Auralization of Document Structure

Steve Portigal and Tom Carey
University of Guelph
Guelph, ON
Canada N1G 2W1
Email: stevep, tcarey@snowhite.cis.uoguelph.ca

Abstract

An experiment compared the effectiveness of auditory, visual, and combination cues to convey document structure. Subjects demonstrated an equivalent level of understanding of the document structure and its content with either a combination cue or a visual cue. Subjects required more time to answer questions in the combination condition than in the visual condition. This suggests a greater cognitive effort is required. A sound-only condition has the poorest performance both in response time and in the subject's answers to questions about the document's structure and its content. Subjects were grouped based on whether or not they replayed sounds as a retention tactic. Subjects who replayed sounds did better than subjects who did not. These results contribute to our understanding of potential uses of sound in user interfaces. The cues used for this task do not appear promising. Future research to determine how to make use of sound must consider user tactics for processing sound cues.

1 Introduction

Computer-based electronic text has become ubiquitous. Kiosks in shopping malls and tourist attractions use electronic text to present information to patrons. Universities and schools are making use of computer-based instruction. Supermarkets use computers to help shoppers find the correct aisle, or to print out coupons.

In this research we have used audio cues as an analog or replacement for visual reading cues. We used these audio cues to provide navigational information as the reader moves through a document. We hypothesized that this will aid in navigation.

1.1 Typographical Structure and Reading

There are a variety of visual signals that are used when reading. Lorch considers typographical signals to be a subset of text signals in general, which also include headings, titles, and summaries [22, 23, 24]. He suggests that each of these signals will aid memory, and that some of them have been demonstrated to assist comprehension, focus attention, improve reading processes, and aid searching within the text.

Other visual factors that can influence the effectiveness of a document include page size, margins, column widths, typefaces, and justification [13, 14]. Typographic cues may influence comprehension. In many cases, readers will provide their own cues (highlighting, underlining, marginal notes and so on) if they are left out of the original text.

1.2 Document Structure and Reading

Many people are familiar with the experience of remembering what part of a document or page contains some information without being able to remember the information itself. There have been several investigations of different aspects of this phenomenon [41, 42, 21]. These researchers found that memory of location (within-page) and memory of content were correlated. There is not a strong indication that memory of within-document location and content are related. Location seems to be a stronger cue for content than content as a cue for location.

1.3 Hypertext Structure and Navigation

Studies of users working with hypertext have shown that they easily become lost in the document [25, 26, 27, 28, 29, 30, 31, 32, 33]. Navigation behavior in a hypertext system can be classified as rambling, touring, or orienteering [12]. Rambling involves loosely directed movement, touring is a system-controlled path through the hypertext, and orienteering involves the use of maps to get from one place to the next.

An investigation by Hendry demonstrated that subjects who were given a tour (based on other users' behavior) through a hypertext document achieved comprehension levels nearly equal to subjects who chose their own path [15, 16, 17, 18, 19, 20]. This shows that the acquisition of content by navigation alone can be simulated by touring.

1.4 Auditory Perception

Understanding the manner in which auditory information is processed is critical when designing an interface which makes use of audio. Listeners can differentiate between many dimensions of sound: pitch, volume, spatial location, duration, timbre, attack, and timing [1, 2, 3, 4, 5]. The greater the difference between two messages along one dimension (i.e., intensity, pitch, etc.) or the greater the number of dimensions along which two messages differ, the more discriminable those two messages will be. There is little evidence regarding the degree to which one dimension may be more important than another [39, 40].

Although cues are heard individually, information obtained in processing one cue may assist in processing another cue. Deutsch observed that subjects performed significantly better in recalling tonal sequences that contained a hierarchical structure [10]. The subjects were able to take advantage of the structure when building their own model of the tonal information.

2 Method

2.1 Subjects

Twenty-seven volunteers (22 male and five female) from the University of Guelph participated in this experiment. All the subjects were comfortable using a computer with a mouse.

2.2 Design

A within-subjects counterbalanced design was used, in which there were three conditions: SVC (sound, visual, combination), VCS (visual, combination, sound), and CSV (combination, sound, visual). The condition determined the order in which the subject would encounter the different documents and cue types. There were nine subjects assigned to each of the three conditions.

