Human-Computer Interaction - INTERACT '99 M. Angela Sasse and Chris Johnson (Editors) Published by IOS Press, @ IFIP TC. 13, 1999

A Principled Design Methodology for Auditory Interaction

Evangelos N. Mitsopoulos & Alistair D.N. Edwards

Department of Computer Science, University of York, York YO1O 5DD, UK.

enm@cs.york.ac.uk

Abstract: When the visual channel of communication is unavailable, non-visual user interfaces must be developed. The proposed methodology aims at the specification of auditory representations for interactive tasks. It consists of three interrelated specification levels. Information and supported tasks are specified in abstract terms at the conceptual level. The structure of the auditory scene is specified at the structural level in terms of auditory streams. The physical dimensions of sound are defined at the implementation level. The design guidelines have been validated by the experimental evaluation of a list of auditory checkbox widgets.

Keywords: extra-ordinary human-computer interaction, auditory display, auditory widgets, blind user.

1 Introduction

Sight is a very powerful means of communication. It is used in everyday communications through diagrams, pictures, gestures and many other forms; it is also the basis of most human-computer interaction. When that channel of communication is unavailable because the person is blind or the environmental conditions are disabling, then non-visual alternatives must be developed.

Numerous methodologies for the specification of user interfaces exist and whether yet another one has to contribute anything of value is indeed the first question to be asked. For instance, consider (Foley et al., 1990). Their framework essentially consists of a medium-independent level of meaning (divided into conceptual and functional design) and two medium dependent levels, the syntactic and lexical design. Although their focus is mostly on graphical user interfaces (GUIs), there is no reason, in principle, why their framework could not be applied to the specification of auditory interfaces. In fact, the methodology presented in this paper consists of three levels of specification which are conceptually similar to those of Foley et al. Despite this similarity, the two methodologies have very different specification objectives.

Visual experience accumulated with every-day communications can be invaluable to GUI design because, combined with a set of empirical guidelines on visual design, it provides the basis for designing visual representations. Assuming that a suitable visual representation can be designed allows the next step forward, that is, the specification of that representation in the user interface. The specification must capture the "... 2D and 3D layout of a display, as well as any temporal variation in the form of the display" - Foley et al.'s (1990) syntactic design. The specification is also restricted at a lexical level by the available system primitives, such as the display primitives of a graphics subroutine package.

However, designers probably have considerably less experience with auditory information and, most of the time, no auditory representations exist to serve as paradigms. Despite the considerable volume of literature on auditory perception and attention, there are just a few guidelines accessible to the auditory interface designer. Even worse, designers might inadvertently apply visual design guidelines, achieving adverse results because of the inherent differences in sight and hearing.

Hence, in the case of auditory interfaces, it is most important to provide guidelines for the design of auditory representations in the first place. Given these guidelines, one may then proceed to address the issue of interface specification. Auditory design guidelines, derived from theories on auditory perception and attention, are the objective of our methodology. These guidelines are organized in a framework similar to Foley and colleagues' to aid a structured design approach. However, our methodology has to satisfy user constraints (perceptual and attentive) rather than system constraints (imposed by the available hardware and software). In this respect, we believe that it represents a valuable contribution to the field of auditory interfaces and, more generally, to multimodal interface design.

1.1 Related Work

A significant amount of research effort has been expended in recent years on the problem of how to substitute the visual communication of the GUI so that it can be used by blind people - see, for instance, (Edwards, 1989; Mynatt & Weber, 1994; Weber, 1993). Most approaches are based on translation of the surface visual representation to a non-visual equivalent. The clearest example of translating at the *surface* level is the GUIB project (Weber, 1993). However, as illustrated by (Edwards, 1995), visual representations naively translated to an auditory form can become very difficult to use. The weak point of surface transformations is their failure to acknowledge the inherent differences between the perception of auditory and visual information. To design an efficient auditory representation, be it either a transformation of a visual representation or a novel form of information ' it is essential to take into account some fundamental principles of auditory perception and attention. This is also the case for transforming visual to haptic representations (Kurze, 1997).

