http://dx.doi.org/10.5392/IJoC.2013.9.2.029

An Empirical Analysis of Auditory Interfaces in Human-computer Interaction

Yoonjae Nam
Department of Culture and Tourism Contents, College of Hotel & Tourism Management
Kyung Hee University, South Korea

ABSTRACT

This study attempted to compare usability of auditory interfaces, which is a comprehensive concept that includes safety, utility, effectiveness, and efficiency, in personal computing environments: verbal messages (speech sounds), earcons (musical sounds), and auditory icons (natural sounds). This study hypothesized that verbal messages would offer higher usability than earcons and auditory icons, since the verbal messages are easy to interpret and understand based on semiotic process. In this study, usability was measured by a set of seven items: ability to inform what the program is doing, relevance to visual interfaces, degree of stimulation, degree of understandability, perceived time pressure, clearness of sound outputs, and degrees of satisfaction. Through the experimental research, the results showed that verbal messages provided the highest level of usability. On the contrary, auditory icons showed the lowest level of usability, as they require users to establish new coding schemes, and thus demand more mental effort from users.

Key words*: Auditory Interface, Verbal message, Earcon, Auditory icons, Usability*

1. INTRODUCTION

The purpose of this study was to investigate the function of auditory interfaces in the personal computing environment. So far, most human-computer interaction (HCI) studies have emphasized the visual aspects of computing environments, including graphic icons, buttons, scrollbars, font designs, elements on screens, and equipment for visual displays. It is true that visual perception is the most memorable and containable.

Though visual sources offer more information than other senses, the visual and auditory senses offer complementary information about the world; they are interdependent. The visual system gives us detailed data about a small area of focus whereas the auditory system provides general data from all around, alerting us to things beyond our peripheral vision [1].

Sounds can be used in more information-rich ways to show what is happening in a system. In particular, sound can be used as a coding method to augment graphical representation. For example, Buxton, Gaver, and Bly [2] suggested that complex systems might benefit from using sounds that had a highly evolved hearing system capable of gathering very detailed information on our environment and that there should be a great potential to improve interfaces by exploiting this capability.

Sounds in HCI have mainly been used to alert users. For example, various forms of beeps and bells are used to indicate that an incorrect command has been issued or that a process needs users' attention. The use of sounds as warning indicators has also been extensive in process control plants [3]. Until recently, few computers could generate deliberately designed sounds other than beeps, but this has changed in the last few years.

As computing technologies develop rapidly, however, auditory interfaces are now considered as an indispensable element for graphic user interfaces. And more and more scholars and engineers have begun to recognize the importance of auditory interfaces. We know that auditory interfaces are sometimes annoying and confusing rather than helpful. So far, however, few studies have focused on the question of how auditory interfaces actually influence users' perceived comfortableness and performance in graphic user interfaces [3].

This study attempts to test and compare usability of three distinct types of auditory interfaces that are commonly used in personal computing environments: verbal messages, earcons, and auditory icons. "Verbal messages" consist of speech sounds (actual or machine-generated human voices), whereas "earcons" use musical sounds (usually with simple melodies) and "auditory icons" adopt natural sounds from the real world (such as sounds of cars or winds). And by "usability," we refer to effectiveness and helpfulness of computing environments.

To test the usability of those auditory interfaces, a set of graphic user interfaces were designed, and the subjects were asked to perform an identical task with the three different auditory interfaces.

** Corresponding author, Email: ynam@khu.ac.kr*

Manuscript received Nov. 22, 2012; revised Jul 30, 2013; accepted Aug 19, 2013

2. THE THREE TYPES OF AUDITORY INTERFACE

2.1 Verbal Messages

Verbal messages are an auditory interface that uses speech sounds. There are two basic ways of generating speech; concatenation and synthesis-by-rule. Concatenation uses digital recordings of a real human voice. The voice may be stored as sentences, phrases, or word segments. Later, they are played back by certain computer programs. New sentences can be constructed by arranging words in a proper order.

Synthesis-by-rule does not use recorded human voices. The synthesis of words and sentences is controlled by rules of phonemics and the contexts of sentences and phrases. Combined with a database, this method may produce a wider range of responses than speech constructed by concatenation. Synthesis-by-rule also allows various levels of pitch and tone. Speech produced by this technology may still sound somewhat "synthetic" and "machine-like." Synthesis-by-rule, however, is useful and strong where larger vocabularies are necessary[3].