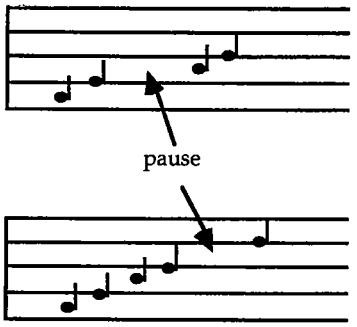

Figure 1: Pictorial representation of an auditory cue for chapter 2.4

2.3 Materials

The documents were presented on an Apple *Macintosh IIsi* computer, using three HyperCard documents having structures that were isomorphic to each other. One document explained the workings of a camera, one of these dealt with automobile maintenance, and the other was about microwave cooking and operation. The three documents were disguised slightly from each other to prevent subjects from recognizing the isomorphic document structure as the experiment progressed. The Flesch Reading Ease scores for the camera, microwave, and automobile documents were, respectively, 57.4, 59.5, and 62.0.

Based on user browsing behavior in hypertext a path was chosen through each documents [15]. Since the documents were structurally isomorphic, the path was the same for all three documents. The 11 section path included behavior such as jumps to related material, linear reading, and rejection of unwanted subject matter. As each new section was opened, a cue would indicate the subject's location within the document. The cues were auditory, visual, or a combination.

The sound cues consisted of two sequences of tones. The number of tones in the first sequence indicated the number of chapters, with a pause after the current chapter. The number of tones in the second sequence indicated the number of subsections in the current chapter, with a pause after the current subsection (see Figure 1 for an example). Each note was an eighth note, with the pause being a quarter note, and the break between each sequence being a whole note. The first sequence began at middle C. The second sequence started one octave above the note that preceded the pause in the first sequence. The first sequence started at middle C and played a scale containing the appropriate number of notes. The second sequence was also a scale, but would not necessarily be a major scale, since the first note was determined by where the pause was located in the first sequence.

The visual cues used boxes to represent the chapters. As the section was opened, the map would be shown. The entire chapter and subsection would be highlighted in gray. After a pause the box for the current chapter would change to black. After another pause the box for the subsection was changed to white and the current subsection was changed to black. The map then remained visible for about 80 milliseconds.

The combination cue had the visual information synchronized with the auditory tones.

We used three sets of 11 isomorphic questions. Each set consisted of three content questions (e.g., What is a method for doing X?), five navigation questions (e.g., How many more subsections remain in the current chapter?) and three content-navigation questions (e.g., Where would you look for information about X?). After reading each section the user would be given a multiple-choice question. After responding, the user would be moved to the next section along the browsing path.

Cue Type	n	Mean	Std. Dev.
Visual	27	.650	.176
Combination	27	.663	.144
Sound	27	.539	.192

Table 1: Means showing effect of cue type on correctness.

Replay Tactic	n	Mean	Std. Dev.
Never recalled any sounds	39	.562	.184
Sometimes recalled sounds	42	.669	.159

Table 2: Means showing effect of recall tactic on correctness.

2.4 Measures

We recorded the time to read each section, the time to answer each question, and the answer to the question. Subjects were debriefed about their preferences, strategies, and background.

2.5 Procedure

Subjects were given an overview of the experiment. Each subject was trained with the visual cues and the auditory cues until they were comfortable with the cues. Then they were given secondary instructions to provide context for the task. Subjects worked through each document in turn, answering questions after finishing each screen. After the final session the experimenter debriefed the subjects.

3 Results

3.1 Correctness

A repeated measures ANOVA was performed showing that cue type had a significant effect on correctness ($F_2, 48 = 6.69, p < .005$). Group means were analyzed using a predetermined contrast of means test which revealed no significant difference between the visual and combination conditions ($F_2, 48 = 0.13$, ns), but a significant effect between both the visual and sound ($F_2, 48 = 8.84, p < .005$) and between the combination and sound conditions ($F_2, 48 = 11.10, p < .005$). The mean correctness and standard deviations are show in Table 1.