There is a linkage to the proposed methodology and the general notion of auditory earcons. Earcons have been designed not only for widgets such as buttons, menus or progress bars but also to convey navigation cues for hierarchical structures (Brewster, 1998). Our methodology is not confined to the design of auditory widgets, either. It is intended to be applicable to the design of *all* aspects of auditory user interfaces. For instance, it could be used to design an interactive presentation of charts to blind users.

Nevertheless, apart from the relationships between the methodology and earcons, there are some important distinctions, too. Earcons have mostly been designed as 'sonic enhancements' to the visual interaction. This methodology primarily address the issue of substituting the visual presentation with a non-visual one, an issue fundamentally different than enhancing visual interaction. More importantly, the design of the 'soundscape' is theory-driven relying on theories of auditory perception and attention. This contrasts with the design of earcons, which is based on empirical guidelines. The theoretical perspective of the methodology provides the means to systematically drive the designer's intuition, to assess the suitability of the resulting design, and to eliminate a considerable number of inappropriate alternatives prior to any experimental evaluation.

The objective of the research effort discussed herein is to develop a more principled basis for the design of non-visual interfaces. Widgets are used as exemplars since: they are essential components of GUIs; their simplicity lends themselves to a compact specification and they clearly capture interaction issues such as ear-hand coordination and auditory-haptic multi-modal integration. However, they only serve to illustrate the fundamental issues involved in the methodology; we understand them to be suitable exemplars rather than the most important design issue in non-visual interaction. The focus of this paper is on the specification of the auditory part of nonvisual widgets and, in particular, the design for interactive tasks where users control the presentation of information depending on the task in hand.

1.2 Auditory Representations

In designing an auditory alternative to a visual representation it is necessary to take into account the inherent differences between auditory and visual information. One of the most important differences is the dynamic character of sound. Visual interfaces are mostly static (although animation is becoming increasingly important). Users can inspect any part of the display at their convenience. A computer-based example is the visual widget in Figure 1.

Figure 1: A visual 'widget' with a number of components, including four checkboxes.

It is a property of this representation that the user can view it in at least two ways. It is possible to *glance* at it to quickly perceive patterns. Based on this information, the user can accomplish a number of tasks such as to identify that several checkboxes are checked (but not all of them). However, the user might choose (depending on the task in hand) a different approach to this item, and to *interact* with it toggling the value of a particular effect or browsing the list of available effects, for instance. All information necessary to the performance of both groups of tasks is accommodated in a single external representation.

This is not the case with auditory displays. Because of the dynamic character of sound and the short-term memory limitations imposed by the serial presentation of auditory information over time, it is difficult to accommodate all tasks into a single representation. Different rates of presentation are required for each group of tasks. Extracting information at a glance should use the fastest possible rate, whereas the rate of interactive tasks must be considerably slower and, most importantly, the user should have control of the information flow.

The distinction between two groups of tasks (die first based on fast rates of presentation and the second involving exploring and interacting with the information at much slower rates) extends to the relevant psychological theories, too. For instance, Auditory Scene Analysis (Bregman, 1990) deals mostly with the organization imposed on the auditory scene by perceptual processes at fast rates of presentation. Cognitive issues, such as selective attention or memory, become more prominent at slower rates of presentation. Issues relevant to fast rates of presentation have been discussed in Mitsopoulos & Edwards (1997; 1998). This paper deals with the design issues of auditory representations for interactive tasks. The methodology as a whole is outlined in Mitsopoulos & Edwards (to appear).

2 Methodology

The design methodology consists of three levels of specification, the conceptual, structural and implementation level. The reasoning for partitioning the specification process in three distinct levels has been discussed in detail in (Mitsopoulos & Edwards, 1998). In the following sections, each level will be illustrated using the example of the checkbox widget in Figure 1.

2.1 Conceptual Level

The set of tasks associated with the widget under design as well as the (abstract) information necessary to perform them is specified at this level. For the checkbox example, these tasks would include: identification of the state of the checkbox column at a glance (whether all or most of the checkboxes are checked or unchecked); browsing of available effects ('Bold', 'Italic', 'Shadow', 'Underline'); identifying and/or toggling of the value of an effect etc. The first task is relevant to fast presentation and has been considered in details in (Mitsopoulos & Edwards, 1998). We will focus our attention on the other two interactive tasks.