Verbal messages into computer programs could enhance users' perception of usability. For example, through five experiments, Nass and Lee [4] showed that people interpret and respond to paralinguistic personality cues in computer-generated speech in the same way as they do human speech. Although the speech content was the same for all participants, when the personality of the computer (extrovert vs. introvert) voice matched their own personality, 1) participants regarded the computer voice as more attractive, credible, and informative; 2) the book review was evaluated more positively; 3) the reviewer was more attractive and credible; and 4) participants were more likely to buy the book. In this sense, we may say that adding human voices to interface will generally help users to treat computers as social actors, which will make users feel more comfortable, safe, and relaxed.

2.2 Earcons

Earcons are the short, distinctive musical motifs that have well-defined rules of construction [5]. Blattner et al. [6] define earcons as "non-verbal audio messages that are used in a computer user interface to provide information to users about computer objects, operations and interactions." There is no intuitive or intrinsic link between an earcon sound and what is represented. And the link is established by the computer interface designer and must be learned by the computer user [1].

A character of rhythm, timbre, register, and dynamic of sound makes a family that has the same category of function. Pitch identifies a constituent from its family. Combining a family with others, we may extend the meanings of earcons.

Earcons could significantly reduce the workload, duration, or mental effort involved in a task [1], [7], [8]. For example, Brewster et al. [9] found that earcons were effective, especially if musical timbres were used. And sonically-enhanced scrollbars, buttons, and windows improved usability by increasing performance, reducing time to recover from errors and reducing workload. In another study, Brewster and Catherine [10] added earcons to tool palettes to indicate the current tool and tool changes so that users could tell what was in use, wherever they were looking. The tool palettes with earcons significantly reduced the number of tasks performed with the wrong tool, as users knew that the current tool was and did not try to perform tasks with the wrong one. Crease and Brewster [11] also added earcons to progress bars to indicate the current state of the task as well as the completion of the download. The results showed a significant reduction in the time taken to perform the task in the audio condition, since the participants were aware of the state of the progress bar without having to remove the visual focus from their foreground task. Brewster et al.[9] attempted to show that compound earcons are an effective way of representing hierarchies in sound. For this, an experiment was conducted in which participants had to identify their location in the hierarchy by listening to the earcon. The results showed that participants could identify their location with more than 97% accuracy.

2.3 Auditory Icons

Auditory icons are familiar real-world sounds that have an intuitive mapping in the interface [5]. Human beings have a magical ability to identify sources of everyday sounds very accurately, up to 95% in some cases [1].For example, by just hearing the sound of raindrops on the roof, we can tell that it is raining. We often immediately understand what is going on only by hearing sounds of tearing a paper or juggling keys.

Another important property of everyday sounds is that they can convey multidimensional information. By only hearing a door slam, a listener can get multiple layers of information such as the size and material properties of the door, the force that was used, the size of the room, etc. [1].

A visual icon is being dragged to make a sound of scratching a surface and then when it is moving another folder window, another sound of scratching would be produced. For a file closing, a sound of door shutting can be attached; the size of the files can be represented by degrees of the depth of the sound.

Auditory icons could have advantages [12]-[16]. For example, Rigas, Hopwood, and Memery [17] showed "structured auditory stimuli (environmental and musical)," or auditory icons and earcons, were particularly effective in communicating "building layouts" (the number of floors, location of rooms, hallways, etc.). And Mynatt [18] attempted to show how well people can identify auditory icons. For this, subjects were asked to describe a collection of short everyday sounds. The content and accuracy of their identifications offer guidelines for the use of auditory cues.

3. HYPOTHESIS

This study analyzes auditory interfaces as a sign system. According to Peirce [19] "a sign is something standing for something to somebody in some respects or capacity." As a sign, an auditory interface also stands for an event or a message.

A computer program has many sign elements such as buttons, scroll bars, visual icons, and auditory icons and languages. They are expected to help users interact with a computer intuitively and easily. When a program's usability is high, users understand the programmer's intentions well. In this sense, when we work on a computer, there are always ongoing exchanges of signs between programmers and users. To click a mouse button or strike a keyboard, the users should interpret and understand the signs provided by the programmers. Auditory interfaces (verbal messages, earcons, auditory icons) are also signs representing certain events in computing. Each kind of auditory interface has its own references and meanings.

