Viability of Auditory Display Using Sonifications and Affecting Factors

Ada Y. Shou
Department of Electrical and Computer Engineering
University of Auckland, Auckland, New Zealand
adashou@me.com

ABSTRACT

Traditional human-computer interfaces (HCI) heavily rely on visual means of information communication, from which an increasing amount of information presented would eventually fatigue and overload users. As a result, auditory display using sonifications is proposed and investigated by many researchers, with the hope of transforming HCIs beyond the pure visual modality. "*Sonification is the use of non-speech audio to convey information.*" It has the advantages of allowing increased multitasking capabilities and mobility due to the omnidirectional nature of hearing. The field of sonification currently consists three key components: perceptual research on sonifications, tools development and application designs. This study focuses on reviewing the current status of perceptual research on sonifications. Experimental results on studying listener comprehension of sonifications are used as evidences to justify viability of using sonifications. Factors affecting listener perception of sonifications, including data types, mapping techniques, use of context, individual differences and training, are summarized and discussed in this study.

INTRODUCTION

Visual display has a long successful history, and it has been employed widely and commonly in most of the traditional HCIs. With the rapid growth of information technologies in recent years, a tremendous amount of information has been generated, stored and interpreted from various fields. The demand of presenting large quantity of information becomes a rising challenge for visual displays. Providing excessive information visually would eventually fatigue and overload users, which leads to unnecessary and costly human errors. Subsequently, auditory display came into consideration due to its integrative properties.

Auditory display uses human listening system as the primary channel and relies on sound for communicating information. Sonification is a subset and a core component of auditory display. Auditory display using sonifications has the advantage of allowing increased multitasking capabilities due to the integrative properties of sound. Users can easily monitor a number of variables through sound without attentional focus. It also permits user mobility because the omnidirectional nature of sound does require specific user orientations in relation to the display. Furthermore, auditory display makes it possible to perceive and interact for the visually impaired or when visual system is occupied. It is expected that the development of effective auditory displays will expand the HCIs beyond the pure visual modality towards some multimodal interfaces.

The field of sonification has a relatively young history. Although early researches can be traced back to the mid 20th century, studies in the field started to really grow only in the late 80s when sound card became commercially available on personal computers. Until 1999, the International Conference on Auditory Display (ICAD) members generated a sonification report [5], in which the status of the field was reviewed and a research agenda was proposed. This report initiated abundant researches in the following years, in a more systematic and cooperative manner. In the same conference, three key components of the sonification field were identified and they were perceptual research on sonifications, tools development and application designs. Perception of sonifications is of no doubt the foundation of the field. Before going into any tool or application developments, it is crucial to verify listener comprehension of sonifications and obtain thorough understandings about factors that may influence listener comprehension. This study focuses on reviewing progress and findings of the recent perceptual researches on sonifications.

LISTENER PERCEPTION OF SONIFICATIONS

Psychological studies on music cognition provided substantial evidence that human is innately sensitive to small changes in auditory dimensions, such as pitch, loudness and timbre. To make use of this nature of human hearing, the field of sonification is explored. Recent researches focuses on investigating listener comprehension of complex sonified data and patterns.

Flowers et al completed a series of experiments to investigate the perceptual differences between visual display and auditory display. One of the experiments [3] was conducted on 45 undergraduate psychology students, who had normal or corrected-to-normal vision and hearing. 18 sets of bivariate data with Pearson's correlations span the range of -0.98 to 0.98 were used. Each set of the bivariate data contained 50 data points. Visual scatter plots of these bivariate data sets were created on overhead transparencies as visual representations. To construct auditory representations of the data sets, different pitches were used to reflect data point magnitudes along the y-axis with positive polarity (high pitch for large magnitude and low pitch for small magnitude). Data points were played in order for an equal duration of 0.1s. Therefore, the total length of the audio scatter plots for a data set was 5s. 26 subjects were tested with the visual scatter plots for all 18 sets of bivariate data, while 19 subjects were tested with the auditory scatter plots. In the visual test, each overhead transparency was displayed for 10s, whereas in the auditory test, each audio clip was played twice. Pearson's correlations between judged and actual correlation of the bivariate data sets were computed. Subjects tested under the visual condition scored a mean Pearson's correlations of 0.91, while subjects tested under the auditory condition scored a mean Pearson's correlations of 0.92. The experimental results strongly suggested that judgment of visual and auditory scatter plots were highly equivalent (see Figure 1).