The necessary information to perform the related tasks can be defined in terms of abstract dimensions which can be of non-final, ordinal, interval, or ratio type (Zhang, 1996). For example, because the two states of a checkbox (checked, unchecked) are distinct but not ordered they can be represented by two values along a nominal dimension. When physical dimensions of sound are selected to convey the information specified at the conceptual level, they should be of the same scale as their abstract counterparts so that they convey no more or less information. Timbre is a nominal dimension and it would be appropriate to represent the state of a checkbox. Pitch (on its own) would be a less appropriate dimension since it is ordinal and the user might erroneously deduce that the two states are ordered (for example, low - high).

Use of an interaction device introduces a number of navigation tasks which require additional information. For example, for a user browsing the list of checkboxes using the keyboard cursor keys, it is necessary to provide information about whether either end of the list has been reached. However, the keyboard does not provide this information. The 'missing' information is defined in conceptual terms so that it can be accommodated in the auditory representation. For instance, another abstract dimension would be the position relative to either end of the list, which is ordinal and has three values: first item in the list, last item, elsewhere in the list.

Moreover, ear-hand coordination may pose additional requirements (such as advance auditory feedback) which affect conceptual specification. For example, it is necessary to warn the user in advance when approaching a target such as either end of the list; otherwise, the user is very likely to overrun the ends at fast browsing rates. The information about position would include two more values: next-to-first item and next-to-last item.

Abstract dimensions are the constituents of semantic entities. A checkbox widget is an entity that has three dimensions, two nominal ones representing its *state* and its *label* ('Bold', for instance) and an ordinal one for its *position* in the list.

The fundamental difference between the conceptual and the other levels of specification is that only the former is medium-independent. Because it mediates between the visual and the auditory representation, had it not been medium-independent, it would be possible to erroneously include in the specification of the auditory representation some information necessary to the visual widget only. Another reason is the number of multi-modal issues (missing haptic information and advance feedback) considered and resolved at this level.

2.2 Structural Level

At this level, the structure of the auditory scene is defined in terms of auditory streams. The term stream here implies a series of auditory events with some common physical characteristic. The term 'channel' used in the early literature on auditory attention corresponds to the notion of a stream (Neumann et al., 1986).

One distinction found in the auditory attention literature is that between voluntary and involuntary attention. In general, we are able to control our attention and selectively attend to a particular stream, or divide our attention among a number of streams at will. However, our auditory system can detect changes in sounds, especially when these are sudden ones or when a new sound is introduced. Then, depending on task load, attention can be involuntary drawn by these sounds, even if they have not been attended to previously.

Most of the debate in auditory attention has been focused on voluntary attention issues. There have been a number of theories on voluntary auditory attention - see ten Hoopen (1996) for a review. According to Hawkins & Presson (1986):

> "To a certain point in the information processing sequence it appears that information from several sources ... can be processed in parallel. Beyond this point, however, inputs usually must be processed in series."

Nevertheless, there is no consensus on where this point lies in the information processing - theories of early selection, late selection and multiple loci theories (ten Hoopen, 1996).

It is a common finding that subjects are able to focus their attention on a particular stream of sounds (selective attention tasks) quite effectively. Most of the time, they are able to tune out all the other sounds. They will notice only gross changes in the physical characteristics of the non-attended auditory material, such as changes from male to female voice or from voice to a tone. Subjects are usually unable to report the semantic content of non-attended sounds. The consequence for auditory design is that, if users are attending selectively to a stream of sounds while performing a task, then it is quite likely that they will miss information conveyed in the non-attended streams. In other words, when performing a selective attention task, all the required information should be contained in the stream attended to, otherwise users might fail to integrate information presented in non-attended streams.

