Listener Comprehension and Reaction-Time Study Of Sonification

Neel S. Patel and Darin E. Hughes
Media Convergence Laboratory, Institute for Simulation and Training, University of Central Florida
Orlando, Florida
npatel@ist.ucf.edu, darin@cs.ucf.edu

ABSTRACT

Sonification is the use of non-speech audio, encompassing the parameters of pitch, intensity, and tempo, to represent patterns and trends in data. Sonification becomes increasingly relevant as visual senses become overloaded, fatigued, or impractical for information communication (combat scenarios, large dataset exploration, etc) due to advantages of audio communication over visualization (immersive, not limited by device size). Creating a theory of successful sonification design based on user audience was the purpose of this study, which focused on understanding the key parameters that drive the comprehension of and reaction time to sonifications that are a reasonable alternative to the same information being presented through visualization.

Two experiments were formed, testing listener comprehension and listener reaction time. The comprehension test involved matching patterns of data presented through sonification to a visual counterpart. Findings show that untrained subjects are accurately able to sonification to the corresponding visualization over 60% of the time, but are unable to map the temporal location of a pattern presented through audio to its spatial location on a corresponding visual representation. The reaction time test involved assessing how quickly subjects were able to recognize and react to a pattern of tones within a background audio stream. Findings show that accurate recognition and reaction time to a 5-tone pattern is noticeably higher that for a 1- or 9-tone pattern. The researchers found that the larger the temporal gap between the relevant patterns inserted into the background noise, the higher the likelihood that the subject would recognize and react to the pattern.

This study starts to lay the foundation for how to successfully sonify data that is currently presented in visual form, and holds the potential of driving breakthrough changes in human-computer interfaces similar in impact to those realized by moving from character based screens to graphical user interfaces.

ABOUT THE AUTHORS

Neel S. Patel is a research associate in the Media Convergence Laboratory at the University of Central Florida's Institute for Simulation and Training Neel's primary focus of research over the last 3 years has been in understanding the human perception of auditory information. His most recent work in the area of applied sonification explores ways in which presenting non-voice information through sound can be used as an effective alternative to presenting that same information through visual displays with a specific focus on understanding factors around human perception and interpretation of the information being presented.

Darin E. Hughes is research faculty at the University of Central Florida's Institute for Simulation and Training. He is the lead audio researcher and sound designer for the Media Convergence Laboratory. His research interests include auditory perception, spatial audio, sonification, and biofeedback. He has worked on simulations that have been funded by the US ARMY, NAVY, and Airforce, the National Science Foundation, and many others.

Listener Comprehension and Reaction-Time Study Of Sonification

Neel S. Patel and Darin E. Hughes
Media Convergence Laboratory, Institute for Simulation and Training, University of Central Florida
Orlando, Florida
npatel@ist.ucf.edu, darin@cs.ucf.edu

INTRODUCTION

Visual fatigue is an increasingly common problem in situations where large quantities of information are being communicated to observers through the use of multiple and complex visual displays. In many cases, the observers are expected to comprehend and react to this information in a timely manner. Visual overload and visual fatigue can lead to undesirable outcomes and in extreme conditions can mean the difference between life and death. Creating reasonable alternatives to visualization in human-computer interfaces is becoming an important area of study. The immersive nature of sound, particularly in comparison to sight, has many advantages that make sonification a viable alternative to visualization. Encompassing the parameters of pitch, intensity, and tempo, to represent patterns and trends in data, sonification can become increasingly relevant as visual senses become overloaded, fatigued, or impractical for information communication.

With the growing limitations of visualization, developing alternative means of information communication becomes crucial in situations ranging from armed forces in combat to stockbrokers identifying dynamic market trends. This paper presents ***sonification, the use of non-speech audio to convey information or identify patterns in data, as a viable alternative to visualization***, (see Figure 1) as well as outlines a theory for sonification design based upon user audience.