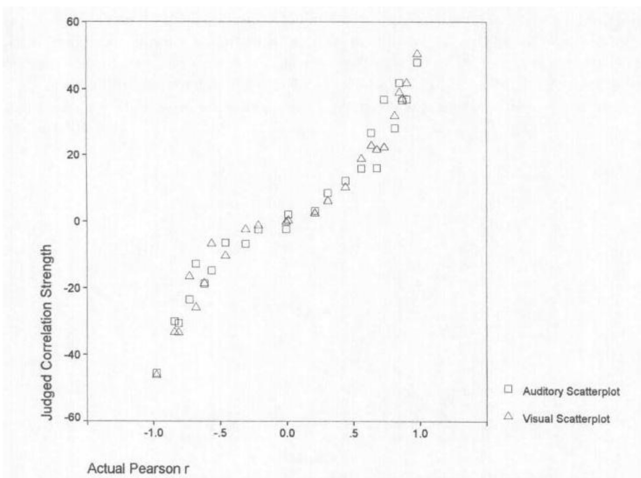

Figure 1: Mean correlation estimates as a function of actual Pearson's r for both modalities [3].

Petal and Hughes [8] completed an experiment to studied listener comprehension of sonification through visual replication (see Figure 2). In the experiments, subjects were given a series of auditory patterns and were asked to visually replicate the given patterns. As experiment result, recognition accuracy and magnitude accuracy were measured. Recognition accuracy assessed listener accuracy on recognizing a tone and indicating it visually. Magnitude accuracy assessed listener accuracy on identifying a specific tone and indicating it to the correct magnitude visually. Auditory patterns of the same length containing the same range of pitches were used. The auditory patterns were created by mapping distinct tones of same duration to different magnitudes of visual graphs. Subjects were randomly allocated into either control or experimental group. The control group was given blank sheets to replicate pattern visually for all 12 trials, while the experimental group was given blank sheets, axes, axes with grid points, for 4 trials each. The results indicated an overall recognition accuracy rate of 76% on replicating auditory patterns. The overall magnitude accuracy rate, measuring the percentage of time that a subject identified a specific tone correctly when it was recognized, was 89%.

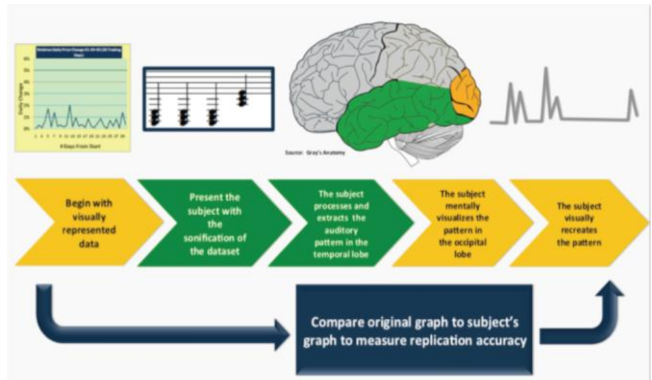

Figure 2: Comprehension Flowchart [8].

Apart from the two experiments described above, there were a number of other experiments on verifying the perception of sonifications. The overall experimental results provide substantial evidence that listeners are able to comprehend information conveyed in sonifications with a reasonably high accuracy. With data sets of certain characteristics, auditory display using sonification can perform equivalently to visual display. Therefore it is feasible to develop auditory display using sonifications, either as an alternative to the traditional visual display or towards a multimodal display.

AFFECTING FACTORS ON SONIFICATION PERCEPTION

There have been growing researches on exploring and addressing factors that may affect sonification perception in recent years. This section discusses those major affecting factors and their impacts.