Nevertheless, a number of studies have reported that non-attended sounds are processed to some extend. Despite the fact that the so-called breakthrough of the non-attended sound is very limited (as low as 6%), the phenomenon has been used as evidence for late-selection theories. These imply that full semantic processing of all sounds in the auditory scene takes place before one of these sounds is selected for further processing. However, the interpretation of these findings is not unequivocal, since most of these studies have not adequately controlled where subjects had been directing their attention. Early-selection theories could also account for these findings. It might have been the case that subjects were relaxing their attention and sampling the non-attended stream(s) out of curiosity or that some non-attended sounds introduced a change in the auditory scene large enough to trigger involuntary attention. In either case, the breakthrough of the non-attended is very limited and it would not be safe practice for a designer to assume that users will reliably integrate non-attended information.

It is not necessary that users will selectively attend to a stream. They may also divide their attention among a number of streams. In this case, they will be able to integrate information among a number of streams. But then, they are usually subject to 'divided attention costs', that is, liable to more errors and increased reaction times. Moreover, even given considerable practice:

> "...it appears doubtful, practically speaking, that practice under divided attention conditions ... would ever bring performance up to levels achieved under focused [selective] attention conditions." (Hawkins & Presson, 1986).

Consequently, it appears that all information required for performing a task should be bound in a single stream, although a stream may support more than one task. If information for a task is distributed over a number of streams then users would find the task more difficult; designing for selective attention, optimizes performance.,

Returning to the checkbox widget example, we have to examine whether the auditory scene complies with the above guideline. It seems reasonable to map each of the three dimensions of a checkbox entity to a distinct stream. This is mostly because we would like the user to be able to attend to any of these independently, but also because the sounds used would be substantially different from each other and hence they would be perceived as distinct streams.

Each stream has to be examined with respect to the tasks associated with it. The stream corresponding to the abstract dimension of position contains all the information required to locate either end of the list. Nevertheless, most tasks rely on information about the state and the label of each checkbox, too. For instance, the task of browsing the labels of the checkboxes would result in the stream with that information becoming the attended one. What would happen if the last checkbox in the list had been reached? From the above discussion it follows that if users were selectively attending to the label stream they would be likely to miss the position information. If they were dividing their attention between the label and position stream, it would take more time to complete the task and they would occasionally overrun the end of the list. Consequently, the abstract dimension of labels must be re-defined at the conceptual level to include the information: first item in the list, last item, elsewhere in the list. Identifying the state of each checkbox is also liable to the same problem. The conceptual level has to be re-specified until the structural level constrains are satisfied.

2.3 Implementation Level

The designer has also to decide on the physical dimensions of sound that will implement the auditory representation. A number of dimensions have been suggested for making a stream easier to attend to. These include pitch, timbre, spatial origin, temporal organization, onset asynchronies, intensity, as well as familiarity with the sound -see for instance (ten Hoopen, 1996; Deutsch, 1996; Bregman, 1990).

It seems that a good design principle is to use a number of physical dimensions to implement an abstract one. This redundancy is necessary to accommodate individual differences and hearing impairments users might have, as well as to make sure that a number of auditory illusions are avoided. For instance, consider the octave illusion (Deutsch, 1986) where frequency and spatial origin cues are opposed to each other. Frequency dominates, giving rise to a number of illusory percepts. However, if the spatial origin information is supported by a number of additional cues such as timbre and onset asynchronies, then the illusion disappears.

The following would be a possible implementation for the checkbox widget. It is not intended as an example of an artefact; the aim here is to illustrate the specification process at the implementation level and demonstrate how the abstract specifications derived at the conceptual and structural level are applied to the realization of the auditory widget.

Labels are presented using synthetic speech which is interrupted as soon as the user moves to an adjacent checkbox. The list is portrayed in a horizontal orientation, so that the speech is displaced to the left or right for the first and last checkboxes respectively. Since the other sounds are non-speech ones, labels form a distinct stream easily attended to.

The state of a checkbox is represented by a sampled sound produced by scratching a pen on a paper as if ticking a checkbox. The pitch of the sampled sound is varied to produce a dull or bright sound so that the two states are distinguishable (unchecked or checked respectively). Conforming to the re-definition in the conceptual level imposed by the structural level, the state sound of the first or last checkbox in the list is displaced to the left or right.