The subjects were grouped based on whether or not they indicated during the debriefing that they recalled the sound cue or the sound portion of the combination cue after it had finished playing (referred to as replay tactic). A repeated measures ANOVA indicated that this replay tactic had a significant effect on correctness ($F_{1,23} = 8.25, p < .01$). The mean correctness and standard deviations are show in Table 2.

3.2 Reading Speed

A repeated measures ANOVA indicated that cue type had a significant effect on reading speed ($F_2, 48 = 5.90, p < .01$). Group means were analyzed using a predetermined contrast of means test which revealed no significant difference between the sound and combination conditions ($F_2, 48 = 0.53$, ns), but a significant difference between visual and combination ($F_2, 48 = 6.47, p < .05$)

Cue Type	n	Mean	Std. Dev.
Visual	27	206.54	70.53
Combination	27	190.30	66.47
Sound	27	185.65	68.32

Table 3: Means Showing effect of cue type on reading speed in words per minute.

Cue Type	n	Mean	Std. Dev.
Visual	27	8.70	2.58
Combination	27	9.50	2.57
Sound	27	10.91	4.16

Table 4: Means showing effect of cue type on response time in seconds.

and between visual and sound ($F_2, 48 = 10.71, p < .005$). The mean correctness and standard deviations are show in Table 3.

3.3 Response Time

The time taken for the subject to respond to each question was averaged, and a repeated measures ANOVA was performed. Cue type had a significant effect on response time ($F_{2,48} = 21.41, p < .005$).

4 Discussion

4.1 Summary of Results

The results indicate that sound cues alone were not sufficient to provide navigational information at the same level as visual or combination cues. The presence of sound cues demanded more effort from the user. The slower reading speeds accompanying sound and combination cues, and the slower response times for the same cues, indicate that more time and presumably more effort was required. This is consistent with previous research. The presence of sound in the combination cue did not affect correctness, but more effort was required to achieve the same correctness as in the visual condition.

4.2 Interpretation of Results

Since correctness results were not different between the combination and visual conditions, it appears that the presence of sound has no effect upon the use of the visual information. The results for reading speed and response time belie this, however. The times for the combination condition were longer than for the visual condition, indicating that sound distracted the user. This is confirmed by subject comments indicating that the combination cue divided their attention as they tried to follow both cues at once. Other users indicated that they had to put in more effort with the combination cue by verifying the sound cue against the visual cue. Many subjects reported that they ignored the audio portion of the combination cue. It seems possible that actively ignoring the sound required some effort, which explains the slower reading time in the combination condition.

A deeper examination that takes into account user tactics reveals that users who reported that they recalled sounds did much better in the sound condition, but not in the combination

condition. This confirms that the combination condition did not benefit from the presence of sound, and again many of the subjects ignored the sound in that condition.

The usefulness of sound depends on both the design of the sound and the task involved. Some subjects reported that the sounds were too fast to count and that they were unable to decode them. Yet in the training sessions the subjects became proficient at using the cue. When placed in the task context, remembering the cue, or decoding the cue and remembering its meaning may be overloading the subjects' processing capacity. Other research demonstrated that a combination cue was better than either an auditory or visual cue [6, 7, 8, 9], or that auditory only cues were better than visual or combined [11]. The task in that study was quite different.

4.3 Implications

Designers who wish to add auditory information to their interface should think carefully before doing so. The results for the combination condition show that the inclusion of sound can increase the mental workload and slow down the user. Sound should not be dismissed as useless, however. Some subjects reported that having the sound cue was useful.

Subjects reported using two different tactics to make use of the different cue types. This suggests that certain behaviors (if properly encouraged) would enable users to make much better use of the auditory information. Some subjects found the combination cue confusing while others found the redundant information useful. The strategies employed affected the subjects' ability to make use of the different types of cues.

Conclusions

We chose not to use more sophisticated sound hardware because we felt that it was important to investigate sound using generally available and inexpensive equipment.

The sound cues used in this study were not as effective as the visual cues, while the combination of sound and visual was as effective but required more effort to achieve the same level of performance. This research shows the need for more work to identify the different types of user strategies and tactics, and how they might be encouraged. Sound cues that were designed to take advantage of those tactics could be a useful supplement to visual cues and even a replacement for visual cues.