Data Type

Human hearing is sensitive to changes in sounds over time. Because of this, it is often easier for listener to correctly interpret sonified data that reflecting changes in the time domain, such as trends of data, periodic and aperiodic events.

In terms of presenting the sonified data, Harrar and Stockman [4] conducted an experiment on evaluating effects of continuous (sound was presented in sine waves) and discrete (sound was presented as simple scaled note) rendering on sonified line graph perceptions among sighted and visually impaired subjects. In the experiment, subjects were instructed to visually or verbally (for the visually impaired subjects) replicate the line graphs and estimate variance of events on the x and y-axis for line graphs of different complexity. The experimental results revealed that continuous representations of the same line graph was perceived with a significantly higher accuracy, compared to the discrete representations. The study also concluded that discrete data representation was more usable in point estimation or point comparison tasks.

Human hearing is used to integrated sound environment, where different sounds occur simultaneously. Listeners have little difficulty in perceiving the combined information in sound and extracting the distinct sounds. To apply this hearing nature to the field of sonification, Brown and Brewster [1] experimented with sonifications containing two data series. The experimental results suggested that listeners were able to perceive both data series simultaneously with a high average accuracy of more than 80%. The study showed it was feasible to sonify two data series and potentially even more.

Mapping Techniques

In sonification, data mapping is the process that determines how conceptual information is translated into auditory displays. It consists of three key aspects – the selection of sound dimension, the choice of polarity and the determination of scaling. The type of data mapping used to sonify data has direct impact on the listener perception.

Various existing studies provided information on how some data dimensions (e.g. temperature, price and mass) are best fitted with certain auditory dimensions (e.g. loudness, speed and pitch). Flowers [2] summarized a number of general effective data mapping approaches. Mapping numerical data to pitch worked well even for untrained listeners; using temporal resolution, the duration of sound streams can be successfully mapped to numeric quantities; and timbre differences were useful in distinguishing several continuous data streams when used in groups. Flowers also concluded that mapping multiple continuous data variables to similar timbres and using loudness changes to represent a continuous variable were proven ineffective. More specially, Walker [11] summarized a few matching pairs that are commonly accepted as being good fits, based on their separate studies. They also pointed out that whether an auditory dimension provided a good fit to a particular data dimension depended on many factors such as training and user demographics and was hence best determined empirically through testing of potential users.

	Loudness	Pitch	Tempo	Onset
Temperature	**good**	okay	okay	poor
Pressure	okay	okay	poor	poor
Size	okay	okay	poor	**good**
Rate	okay	**good**	okay	poor

Table 1: Summary of Mapping Effectiveness [11].

Polarity of mapping determines the direction of corresponding change in sound dimension when the variable in data dimension moves in a particular direction. For instance, whether an increase in “dangerous level” is better represented by an increase or decrease in pitch? There are generally accepted polarities for mappings between certain data dimension to sound dimension (refer to table 1) but testing is strongly recommended as a way to find out with a level of certainty exactly what the polarity should be.

Scaling determines quantitatively the corresponding amount of change in auditory dimension when the variable in data dimension changes by a certain amount. An accurate scaling function is important where a quantitative estimation based on auditory graph is required. However, the “most accurate scaling function” in any particular case is dependent on each individual user’s interpretation and thus there is not a “most accurate scaling function” that applies universally. Therefore, testing is again required to determine the optimal scaling function. The optimal scaling function may very often be non-linear meaning for instance, a temperature rise from 80 degrees to 90 degrees and from 20 to 30 degrees may be represented by a different amount of pitch change.

In order to test and choose the optimized mapping, Walker [13] recommended and has successfully employed the magnitude estimation task to establish mapping relations, polarities and scaling. Ideally, subjects spanning the entire demographics and different training levels are required for undertaking the task, in order to obtain a true representation of the actual population. Subjects are asked to observe an auditory graph and estimate the value on the data dimension. Without any quantitative cues or references this estimation is dependent completely on subjects’ interpretation. The optimal mapping relations, polarities and scaling are then statistically determined from these “blind interpretations”.