Position in the list is represented by a brief tone. For advance feedback, a simultaneous dissonant tone is presented and both tones are displaced to the left or right, appropriately. For the first and last item, the sound is again displaced as above. In this way, five values are created as prescribed in the conceptual level. Importantly, the user can not only discriminate one value from the other but may also identify each value on its own and hence, his or her position in the list at any time. Thus, even if disrupted, the user may resume a task without having to remember the current position in the list. It should be noted that in order to derive a number of absolutely identifiable values it might be necessary to use more than one dimension - yet another reason for redundant dimensions in the implementation. For instance, it is easy to distinguish a large number of pitches from one another but only few people have the ability of absolute pitch.

3 Experimental Evaluation

The aim of the experiment described here was to investigate the validity of a basic guideline of the methodology for the structural level, namely that all information required for performing a task should be in one stream. Numerous psychological experiments have demonstrated that performance deteriorates when attention is divided to more than one stream. However, the stimuli typically used in these experiments are austere and experimental conditions tightly controlled. It might be the case that, in practice, the design guideline in question does not introduce any particular improvements just because there is enough time to switch attention between streams, as in some psychological experiments that failed to show performance deterioration.

Also, most psychological experiments on selective attention have been employing presentation of spoken messages. It is quite likely that non-speech sounds can be semantically processed faster; this would largely eliminate any costs of dividing attention. Consequently, we would like to examine the validity of the guideline proposed in Section 2.2 in a situation more ecologically appropriate to the auditory user interfaces.

3.1 Method

Subjects: Eight subjects, from the academic community of the University of York were recruited. Most subjects were experienced musicians apart from two who were less experienced, though they used to play a musical instrument. The reason for using experienced musicians as subjects is that we expect them to be more practiced in attending to a number of instruments at a time. Consequently, they could be better in divided attention tasks than subjects with no significant experience in music. If musicians' performance in selective attention tasks is found to be superior than in divided attention tasks, then it is quite likely that the difference will be more pronounced for other users and even persist despite any practice given.

Stimuli and Apparatus: Sounds were generated with a SB-AWE64 Gold sound card and then presented over headphones. Subjects had to use the left and right cursor key and space bar of a PC keyboard; there was no visual contact with the computer monitor. Subjects' answers and presented sounds were taped using a Sony DAT recorder.

The experiment was a repeated-measures 2x2 factorial design. The two factors were the condition and the task factor. After having completed a training session, each subject served in all four blocks, the order of which was determined by a balanced Latin square. Each block contained 24 lists of checkboxes randomly generated by the computer. The first four lists in a block were discarded prior to any analysis.

Each condition level was a variation of the checkbox sounds (label, state and position) described in Section 2.3. The only difference between the two conditions was that, in the selective condition, all required information was in the label stream whereas, in the divided condition, information had to be integrated across the label and the position streams. In more detail, for the selective condition, the position stream was kept to the centre providing no spatial information. The sounds in the label stream were displaced to denote the leftmost and rightmost checkbox in the list. For the divided condition, labels did not provide any spatial information; this had to be extracted from the position stream which was displaced following the same rule as labels in the selective condition. The checkbox stream did not provide any spatial information and was kept the same under both conditions; information contained in this stream was not required by any of the tasks in the experiment.

Sounds departed slightly from those described at the implementation level. Digits (0, 4, ..., 9) were used for labels. Advance feedback was substituted by random variations in the position sounds, to keep the spatial information provided by the label stream in the selective condition *equivalent* to that of the position stream in the divided condition. Position sounds were louder and prolonged to make attending to them as easy as possible. These changes were introduced following a pilot study.

The first task level was the *target* task. For each list, its leftmost or rightmost checkbox was presented at random. Subjects had to move towards the other end of the list, one checkbox at a time, using the cursor keys. They had to press the space bar as soon as they had reached the other end of the list or a target checkbox (labelled V) had been encountered.

The second task level was the *addition* task which was essentially a target task with the variation that numerical labels had to be summed for each list. At the end of each list, subjects had to say aloud the sum.

Three dependent variables were of interest for each list: the mean key-press rate, the first-key-press time and errors (target and list-end overruns). For time measures the upper and lower quartiles for each subject and each block were discarded to control for outliers.

3.2 First-Key-press Times

Figure 2: First-key-press reaction times (ms).