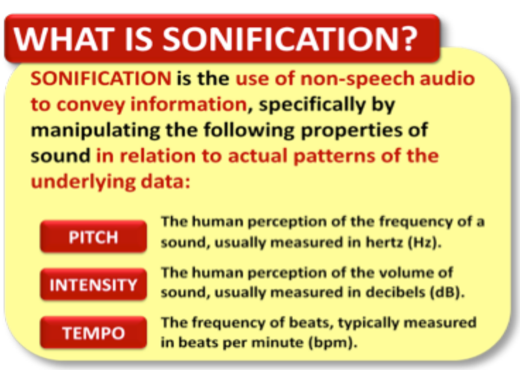

Figure 1. Breakdown of sonification

The underutilization of the auditory cortex, used for processing information communicated through sound, as well as the immersive nature of sound, which allows 360° mobility and multitasking prohibited by traditional visual displays, serve as grounds for the use of sonification in human-computer interfaces. This approach has the potential to both eliminate the aforementioned visual fatigue by using alternate modalities of communication and also to allow for information communication in areas previously deemed impossible.

STATE OF THE FIELD

With the focused, in-depth study of sonification having been pioneered only 20 years ago, sonification is a relatively young field of science. The field has yet to include rules and guidelines for designing sonifications, and the majority of experimental sonifications created thus far have been either scenario-specific or ad-hoc, without applying to a general format of design.

Auditory Displays – A Precursor To Sonifications

The distinction between different types of auditory displays was made by Herman (Herman, 2002). Auditory displays are typically defined as any transmission that leads to audible perceptions for the user, including speech interfaces and alarms or other auditory notifications. Sonification is distinguished by the specific mapping of changes in a dataset to auditory changes, and can be further broken into model-based sonification and parameter-based sonification. Model-based sonification involves using data to design a virtual 'instrument' that can be played to explore a dataset. This study focused instead on parameter-based sonification, in which a specific data value is mapped to acoustic attributes of a sound (Ritter, 1999).

Auditory displays have gained prominence in many practical applications that would be otherwise limited by traditional visual displays. The primary example is in alarm notifications, which communicate to an observer that a specific event has occurred, allowing the observer to act upon that event. Two more advanced uses, the pulse-oximeter and the Geiger counter, are examples of auditory displays commonly used on a daily basis. Both allow for the communication of information without necessitating the user to continually watch a visual display, and are very effective at signaling the user of a change occurring (Hermann, 2002). While too rudimentary to qualify as sonifications, these auditory displays demonstrate the potential applicability of sonification and sound for purposes more complex than alarming a user of change taking place in the data.

Early Sonification Research

The potential viability of communicating information through sound as an alternative to visual communication was demonstrated by Pollack and Ficks in their 1954 experiment, a precursor to modern sonification research. The experiment tested the use of audio variables to represent simple changes in quantitative information, and the results of the study indicated that multiple dimensions of sound used in conjunction with one another were more effective than a uni-dimensional display (Kramer, 1994). Chambers et al. (1974) demonstrated the first evidence of the positive impacts of sonification in human comprehension by supplementing visual graphs with audio. Bly's 1982 experiments showed that comprehension of auditory displays was as accurate as comprehension of visual displays. Furthermore, comprehension of the two combined was significantly more accurate than comprehension of either alone. These research papers formed the basis upon which the core theory of sonification has been built since 1987.

A study by Kramer et al. (1997) outlined current trends in sonification and areas lacking in development, finding that establishing practical and real-world applicable uses of sonification was a neglected area. Instead, research continued, and continues, to focus on either theoretical or extremely narrow/specialized sonification, abandoning widespread application of sonification. Before the practical application of sonification can be realized, it is necessary to create a ***theory for sonification design*** that takes into account situation-specific (the conditions under which sonification is being applied) and demographic (target audience for the sonification) characteristics of the target audience.