According to Eco [20] and Rossi-Landi [21], a sign is produced through a three-stage process. In the first stage, a ‘percept’ is produced through ‘perceiving’ material sensory data of an external object. In the second stage, a ‘sign’ is produced from the percept by the action of ‘signifying.’ Lastly, a ‘meaning’ is produced from the sign by the action of ‘interpreting.’ Relying on the theories of sign production, this study suggests that degrees of interpretability could determine the usability of an auditory interface: an auditory interface that can be more easily interpreted and understood has a higher level of usability.

H: Verbal messages would have higher usability scores than other types of auditory interface.

As computer systems and their interfaces become more complex, usability has become a key concept in HCI [3]. It is a comprehensive concept that includes safety, utility, effectiveness, and efficiency, which is measurable in terms of accuracy, time, and satisfaction with the subjective workload. Thus, it is important that system designers adequately evaluate the usability of system designs [22].

This study hypothesized that verbal messages, as “symbols,” would provide the highest level of usability, since users may use already-familiar coding schemes (natural languages) to interpret the meaning. On the contrary, auditory icons require users to establish new coding schemes to interpret the meanings of auditory icons, and therefore, they demand more mental effort from users; as a result, the auditory icons would show the lowest level of usability. Also, previous studies have supported this hypothesis because of following reasons; First of all, personal computers are generally equipped with much more powerful processors and faster sound cards. A few seconds of human voices can be processed seamlessly. Second, as Reeves and Nass [23] have shown in many studies, users tend to treat personal computers as if they were social actors. Users will perceive a higher level of usability and friendliness when they hear human voices rather than machine-produced synthetic musical sounds. Third, even though auditory icons and earcons are shorter, expressive, and intuitive, they require more mental effort to be understood, since they have their own coding scheme for interpretation. Therefore, it is hypothesized that the interfaces with verbal messages would show a higher level of usability than those with earcons and auditory icons.

4. METHOD

4.1 Participants

Forty-eight college students (juniors and seniors; 16 males and 32 females) enrolled in film theory class were recruited at a major private university. All subjects had more than 3 years of experience in computer use.

4.2 Apparatus

With Adobe’s Flash, a fake program was created that requested the following tasks: (1) calculating and clicking the OK buttons, (2) sending an e-mail message to the course instructor, (3) downloading a file from an Internet site, (4) locating a file with the Windows Explorer, and (5) clicking two series of numbered buttons in sequence. The image captures of the tasks were provided in Appendix 1.

Then, three versions of the program were made with the same visual interfaces but with three different auditory interfaces: verbal messages, earcons, and auditory icons. Verbal messages were recorded by an expert, and the voice was taken from a female professional newscaster. Earcons and auditory icons were created and edited with the KORG NS5R sound module, the MIDI program, and the sound-editing program. Voice and sounds were recorded at CD-quality (the sampling rates of 44.1 KHz). (The fake program with the three auditory interfaces and the questionnaire can be downloaded from one of the authors’ Website, but the URL was not reported here due to author’s identification.) Detailed information about the verbal messages, earcons, and auditory icons are provided in Table 1.

4.3 Procedures

The experiment was performed in a computer lab where all computers were equipped with 16-bit soundcards and headphones. Each subject completed the task three times under the three auditory interfaces, which were given in a random order. After completing each of the three phases, the subjects answered a questionnaire for the given interface.

4.4 Measures

The questions measuring usability were adopted from the Questionnaire for User Interaction Satisfaction and the NASA-TLX Task Load Index [24] and modified for this study.

In this study, usability was measured by a set of seven items: ability to inform what the program is doing, relevance to visual interfaces, degree of stimulation, degree of understandability, perceived time pressure, clearness of sound outputs, and degrees of satisfaction/frustration. Each of the seven items was measured with a 9-point scale (see Appendix 2).