For the first checkbox presented in a list subject had to integrate two pieces of information:

- First, whether or not the checkbox was a target. If it was, they had to press the space bar immediately.
- Second, subjects had to work out whether they were at the leftmost or the rightmost checkbox, in order to move towards the other end of the list.

For the selective condition this information was in the label stream; for the divided, in the position stream. For the target task, subjects had to compare the label to the target value (zero). For the addition task, the value of the number had to be identified and remembered for the additions to follow though as yet no mental arithmetic took place. First-key-press mean times are shown in Figure 2. The following results are obtained running the repeated-measures ANOVA test:

Source	F	Sig.
Task	58.088	0.000
Condition	12.110	0.010
Interaction	6.555	0.038

The significant condition main effect reveals that subjects were reacting faster in the selective condition. But this is not necessarily attributed to divided attention costs. An alternative hypothesis might be that extracting spatial information was more difficult from the position stream than the label stream. Since the same mapping rules were used, the problem would not be in interpreting the values but in perceiving the physical dimensions of the position sounds in the first place. However, this is unlikely for a number of reasons. First, the position sounds were designed to be as easy to attend to as possible. Second, during the training session, subjects had to selectively attend to the label stream in the selective condition and to the position stream in the divided condition to extract spatial information only. When asked to compare the difficulty of selectively attending to either stream, most subjects found no difference though a few preferred either the label or the position sounds to convey spatial information. Overall, there was no trend revealing preference of either condition.

The alternative hypothesis is refuted by the significant interaction effect. If spatial information was harder to extract from position sounds then the added amount of difficulty should remain constant for either task since:

1. the processing of digits remains unaffected by the condition factor (they are always on the label stream); and
2. extracting spatial information is the same for both tasks under divided conditions, since the addition task is essentially a target task as far as processing of spatial information is concerned.

A constant level of difficulty implies no interaction between tasks and conditions. Hence, the alternative hypothesis is rejected.

Thus, an examination of the first-key-press reaction time allows us to conclude that differences in performance can be attributed to costs of divided attention rather than the relative difficulty, if any, of extracting spatial information from the position sounds. This would validate our guideline from the designer’s point of view if significant differences exist in overall performance measured by the mean keypress rate variable.

3.3 Mean Key-press and Error Rates

Figure 3: Mean key-press rates (ms/key-press).

Mean performance is a more important factor to the designer. Mean key-press rates are shown in Figure 3. The following results are obtained running a repeated-measures ANOVA test:

Source	F	Sig
Task	44.768	0.000
Condition	7.391	0.030
Interaction	1.166	0.316

The two main effects (task, condition) are significant. The insignificant interaction cannot be related to absence of divided attention costs. First, while browsing the lists in the divided condition, subjects may be able to successfully accomplish the tasks without having to divide their attention for every checkbox, as long as the end of the list has not been reached. Thus, it is difficult to unequivocally associate results with the underlying psychological processes. Moreover, the variance in key-press rates attributed to the mental arithmetic task would by large enough to absorb any divided attention costs.

The target task is particularly interesting from the designer’s point of view, since it is a fundamental constituent for most tasks in the auditory interface. Performance was worse by 20.7% or 115 msecs per key-press in the divided condition. Subjectively, only one subject weakly preferred the divided condition to the selective one, although all subjects performed significantly better in the selective condition. Most subjects felt that the selective condition was easier; two of them believed that it was much easier than the divided condition. Individuals were affected to different extents. For two of them, divided attention costs were less than 10%, for three it ranged between 15% and 20%, for the rest between 25% and 55%.

The results for error rates are not significant in general because four out of eight subjects had almost errorless performance. The trend was a total of 14 errors in the selective condition and 28 in the divided. The addition task errors were significantly more than the target task errors (1.5 vs. 3.75 errors per subject, significant at 5%).

4 Conclusions

The proposed methodology has been applied to the specification of a number of widgets such as checkboxes, radio buttons and listboxes. The basic psychological principles underlying the methodology have been experimentally evaluated. It appears that significant improvements in performance can be gained by applying the methodology in the design of auditory representations. Because of space limitations, a detailed discussion of several issues considered in the methodology has been omitted (such as how involuntary attention can be used appropriately or how the physical dimensions of sound can be manipulated to implement a desirable auditory scene structure). However, we believe that the main issues have been highlighted and validated, proving the methodology to be a valuable tool for auditory interface designers.