Advantages and Importance of Sonification

Several properties of sound give audio information communication, or sonification, advantages over traditional visual communication. Sonification can "provide background information about changed states without a major disruption of attentional focus" (Watson and Sanderson, 2004). Sonification can be used to present high-dimensional data without creating information overload for users as well as allow simultaneous background (peripheral, rather than direct observational) tracking of changing temporally complex variables. The advantageous properties of sound can be summarized as follows (Kramer et al., 1997):

1. Sound is immersive, and allows information communication without having to directly observe a visual display
2. Sound allows for information communication without overloading the visual sense (multimodal information communication)
3. The sense of hearing is particularly attuned to temporal changes, making dynamic trends easy to notice

GOALS OF STUDY

As outlined earlier, gaps exist in the current state of sonification knowledge in terms of a theory of sonification design and factors that affect the use or design of a sonification. This study focuses on ***the conditions under which sonification is a viable alternative to visualization*** in terms of two metrics – comprehension of patterns represented in the sonification, and reaction time to specific patterns of sound. The study aimed to assess these metrics in untrained listeners in order to determine the factors that drive or ***influence a listener's ability to accurately comprehend and quickly react to a sonification***, in order to create the aforementioned theory for sonification design.

COMPREHENSTION TEST

The first of the two experiments conducted as part of the project studied **listener comprehension of sonifications,** typically via identification of a pattern represented through sonification and conversion to a pattern represented through visualization. The design of the test involved exposing a user, for 12 trials, to a sonification and four visual graphs, one of which represented the same pattern as the sonification. Accuracy was defined as the number of times the user correctly identified the graph divided by the total number of trials. See Appendix A for detailed procedures.

Design Considerations

A metric called the Visual Gap Index (VGI) was used to quantify 'difficulty' of each trial and allow for comparisons across multiple trials (see Figure 2)**.**

It was premised that the more alike two graphs looked, the more difficult it would be to differentiate between them, while the greater the visual distinction between two graphs, the easier it would be to distinguish between them. The VGI formula estimates the visual difference between two graphs by comparing the relative locations of individual points along each graph. A low VGI indicates two very similar graphs, whereas a high VGI indicates more distinct graphs. For each trial, the total VGI was calculated by adding the VGI of each of the three incorrect graphs, allowing for comparisons of difficulty across trials. Additionally, comparison of difficulty within a trial could be estimated by the relative VGI scores of the three incorrect graphs to the correct graph.

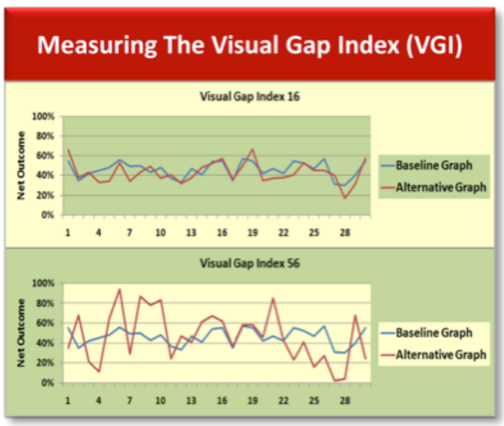

Figure 2. Illustration of the Visual Gap Index

Human Subject Pool and Testing Procedure

Seventy five (n=75) subjects were exposed to each of the twelve trials of the comprehension test. Demographic data covering age, gender, musical experience, and hearing impairments was collected. Tests were administered in a computer-lab setting, with each subject at an individual computer with an individual set of headphones. The order that each of the 12 trials were presented in was randomized as to eliminate the chance of subjects sitting adjacent to one another sharing answers.

Results

The results indicated an overall accuracy rate of 60% across all demographic groups. A set of 75 results was randomly generated by a computer such that for each trial, the computer would randomly select one of the four possible graphs, and repeat this for all 12 trials/all 75 iterations. The overall accuracy of the human results were compared to that of the randomly generated results, and **it was found that there was a statistically significant difference between the two datasets (p<0.001),** with the human results having a significantly higher comprehension accuracy rate relative to what would be expected from random guessing or random chance.

ANOVA: Single Factor

Groups	Count	Sum	Average	Variance
Computer Simulation	75	13.083	0.242	0.015
Human Subjects	75	32.167	0.596	0.026

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3.372	1	3.372	162.186	0.000	3.931
Within Groups	2.204	148	0.021			
Total	5.576	149				

Figure 3. Analysis of human and computer trials

The results were next analyzed in terms of VGI scores. It was revealed that the higher the VGI score (and the more dissimilar the graphs were), the higher the overall accuracy for that trial was. Correspondingly, the lower the VGI score (and the more similar the graphs), the lower accuracy was for that trial (see Figure 4). **This supports the reasoning that when presented with a sonification, a subject will extract the key pattern from the audio, and attempt to find the closest match to that pattern in one of the four visual graphs, giving insight into the exact process of comprehending a sonification**.