Typically, when information to be sonified is of multi-dimension, 2 or more variables in the sound dimension are to be used to represent the data. This would further complicate the mapping process due to the possible interactions between different sound dimensions, in which change in one dimension affects the perception of the other. For example, Neuhoff et al [6] found that changes in pitch can influence how listener estimate changes in loudness, and vice versa. However, the interaction can also be applied favorably in auditory display. Multiple sound dimensions can be used together to reflect a single data parameter change, and this is known as redundancy mapping. Redundancy mapping is effective in emphasizing the change of variable to listeners. Peres and Lane [9] showed in their study that using integral dimensions of sound (pitch and loudness, where interaction between dimensions exists) in data mapping improved listener performance in an auditory monitoring task, in which subjects were asked to determine the status of box plots (on target, off target and skewed) based on sonified data and provided their response visually through buttons. Whereas using separate dimensions (pitch and tempo) in data mapping showed no differences, compared with when only a single auditory dimension was used.

Use of Context

It is of no doubt that the suitable use of context benefits visual representations of data. To extend to the sonification field, the effectiveness of use of context was investigated with sonified data. Context information includes the labeling of x-axis, y-axis and gridlines in the graph. Information of interest (whether we are interested in the general trend or values of some specific points in the graph), data types (continuous or discrete) and mapping techniques (the way context information is presented) are all expected to influence on the success of the use of context. However, current researches are limited and not all those factors have been explored in applying context in sonification. Smith and Walker [10] studied the effect of adding context to auditory graphs and concluded that adding useful context helps to enhance the user perception of auditory graphs.

Providing x-axis context could help user to identify current location relative to the complete data block, this is particularly useful if the information of interest are concentrated in a particular section within the data block. The user needs to first locate the section on the x-axis before attempting to translate the sound played in that section into the wanted data. This axis context is usually added as a short clicking sound after every interval of a certain length (e.g. gridlining the x-axis).

In the scenario where the data series is of a continuous and varying nature, it can be divided into arbitrary blocks within each of which the data variable (pitch, speed etc.) is averaged to a constant level. This process practically converts the data from continuous to discrete. With the data blocks being of reasonable length and each of them being of different value from its neighbors, x-axis context is effectively constructed. Users can simple grasp the x-axis context by counting the number of blocks. This eliminates the need of adding extra clicks to provide x-axis context.

Providing y-axis context could help users grasp the relative values of data, this could be the y-value of some particular points (starting point), maximum or minimum within the data series. It is usually implemented by predefining particular points on y-axis to distinctive tones. This approach effectively provides the users with gridlines on the y-axis from which data values can be more easily estimated.

Nees and Walker [16] compared different loudness of the auditory context with respect to that of the sonified data and concluded that context was most effective when its loudness was either higher or lower than that of the sonified data.

Compared with visual displays, providing context information in auditory graphs is not as straightforward as drawing and labeling x-axis and y-axis and the use of the context may not be as intuitive. As discussed previously in the example of providing x-axis context, user needs to count the number of clicks (or counting the number of discrete blocks) in order to grasp the x-axis. In visual display, this is easily achieved by labeling quantitative numbers along the axis. More research is needed to explore how to quantitatively label auditory graphs.

Individual Differences

Individual limitation, capability and experiences are believed to have impact on comprehension of sonification. From the limited researches that have been completed, effects of individual's physical constraints, cognitive ability, musical ability and demographics on comprehension of sonifications were studied preliminarily.