Acknowledgements

The work presented in this paper is part of a long-term project sponsored by EPSRC (grant number GR/M12063), which aims to derive a principled design methodology for the specification of multi-modal (auditory and haptic) non-visual interfaces.

References

Bregman, A. (1990), *Auditory Scene Analysis,* MIT Press.

Brewster, S. A. (1998), “The design of sonically-enhanced widgets”, *Interacting with Computers* 11(2), 211-35.

Deutsch, D. (1986), Auditory Pattern Recognition, *in* K Boff, L. Kaufrnan & J. R Thomas (eds.), *Handbook of Perception and Human Performance,* Vol. 2, J Wiley & Sons, chapter 32, pp.32.1-32.49.

Deutsch, D. (1996), The Perception of Auditory Patterns *in* W. Prinz & B. Bridgeman (eds.), *Handbook Perception and Action,* Vol. 1, Academic Pi pp.253-96.

Edwards, A. (1989), “Soundtrack: An Auditory Interface Blind Users”, *Human-Computer Interaction* 4(1), 66.

Edwards, A. D. N. (1995), Outspoken Software for B Users, *in* A. D. N. Edwards (ed.), *Extra-0rdinary Human-Computer Interaction,* Cambridge University Press, pp.59-82.

Foley, J. D., van Dam, A., Feiner, S. K. & Hug J. E (1990), *Computers Graphics: Principles Practice,* second edition, Addison-Wesley.

Hawkins, H. & Presson, J. (1986), Auditory Inform, Processing, *in* K. R. Boff, L. Kauffnan & I Thomas (eds.), *Handbook of Perception and Human Performance,* Vol. 2, John Wiley & Sons, chapter pp.26.1-26.64.

Kurze, M. (1997), Rendering Graphics for Interactive Haptic Perception, in *S.* Pemberton (ed.) *Proceedings of CHI’97 Human Factors Computing Systems,* ACM Press, pp.423 http://www.acm.org/sigchi/chi97/proceedings/paper/mk.htm

Mitsopoulos, E. & Edwards, A. D. N. (1997), Auditory Scene Analysis as the Basis for Designing And Widgets, *in* E. Mynatt & J. A. Ballas (c *Proceedings of the International Conference Auditory Display (ICAD’97),* Xerox Corpora pp. 13-8.

Mitsopoulos, E. & Edwards, A. D. N. (1998), A Principled Methodology for the Specification and Design of 1 visual Widgets, *in S.* Brewster & A. D. N. Edwards (eds.), *Proceedings of the International Conference Auditory Display (ICAD’98),* The British Computer Society.

Mitsopoulos, E. & Edwards, A. D. N. (to appear Methodology for the Specification of Non-Visual Widgets, *in Adjunct Conference Proceedings of International ‘99.*

Mynatt, E. D. & Weber, G. (1994), Nonvisual Present of Graphical User Interfaces: Contrasting Approaches, *in B.* Adelson, S. Dumais & J. (eds.), *Proceedings of CHI’94 Human Factors Computing Systems,* ACM Press, pp. 166-172.

Neumann, 0., van der Heijden, A. H. C. & Allport, 1 (1986), “Visual Selective Attention: Introduction Remarks”, *Psychological Research* 48(4), 185-8

ten Hoopen, G. (1996), Auditory Attention, *in 0.* Neumann & A. E Saunders (eds.), *Handbook of Perception and Action, Vol.* 3, Academic Press, pp.79-112.

Weber, G. (1993), Access by Blind and Partially Sighted People to Interaction Objects in MS-Windows, *in* The Swedish Handicapped Institute (ed.), *Proceedings of* the 2nd European Conference on the Advancement of Rehabilitation Technology (ECART), p.2.2.

Zhang, J. (1996), “A Representational Analysis of Relational Information Displays”, *International Journal of Human-Computer Studies 45(1),* 59-74.