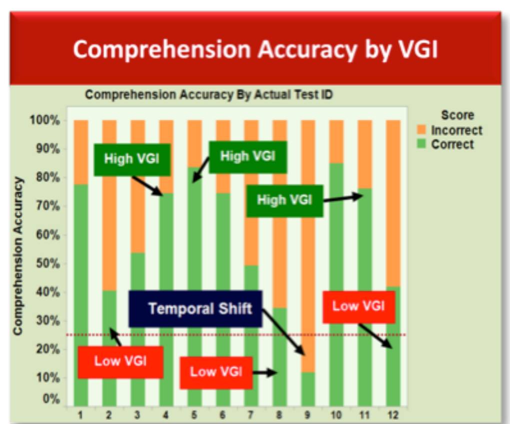

Figure 4. Comprehension accuracy by VGI

One trial was specifically designed to test the phenomenon of **temporal dislocation** (see Figure 5). In this trial, the same key pattern was inserted into each of the four graphs, but at a different location in each graph. This pattern was also present in the sonification. To correctly choose the corresponding graph, the subject would have to identify both the pattern and its relative temporal location, or spacing relative to other sound patterns, and translate that to its relative spatial location (to other patterns in the graph). A 12% accuracy rate was found for this trial, showing poorer performance than would be expected from guessing.

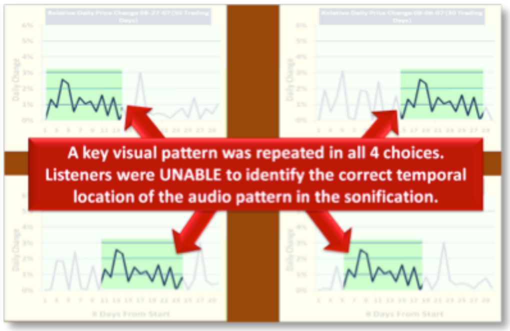

Figure 5. Illustration of temporal dislocation

It was found that there was a statistically significant (p<0.01) difference between the younger (age<25, 65% accuracy) and older (age>40, 55% accuracy) demographic groups, with the younger demographic group showing a higher overall accuracy rate than the older group (see Figure 6). Male subjects had a similarly higher accuracy rate than females, and the subject group without any hearing impairments showed higher overall accuracy than those with hearing impairments. Prior musical experience seemed to have no impact on the comprehension accuracy rate.

Comprehension Test - ANOVA By Demographic Factors

	Groups	Comprehension Accuracy	P-value
By Age Demographic	>40 yrs	54%	0.009
	<25 yrs	66%	
By Gender Demographic	Male	69%	0.011
	Female	56%	
By Age/Gender Demographic	Female <25	64%	0.017
	Male <25	70%	
	Female >40	53%	
	Male >40	56%	
By Musical Experience	No Musical Experience	60%	0.905
	Musical Experience	61%	
By Hearing Impairment	No Hearing Impairment	63%	0.006
	Hearing Impaired	42%	

Figure 6. Analysis by demographic factors

REACTION TIME TEST

The second of the two experiments conducted involved training subjects with a specific pattern of sounds (either 1, 5, or 9 tones in length), and having them identify each time their specific pattern of sounds was played within a larger 20-minute stream of sounds. The goal of this experiment was to study both the speed and accuracy of a user reacting to a specific auditory pattern of tones, as well as study demographic differences that affect this speed or accuracy. See Appendix B for detailed procedures.

Design Considerations

A program was created to randomly insert a 1, 5, or 9 tone pattern into a 20 minute stream of 'background' tones a total of 50 times. The notes were created such that the patterns were 1 octave higher (had a pitch of roughly 440 Hz higher) than the ambient sound. In order to identify a pattern, the subject had to indicate (via a mouse click) recognition of the pattern within the tone recognition window defined as the time from the start of the pattern to 1 second after the termination of the pattern (see Figure 7).