There is still great potential for nonspeech audio at the interface. The results here show a limitation of the methodological approach and the design of the cues themselves. Further research can overcome those shortcomings and determine ways in which sound can be of benefit.

This work shows that there are limits to what sound can do in an interface, and that a better understanding of these limits is needed so that guidelines can be created for designers who wish to take advantage of what sound can do.

Acknowledgments

This research was aided by a Grant-in-Aid of Research from Sigma Xi, The Scientific Research Society, and by funding from NSERC.

References

[1] Buxton, W., S. Bly, S. Frysinger, D. Lunney, D. Mansur, J. Mezrich, and R. Morrison. "Communicating with Sound." In *Proceedings of CHI'85 Conference on Human Factors in Computing Systems*, edited by L. Borman and B. Curtis, 115–119. New York: ACM, 1985.

[2] Buxton, W., W. Gaver, and S. Bly. (1995): in press.

[3] *The Use of Nonspeech Audio Displays in Human Computer Interaction.* Cambridge: Cambridge University Press.

[4] *Chilton's Repair and Tune-Up Guide, Dodge 1968-77.* Radnor, PA: Chilton Book Company, 1977.

[5] Colovita, F. B. "Human Sensory Dominance." *Perception and Psychophysics* **16 (2)** (1974): 409–412.

[6] Bly, S. "Presenting Information in Sound." In *Human Factors in Computer Systems*, edited by M. Schneider, 371–375. New York: ACM, 1982.

[7] DiGiano, C. J. *Visualizing Program Behavior Using Non-Speech Audio.* Unpublished M.Sc. thesis, University of Toronto, 1992.

[8] DiGiano, C. J., and R. M. Baecker. "Program Auralization: Sound Enhancements to the Programming Environment." *Proceedings Graphics Interface '92*, 44–52. Toronto: CIPS, 1992.

[9] Dixon, N. F. *Preconscious Processing.* Chichester: Wiley, 1981.

[10] Deutsch, D. "The Processing of Structured and Unstructured Tonal Sequences." *Perception and Psychophysics* **28** (1980): 381–389.

[11] Fitch, W. T., and G. Kramer. "Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System." In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, 307–326. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[12] Hammond, N., and L. Allinson. "Travels Around a Learning Support Environment: Rambling, Orienteering, or Touring?" In *Proceedings of CHI'88 Conference on Human Factors in Computing Systems*, edited by E. Soloway, D. Frye and S. B. Sheppard, 269–273. New York: ACM, 1988.

[13] Hartley, J. "Planning the Typographical Structure of Instructional Text." *Educational Psychologist* **27** (1986): 315–332.

[14] Hartley, J. "Typography and Executive Control Processes in Reading." In *Executive Control Processes in Reading*, edited by B. K. Britton, and S. M. Glynn, 57–80. Hillsdale, NJ: Lawrence Erlbaum, 1987.

[15] Hendry, D. "The Relationship Between Navigation and Comprehension in a Hypertext Environment." M.Sc., University of Guelph, 1989.

[16] Jensen, P. *Microwave Cookbook—The Complete Guide.* Tuscon, AZ: HP Books, 1986.

[17] Jones, S. D., and S. M. Furner. "The Construction of Audio Icons and Information Cues for Human Computer Dialogues." *Contemporary Ergonomics, Proceedings of the Ergonomics Society's 1989 Annual Conference*, 436–441. Reading, England: Taylor & Francis, 1989.

[18] Kesterton, M. "Wired World." *The Globe and Mail* (1993): A22.

[19] Kirk, R.E. *Experimental Design: Procedures for the Behavioral Sciences*, 2nd ed. Belmont, CA: Brooks/Cole, 1982.

[20] Lachman, R., J. L. Lachman, and E. C. Butterfield. *Cognitive Psychology and Information Processing: An Introduction.* Hillsdale, NJ: Lawrence Erlbaum, 1979.

[21] Lovelace, E. A., and S. D. Southall. "Memory for Words in Proseand their Locations on the Page." *Mem. & Cog.* **11** (1983): 429–434.