Walker and Lane found in their experiment that in some situations, visually impaired individuals may respond differently to polarity of sonified data, compared to sighted individuals [12]. For example, sighted individuals preferred the use of positive polarity to map the "number of dollars" variable to frequency, whereas visually impaired individuals preferred the opposite polarity. Neuhoff et al [6], in their experiment determining appropriate mapping of pitch change to changes of variable, found individuals with musical expertise scaled pitch change differently than individuals with no musical experiences. Walker and Mauney [14] completed a specific experiment to study effect of individual differences on comprehension of sonification. Subject cognitive ability, including working memory ("the system which actively holds information in the mind and to make it available for further information processing") and spatial reasoning ("ability to visualize spatial patters and mentally manipulate them over a time-ordered sequence of spatial transformations"), and demographics, such as gender, age, handedness and musical experiences were assessed and judged in correlation with their performance on magnitude estimation of sonified data. The experimental results suggested that individual working memory capacity and gender seemed to have substantial influence on comprehension of sonified data, although the test results were not completely consistent. However, musical experience was not seen as an effecting factor, which conflicted with Neuhoff's finding. In Petal and Hughes' experiment discussed earlier, demographical analysis of the experimental results revealed similar gender difference, as males outperformed females.

Although the preliminary results from limited studies in individual differences supported that the differences in individuals have impact on comprehension of sonifications, researches employing more appropriate testing methodologies are required to obtain consistent findings.

Training

Unlike visual display, where applications are pervasive and the user cognition is well established, auditory display is relatively unfamiliar to most of the users. Early research in the psychological field indicated that perceptual performance can be improved by proper training. Therefore, training is identified as one of the factors that can benefit novice users of auditory display. Recent researches focused on investigating the effect of different training methods, mainly divided as conceptual training and perceptual training.

Conceptual training involved analyzing and breaking down a task and offering strategies for each of the steps. Walker was the first to experiment the conceptual training, with a point estimation task through auditory display. In their experiment, subjects were pretested and divided into training and no-training groups. After either a conceptual training session or a task filler session, subjects were tested again with the point estimation task. The experiment results suggested conceptual training made a statistically significant improvement on user accuracy with a point estimation task using sonified data. However, the effect of conceptual training soon became debatable as later experiment completed by Walker and Nees [15] failed to replicate the same results. Walker and Nees experimental results revealed conceptual training improved user accuracy but the improvement was lack of statistical significance. The effect of conceptual training on perception of sonified data remained inconclusive.

Perceptual training involved intensive practice with certain stimuli with correct response provided. Classic perceptual training methods included the use of prompting and feedback. With prompting, a cue of correct response to a stimulus is provided before or during the presentation of the stimulus. With feedback, the correct answer is revealed after user makes a response to a stimulus. Early studies showed mixed results on whether there were differences between prompting and feedback when used in training to aid sonified data perception. In Walker and Nees [15] recent experiment evaluating effect of conceptual and different perceptual training methods, the results indicated that a brief training session (20 minutes) using feedback method significantly improved user performance with point estimation task using sonified data. On the other hand, training using prompting demonstrated some improvements, but the improvements were less than the feedback approach. There is no doubt that perceptual training can improve listener comprehension of sonified data, but further work is required to determine an optimal training approach.

SUMMARY

Existing researches on sonification perception supported that listeners were able to comprehend sonifications with high accuracy and with specific data sonification can perform equivalently to visual display. Studies on the perceptual process of sonified data identified data type, various data mapping techniques, individual differences and training as the factors that may affect the perception of sonifications. The human hearing nature determines that applying sonification to data reflecting time domain behaviors is more effective. Selecting the correct data mapping techniques is of great importance to ensure the correct perception of sonified data. It is beneficial to follow the successful techniques identified to sonify certain data. Individual differences in cognitive ability, music experiences and gender are likely to have impact on comprehension of sonifications, but their impact can be reduced through effective perceptual training.

FUTURE WORK

Research in perception of the auditory display using sonifications has progressed from studying perception of complex and dynamic sonified data to a profound study aiming to identify factors influencing sonification perception and understanding their impacts. A number of important affecting factors have been successfully addressed. In particular, understanding of appropriate mapping techniques and the proper use of context in sonified data have great potential to benefit designs of sonification tools and applications. However, due to the lack of sufficient researches it is still not possible to define a general sonification design guidelines. Some of the existing findings cannot or have not been reproduced through experiments, further studies are required. Various tasks were used by different research groups in the existing studies, which make direct comparison of results difficult. Most of the affecting factors have been studied in isolation, benefit or influence of combined affecting factors has not been studied thoroughly.