Table 1. The Events and the Contents of the Three Auditory Interfaces

Events	Verbal Messages	Earcons	Auditory icons
Opening the program	“Welcome to <this> program.”	A short score	Sound of door opening
For correct answer	“Your answer is correct.”	Quadruple piano sounds (C_2, $E_2$2, $G_2$2, C_1)	Sound of applause
For wrong answer	“Your answer is wrong”	Single sound ($B^b{}_1$)	Sound of a thunder
Clicking button to log on	“Now logging on <this> program”	Church Bell sound	Sound of mouse Clicking
Sending an e-mail	“A mail is being sent”	A short score	Sound of pouring hot

	(Repeated)		water in a cup
Sending an e-mail completed	“E-mail sending completed”	Harp sound (E_2, G_2, B_2, C_1)	Sound of applause
Click the main folder to open it	“Main folder is being opened”	Single string sound (C_2)	Whooshing sound
Click the first subfolder to open it	“First subfolder is being opened”	Double string sound (C_2, D_2)	Whooshing sound (Louder than sound of main folder opening)
Click the second subfolder to open it	“Second subfolder is being opened”	Triple string sound (C_2, D_2, E_2)	Whooshing sound (Still louder than the sound of first subfolder)

5. RESULTS

Mean analyses and one-way ANOVA showed that the interface with the verbal messages achieved higher scores in all of the seven measures of usability, and the differences were statistically significant ($p < .05$) in five of them: ability to inform, relevance, understandability, clearness, and satisfaction. The results are shown in Table 2.

For the ability to inform of the process, the verbal messages were significantly higher than the earcons (F=.743, df=94, $p < .003$) and the auditory interface (F=.169, df=94, $p < .001$). But there was no significant difference between the earcons and the auditory icons (F=2.363, df=94, $p < .83$). For the relevance to the visual interface, the verbal messages were significantly higher than the earcons (F=.052, df=94, $p < .04$) and the auditory interface (F=.096, df=94, $p < .002$). But there was a weak difference between the earcons and the auditory icons (F=.357, df=94, $p < .24$). For the degrees of stimulation, there were no significant differences among the three interfaces, though the verbal messages showed relatively higher scores than the earcons (F=.290, df=94, $p < .57$) and the auditory interface (F=.043, df=94, $p < .06$). As expected, for the degrees of understandability, the verbal messages were much more higher than the earcons (F=.056 df=94, $p < .002$) and the auditory interface (F=1.540, df=94, $p < .002$). But there was no significant difference between the earcons and the auditory icons (F=1.183, df=94, $p < .75$). For the perceived time pressure, there were no significant differences among the three interfaces, though the verbal messages show still relatively higher scores than the earcons (F=1.699, df=94, $p < .68$) and the auditory interface (F=.308, df=94, $p < .34$). The length of the time for the verbal messages was actually slightly longer (1.1 to 1.5 seconds) than for the earcons and the auditory icons (about 0.5 second). For the clearness of sound outputs, the verbal messages were significantly higher than the earcons (F=2.279, df=94, $p < .04$) and the auditory interface (F=1.482, df=94, $p < .002$). But there was no significant difference between the earcons and the auditory icons (F=.558, df=94, $p < .54$). This means that users felt the verbal messages were clearest among the three interfaces, even though all the sounds had the same quality of the sampling rates of 44.1 KHz. For the degrees of satisfaction/frustration, the verbal messages were not significantly higher than the earcons (F=.540, df=94, $p < .17$), and the earcons were not significantly higher than the auditory icons (F=.554, df=94, $p < .11$), either. But there was a significant difference between the verbal messages and the auditory interface (F=.004, df=94, $p < .01$).

We also tested possible gender effects caused by the female voice of the verbal messages. However, it was found that there was no statistically significant difference between the male subjects and the female subjects for the verbal messages as well as for the earcons and the auditory icons for all of the seven measures. (The p values of the t-test ranged from .17 to .95.)