[22] Lorch, R. F. "Text Signaling Devices and their Effects on Reading and Memory Processes." *Educational Psychology Review* **1** (1989): 209–234.

[23] Ludwig, L., N. Pincever, and M. Cohen. "Extending the Notion of a Window System to Audio." *IEEE Comp.* **23(8)** (1990): 66–72.

[24] Lumbreras, M. *A Hypertext for Blind People.* (Hypertext '93 poster), 1993.

[25] Lungu, D., T. T. Carey, J. Mitterer, and B. Nonnecke. "A Comparison of Access Methods for Online Documentation." In *Proceedings IBM Canada 1992 CASCon*, edited by J. Slonim et. al, 194–201. Toronto, ON: IBM Canada Centre for Advanced Studies, 1992.

[26] Mansur, D. L., M. M. Blattner, and K. I. Joy. "Sound-Graphs: A Numerical Data Analysis Method for the Blind." *18th Hawaii International Conference on Systems Science*, 163–174. New York: IEEE, 1985.

[27] Microsoft Corporation. *Microsoft Word 5.0a* [Computer program]. Redmond, WA, 1991.

[28] Muchnik, C., M. Efrati, E. Nemeth, M. Malin, and M. Hildesheimer. "Central Auditory Skills in Blind and Sighted Subjects." *Scandinavian Audiology.* **20(1)** (1991): 19–23.

[29] Mynatt, E. D., and W. K. Edwards. "Mapping GUIs to Auditory Interfaces." *Proceedings of the ACM Symposium on User Interface Software and Technology*, 61–70. New York: ACM, 1992.

[30] Oborne, D. J., and D. Holton. "Reading From Screen Versus Paper: There is no Difference." *International J. Man-Machine Studies* **28** (1988): 1–9.

[31] Paciello, M. G. "Accessible Document Design: The Vision and Goals of the International Committee on Accessible Document Design (ICADD)." *Infor. & Tech. J.* (1995): in press.

[32] Pitt, I. J., and A. D. N. Edwards. "Navigating the Interface by Sound for Blind Users." In *People and Computers VI*, edited by D. Diaper and N. Hammond, 373–383. Cambridge: Cambridge University Press, 1991.

[33] Prior, M., and G. A. Troup. "Processing of Timbre and Rhythm in Musicians and Nonmusicians." *Cortex* **24(3)** (1988): 451–456.

[34] Schulze, H. H. "Categorical Perception of Rhythmic Patterns." *Psychol. Res.* **51** (1989): 10–15.

[35] Teshiba, K., and M. Chignell. "Development of a User Model Evaluation Technique for Hypermedia Based Interfaces." In *Proceedings of the Human Factors Society 32nd Annual Meeting*, 323–327. New York: HFS, 1988.

[36] van Nes, F. L. "Space, Colour, and Typography on Visual Display Terminals." *Behav. & Infor. Techn.* **5(2)** (1986): 99-118.

[37] Vanderheiden, G. C. *A White Paper on the Design of Software Application Programs to Increase Their Accessibility for People with Disabilities.* Madison: University of Wisconsin-Madison, Trace R&D Center, 1992.

[38] Wagenaar, W. A., C. A. Varey, and P. T. Hudson. "Do Audiovisuals Aid? A Study of Biosensory Presentation on the Recall of Information. In *Attention and Performance X: Control of Language Processes*, edited by H. Bouma, and D. G. Bouwhuis, 379–391. London: Lawrence Erlbaum, 1984.

[39] Wickens, C. D. *Engineering Psychology and Human Performance.* Columbus, OH: Merrill, 1984.

[40] Wright, P. "Cognitive Overheads and Prostheses: Some Issues in Evaluating Hypertexts." *Proceedings of ACM Hypertext'91 Conference*, 1–12. New York: ACM, 1991.

[41] Zechmeister, E. B., and J. McKillip. "Recall of Place on The Page." *J. Ed. Psychol.* **63** (1973): 446–453.

[42] Zechmeister, E. B., J. McKillip, S. Pasko, and D. Bespalec. "Visual Memory for Place on the Page." *J. Gen. Psychol.* **92** (1975): 43–52.