ergonomics, human-computer interaction, psycho-acoustics, digital signal processing and information visualization). This effort must be pursued and new design parameters must be integrated in our framework.

Based on our design framework and parameters, one research avenue is to establish design heuristics or design patterns. One example of design pattern could be: in the case of a background monitoring task, sounds must if possible be short and integrated into user's acoustic environment.

Finally another research avenue is to establish links between the sonification design process and the software architecture of an auditory interface. As done for visualization, can we define a pipeline software architecture based on the steps of our framework?

6. REFERENCES

[1] E. Chi, J. Riedl, "An Operator Interaction Framework for Visualization Systems", Proc. Info Vis '98.

[2] P.B.L. Meijer, "An Experimental System for Auditory Image Representations", IEEE Transactions on Biomedical Engineering, Vol. 39 No 2, 1992.

[3] E. Pedersen, T. Sokoler, "AROMA: abstract representation of presence supporting mutual awareness", Proc. CHI'97.

[4] S. Barrass, "Auditory information design", PhD thesis, Australian National University, 1997.

[5] S.A. Brewster, "Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces", PhD thesis, University of York, 1997.

[6] S. Conversy, "Conception d'Icônes Auditives Paramétrées pour les Interfaces Homme-Machine", PhD thesis, Université de Paris-Sud, 2000.

[7] O. Bernsen, "Foundations of multimodal representations. A taxonomy of representational modalities", Interacting with Computers, Vol. 6, No 4, 1994.

[8] M. Blattner, D. Sumikawa, Greenberg, "Earcons, icons : their structure, common design principles", Proc. HCI'89.

[9] S. Barrass, G. Kramer, "Using sonification", Multimedia Systems, Vol. 7, No 1, 1999.

[10] C.M. Wilson, S.K. Lodha, "Listen: A Data Sonification Toolkit", Proc. ICAD'96.

[11] C. Daassi, M.C. Fauvet, L. Nigay, "Multiple Visual Representation of Temporal Data", Proc. DEXA'02.

[12] M. Crease, S. Brewster, "Making progress with sounds - the design, evaluation of an audio progress bar", Proc. ICAD'98.

[13] W.W. Gaver, "Using, Creating Auditory Icons", in Auditory Display: Sonification, Audification, Auditory Interfaces. Reading, MA: Addison Wesley, 1994.

[14] P. Lucas, "An evaluation of the communicative ability of auditory icons, earcons", Proc. ICAD'94.

[15] P. Vickers, "Musical Program Auralisation System to Study the Effects on Debugging Tasks as Performed by Novice Pascal Programmers", PhD thesis, Loughborough University, 2000.

[16] E.R. Miranda, "Computer Sound Design - Synthesis Techniques, Programming - 2nd edition", Focal Press, 2002.

[17] A. Walker, S.A. Brewster, D. McGookin, A. Ng, "Diary in the sky : a spatial audio display for a mobile calendar", Proc. CHI'01.

[18] A.S. Bregman, "Auditory Scene Analysis", Cambridge, MA : MIT Press, 1990.

[19] N. Tsingos, "A versatile software architecture for virtual audio simulations", Proc. ICAD'01.

[20] N. Sawhney, C. Schmandt, "Design of Spatialized Audio in Nomadic Environments", Proc. ICAD'97.

[21] S. Ronkainen, "Earcons In Motion - Defining Language for an Intelligent Mobile Device", Proc. ICAD'01.

[22] M. Barra, T. Cillo, A. De Santis, T. Matlock, U.F. Petrillo, A. Negro, V. Scarano, P.P. Maglio, "Personal webmelody : customized sonification of web servers", Proc. ICAD'01.

[23] H. Fouad, J.K. Hahn, J.A. Ballas, "Synthesis of Complex Sonic Environments", Proc. ICAD'97.

[24] E.D. Mynatt, M. Back, R. Want, M. Baer, J.B. Ellis, "Designing Audio Aura", Proc. CHI'98.

[25] B.N. Walker, D.M. Lane, "Sonification mappings database on the web", Proc. ICAD'01.