Table 2. The Results of the ANOVA

Dependent variable	Independent variable	Mean	SD
Ability to Inform of the process	Verbal messages	6.42	2.32
	Earcons	4.46	1.79
	Auditory Icons	4.33	2.31
Relevance to the visual interfaces	Verbal messages	6.16	2.07
	Earcons	4.87	1.98
	Auditory Icons	4.21	1.86
Degrees of stimulating	Verbal messages	5.63	2.28
	Earcons	5.25	1.86
	Auditory Icons	4.37	2.16
Degrees of understandability	Verbal messages	7.37	2.12
	Earcons	5.21	2.06
	Auditory Icons	5.00	2.43
Perceived time pressure	Verbal messages	5.63	2.63
	Earcons	5.33	2.23
	Auditory Icons	4.91	2.46
Clearness of sound outputs	Verbal messages	7.29	1.87
	Earcons	6.00	2.28
	Auditory Icons	5.33	2.11
Degrees of satisfaction	Verbal messages	5.63	2.28
	Earcons	4.79	1.86
	Auditory Icons	3.83	2.20

Note: Items were measured on a 9-points scale where 1 reflected the lowest possible score and 9 reflected the highest score.

6. DISCUSSION

The main concern of this study was to find which auditory interface of the three has a higher level of usability. This study hypothesized that verbal messages would offer higher usability than earcons and auditory icons, since the verbal messages are easy to interpret and understand. The hypothesis was generally supported. Though the results could not be immediately generalized, they should strongly suspect the general belief among the computer engineers that verbal messages would be less effective than auditory icons and earcons.

The results can be considered as additional support for Reeves and Nass [22]’s series of studies that are based on the general assumption that users tend to consider their computers as partners with whom they interact, rather than simple tools or an electronic apparatus.

This study provides several important implications for designing human-computer interaction. The results suggest that verbal messages are the most conventional, general, and, thus, friendly signs.

According to the sign production theory, a sign should be reproduced as something socially exchangeable and to achieve an objective stage of interpretation [22]. In this sense, interfaces with verbal messages have higher degrees of usability, since they require less mental effort than earcons and auditory icons, which involve more interpretation of abstract metaphors and emotional inspirations.

As voice recognition and text-to-speech technologies are developing rapidly, auditory interfaces with verbal messages will be much more seamless and natural, and therefore, their usability will become much greater than the usability of earcons and auditory icons.

This study, however, is not asserting that verbal messages are always superior to earcons and auditory icons. The main point is rather that sociable and humanized environments will enjoy a natural advantage over machine-like interfaces in human-computer interactions [22]. Although those results can be generalized without any modification, in some cases for routinized and simple tasks and perhaps for emergent warnings, earcons and auditory icons might be better.

REFERENCES

[1] S.A. Brewster, *Providing a structured Method for integrating Non-speech Audio into Human-computer interfaces.* PhD. Dissertation: University of York, 1994.

[2] W. Buxton, W. Gaver and S. Bly, *Tutorial number 8: The use of mon-speech audio at the interface,* In Processing of CHI' 91. New orleans: ACM Press: Addison-Wesley, 1991.

[3] J. Preece, (1994). *Human-computer interaction,* Wokingham, England: Addoson-Wesley, 1994.

[4] C. Nass and K.M. Lee, "Does computer-generated speech manifest personality? An experimental test of similarity-attraction," Proceedings of the CHI 2000 conference on Human factors in computing systems, 2000, pp. 329-336.

[5] A. PappⅢ and M. Blattner, *Dynamic presentation of Asynchronous auditory output*. ACM Multimedia 96, Boston, MA, 1996, pp. 109-116.

[6] M. Blattner, D. Sumikawa and R. Greenberg, "Earcon and icons: Their structure and common design principles," Human computer interaction, vol. 4, no.1, 1989.

[7] S.A. Brewster, A. Capriotti and C.V. Hall, "Using compound earcons to represent hierarchies," In HCI Letters, vol. 1, no.1, 1998, pp. 6-8.

[8] D.K. McGookin and S.A. Brewster, "Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition," ACM Transactions on Applied Perception (TAP) archive, vol. 1, no. 2, 2004, pp. 130-155.

[9] S.A. Brewster, P.C. Wright and A.D.N. Edwards, "The design and evaluation of an auditory-enhanced scrollbar," In B. Adelson, S. Dumais, & J. Olson (Eds.), Proceedings of CHI'94, Boston: ACM Press, Addison-Wesley, 1994, pp. 173-179.

[10] S.A. Brewster and V.C. Catherine, "The design and evaluation of a sonically enhanced tool palette," ACM Transactions on Applied Perception, vol. 2, no. 4, 2005, pp. 455-461.