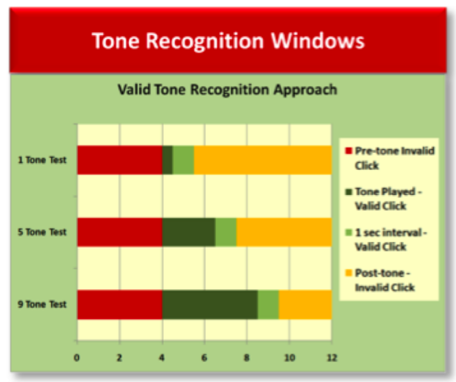

Figure 7. Tone recognition window illustration

Human Subject Pool and Testing Procedure

The same subjects as well as the same settings were used in both experiments. 1/3 of the subjects were exposed to the 1 tone pattern, 1/3 to the 5 tone long, and 1/3 to the 9 tone long pattern, allowing for comparisons between various pattern lengths and an equal split between the three different patterns.

Results

The results indicate that subjects were able to identify the tone pattern within one second of it being played 82% of the time (see Figure 8). It was found that accuracy was highest for the 5 tone pattern, and lower for the 1 and 9 tone patterns, indicating that a mid-length pattern is easy to quickly identify, whereas a shorter pattern is easy to miss and a longer pattern is too long to allow for quick identification. A regression analysis of the accuracy of the 1, 5, and 9 tone trials revealed negative slopes indicating a decrease in reaction time accuracy over time (see Figure 9). This can be attributed to auditory fatigue, or the sounds becoming monotonous to the user over the 20-minute sample. The important results, however, come from the fact that the onset of auditory fatigue (the decrement in performance as the user loses their ability to pick up key auditory changes) was most gradual for the 5 tone sample, and much steeper for the 1 and 9 tone samples. This indicates that using a mid-length pattern can help alleviate the issue of auditory fatigue.

	Tone Present	Tone Not Present
User Reaction Within 1 sec	Correct Identification (Hit) <25 yrs: 76% >40 yrs: 86%	Incorrect Identification (False Hit) <25 yrs: 14% >40 yrs: 7%
No User Reaction	Incorrect Rejection (Miss) <25 yrs: 25% >40 yrs: 14%	Correct Rejection

Figure 8. Listener reactions and results

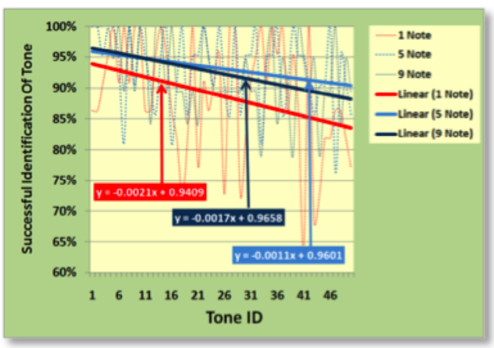

Figure 9. Regression analysis by pattern length

Data was also analyzed in terms of the 'density' of successive patterns being inserted, or how close successive patterns were to one another. It was revealed that as the timing between successive patterns decreased, so did performance, and the onset of auditory fatigue was more rapid. In contrast, as the space between successive patterns increased, performance and accuracy increased as well, and auditory fatigue was slower to appear. This indicates that when designing a sonification that needs to be reacted to quickly and accurately, sufficient space should be left between successive patterns to avoid the impact of auditory fatigue.

Once again, results were analyzed in terms of demographic groups. In contrast to the comprehension test, where younger demographic groups outperformed older groups, the older demographic groups (age>40) exhibited significantly higher performance (in terms of higher accuracy, fewer missed tones, and fewer 'false alarms' of identifying a tone when none was present) than younger (age<25) demographic groups for the reaction time test. Other demographic factors, such as gender, musical experience, or hearing impairments, did not show any statistically significant difference between groups (see Figure 10).