In future, more experiments employing various tasks are suggested to verify some of the existing findings and to validate the findings for different sonification tasks. The affecting factors shall be studied in combination in order to create some general sonification design guidelines that are suitable for a range of data series and tasks.

REFERENCES

1. Brown, L.M. and Brewster, S.A. Drawing by Ear: Interpreting Sonified Line Graphs. In *Proc. of the 2003 International Conference on Auditory Display*, (2003), 152-156. http://www.icad.org/Proceedings/2003/BrownBrewster2003a.pdf
2. Flowers, J.H. Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. In *Proc. of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, (2005), 406-409. http://www.icad.org/Proceedings/2005/Flowers2005.pdf
3. Flowers, J.H., Buhman, D.C. and Turnage, K.D. Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples. *Human Factors*, (1997), 341-351. http://www.ncbi.nlm.nih.gov/pubmed/9394628
4. Harrar, L. and Stockman, T. Designing Auditory Graph Overview: an Examination of Discrete vs. Continuous Sound and the Influence of Presentation Speed. In *Proc. of the 13th International Conference on Auditory Display*, (2007), 299-305.

http://www.icad.org/Proceedings/2007/HarrarStockman2007.pdf

5. Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J.H., Miner, N. and Neuhoff, J. Sonification Report: Status of the Field and Research Agenda. *International Community for Auditory Display (ICAD)*, (1999). http://www.icad.org/node/400
6. Neuhoff, J.G. and Wayand, Joseph. Pitch Change, Sonification, and Musical Expertise: Which Way is Up? In *Proc. of the 2002 International Conference on Auditory Display*, (2002). http://www.icad.org/Proceedings/2002/NeuhoffKnight2002.pdf
7. Neuhoff, J.G. The Sonification Handbook. 1st *Edition*, *Chapter 4*, (2011), 63-85. http://sonification.de/handbook/index.php/chapters/chapter4/
8. Patel, N.S. and Hughes, D.E. Studying Listener Comprehension of Sonifications through Visual Replication. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, (2011). http://ntsa.metapress.com/link.asp?id=1412w481543368w2
9. Peres, S.C. and Lane, D.M. Auditory Graphs: The Effects of Redundant Dimensions and Divided Attention. In *Proc. of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, (2005), 169-174. http://www.icad.org/Proceedings/2005/PeresLane2005.pdf
10. Smith, D.R. and Walker, B.N. Tick-marks, Axes, and Labels: the Effects of Adding Context to Auditory Graphs. In *Proc. of the International Conference on Auditory Display*, (2002). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.196.7184
11. Walker, B.N. Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Transactions on Applied Perception, Vol.2*, (2005), 407-412. http://sonify.psych.gatech.edu/publications/pdfs/2005TAP-WalkerKramer-ICAD1996paper.pdf
12. Walker, B.N. and Lane, D.M. Psychophysical Scaling of Sonification Mappings: A Comparison of Visually Impaired and Sighted Listeners. In *Proc. of the International Conference on Auditory Display*, (2001), 90-94. http://www.icad.org/Proceedings/2001/WalkerLane2001a.pdf
13. Walker, B.N. Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification. *Journal of Experimental Psychology: Applied, Vol8(4)*, (2002), 211-221. http://psycnet.apa.org/journals/xap/8/4/211/
14. Walker, B.N. and Mauney, L.M. Individual Differences, Cognitive Abilities, and the Interpretation of Auditory Graphs. In *Proc. of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, (2004). http://www.icad.org/Proceedings/2004/WalkerMauney2004.pdf
15. Walker, B.N. and Nees, M.A. Brief Training for Performance of a Point Estimation Sonification Task. In *Proc. of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, (2005), 276-279. http://www.icad.org/Proceedings/2005/WalkerNees2005a.pdf
16. Walker, B.N. and Nees, M.A. Relative Intensity of Auditory Context for Auditory Graph Design. In *Proc. of the 12th International Conference on Auditory Display*, (2006), 95-98. http://www.icad.org/Proceedings/2006/NeesWalker2006.pdf