[11] M.G. Crease and S.A. Brewster, "Making Progress With Sounds-The Design and Evaluation Of An Audio Progress Bar," In Proceedings of ICAD'98. Glasgow, UK: British Computer Society, 1998

[12] A.B. Barreto, A. J. Julie and H. Peterjohn , "Impact of spatial auditory feedback on the efficiency of iconic human-computer interfaces under conditions of visual impairment," Computers in Human Behavior, vol. 23, no. 3, 2007, pp.1211-1231.

[13] S.A. Brewster, *Providing a model for the use of sound in user interfaces (Technical Report No. YCS 169)*, University of York, Department of Computer Science, 1991.

[14] A. Dix, E. F. Janet, D. Gregory and B. Russell, *Human-Computer Interaction*, London: Prentice-Hall, 1993.

[15] J. A. Jacko, "The identifiability of auditory icons for use in educational software for children," Interacting with Computers vol. 8, no. 2, 1996, pp. 121-133.

[16] D.K. Palladino and N.W. Bruce, "Learning rates for auditory menus enhanced with spearcons versus earcons," Proceedings of the International Conference on Auditory Display (ICAD2007): Montreal, Canada, 2007, pp. 274-279.

[17] D. I. Rigas, D. Hopwood and D. Memery, "Communicating spatial information via a multimedia-auditory interface," In EUROMICRO Conference 1999, Proceedings 25th, vol. 2, 1999, pp. 398-405.

[18] E.D. Mynatt, "Designing with Auditory Icons," Proceedings of the Second International Conference on Auditory Display ICAD `94, Santa Fe Institute, New Mexico, 1994.

[19] C. Peirce, *Philosophical Writings of Peirce*, New York: Dover, 1955.

[20] U. Eco, *A theory of Semiotics*, Bloominton: Indiana University Press, 1979.

[21] F. Rossi-Landi, *Linguistics and economics*, The Hague: Mouton, 1975.

[22] Y. Nam and J. Kim, "Semiotic analysis of sounds in personal computers: Toward a semiotic model of human-computer interaction," Semiotica, vol. 182 , no. 1/4, 2010, pp. 269-284.

[23] B. Reeves and C. Nass, *The Media Equation*, New York: Cambridge University Press, 1996.

[24] NASA Human performance research Group, *Task load index. V 1.0 Computerized version*, NASA Arnes Resarch Center, 1987.

APPENDIX 1

Task 1. Calculation: Solve the problems and enter the answers in the blanks; then click the OK buttons. Sound output: "Your answer is correct/wrong"; Piano sounds; Sounds of applause/thunder."

Task 2. Sending an e-mail message: Complete a short regarding e-mail message to the course instructor. Sound output: “A mail is being sent”, etc.; Harp sounds; Sounds of water pouring, etc.

Task 3. Downloading a file: Click a file to download from the Net. Sound output: “The file is being downloaded,” etc.; Single string sounds; Whooshing sounds, etc.

Task 4. Locating a file: Locate a file named “thank.jpg” under the specified folder. Sound output: “The first subfolder is being opened,” etc.; Double string sounds; Smaller/louder whooshing sounds.

Task 5. Clicking the numbered buttons in sequence following the numbers given in the box. Sound output: “One, two, three, …,” Sounds of Do Re Mi…, and the Sounds of telephone number pad. If errors occur, warning messages are given (“Error,”etc.; Beep sounds, and Crashing sounds.

APPENDIX 2

Dependent variables	Questionnaires (9 point scale)
	(During doing the task, you felt that the sounds from the PC …)
Ability to Inform of the process	Provided with appropriate information for the task
Relevance to the visual interfaces	Had relevance to the visual presentations
Degrees of stimulating	Were Boring/Exciting
Degrees of understandability	Made easy for you to understand what you were doing
Perceived time pressure	Rushed you into doing the task
Clearness of sound outputs	Were delivered clearly
Degrees of satisfaction	Satisfied you for doing the task

Yoonjae Nam

He is an Assistant Professor in the Department of Culture and Tourism Contents, Kyung Hee University, South Korea. He received his Ph.D. from University at Buffalo, The State University of New York. He is interested in current digital media, in general, and corporate communication, diffusion of media contents and social networks, in specific.