Reaction Time Test - ANOVA By Demographic Factors

	Groups	Reaction Accuracy	P-value
By Age Demographic	>40 yrs	86%	0.024
	<25 yrs	76%	
By Gender Demographic	Male	81%	0.623
	Female	83%	
By Age/Gender Demographic	Female <25	74%	0.150
	Male <25	77%	
	Female >40	86%	
	Male >40	89%	
By Musical Experience	No Musical Experience	85%	0.389
	Musical Experience	80%	
By Hearing Impairment	No Hearing Impairment	82%	0.423
	Hearing Impaired	88%	

Figure 10. Analysis of demographic factors

Conclusions and Recommendations

In conclusion, it turns out that a **user really can 'see what they hear'**. This research has shown that the human brain can comprehend and correlate information in audio patterns to visual patterns with a high degree of accuracy, effectively opening the doors to the use of sound as an alternative to visual displays to communicate information in human-computer interfaces. It was demonstrated that people can visualize patterns communicated by a series of tones and successfully match to their visual counterparts, as well as react to these patterns with a high degree of accuracy and short reaction time. The onset of auditory fatigue, and its differences across tonal patterns, has been demonstrated, as have specifics of sonification comprehension, including the nature and granularity of temporal-to-spatial data translation. The study also shows that results can vary based on underlying demographic factors such as age and gender, effectively creating a theory for sonification design.

To extend and apply that theory, three further extensions to this research are proposed. The first is to study the impact of training and repeated exposure to sonifications on comprehension and reaction time. The next step would be field-testing sonification in a real-life situation or environment to create a model for the practical application of sonification, accounting for factors specific to the scenario. The final step would be generalizing that model for use in all human-computer interfaces.

APPENDIX A – COMPREHENSION TEST PROCEDURES

1. Create a total of 12 test trials, each with:
   - *4 visual line graphs, labeled A, B, C and D*
   - *1 sonification of one of the four graphs used*
   - *The object of the trial to correctly match the sonification to its corresponding visual representation.*
2. Trials should have differing degrees of closeness, with closeness being defined as the Visual Gap Index (VGI), which assesses the visual difference between two graphs such that as VGI increases, visual difference between the two graphs increases as well. Trials should measure specific aspects of sonification comprehension such as granularity of temporal-to-spatial data translation.
3. Collect a large ($n > 60$) sample of human subjects spanning (with relatively equal distributions between) multiple demographic factors, including age, gender, and musical experience.
4. Have each subject fill out a demographic sheet, including the fields of: age, date, musical experience, prior sonification experience, hearing and visual impairments, etc, as well as a Parental/Informed Consent Form
5. Expose the subject to the Comprehension Test created in steps 1-3. Data collection should include, for each trial:
   - *Letter of graph that corresponds to the sonification*
   - *Number of times the sonification was played*
6. Analysis should include overall accuracy across trials and per trial, as well as most frequent incorrect responses and their proximity scores.

APPENDIX B – REACTION TIME TEST PROECEDURES

1. Create an algorithm to generate a 20 minute stream of random tones all in the 4th octave (frequency = 260 Hz) at the rate of two tones or beats per second (each tone lasts .5 seconds).

2. Create randomly generated patterns of 1, 5, and 9 tones in length in the 6th octave (frequency = 1050 Hz) at the rate of two tones or beats per second.

3. Randomly insert the 1 tone pattern into the original 20 minute set of random tones 50 times. This is the One Tone Test. Repeat with the 5 and 9 tone patterns to create the Five and Nine Tone Tests.

4. Collect a large ($n > 60$) sample of human subjects spanning (with relatively equal distributions between) multiple demographic factors, including age, gender, and musical experience.

5. Have each subject fill out a demographic sheet, including the fields of: age, date, musical experience, prior sonification experience, hearing and visual impairments, etc, as well as a Parental/Informed Consent Form

6. Choose the 1, 5, or 9 tone test to administer to the subject, and train the subject by playing the 1, 5, or 9 tone pattern for them 5 times. Instruct them that they will be asked to react as quickly as possible each time they hear their pattern within a larger 20 minute stream of sounds. This step should be carried out as to ensure as even a distribution as is possible between the 1, 5, and 9 tone tests.

7. Play the appropriate One, Five, or Nine Tone test for the subject, having them click a button that records the timestamp each time they hear their specific tone.

ACKNOWLEDGEMENTS

The authors would like to thank the entire Institute for Simulation and Training at the University of Central Florida for its assistance, with particular acknowledgement to the Media Convergence Laboratory, Dr. Charles E. Hughes, and Dr. Stephen M. Fiore. They would also like to acknowledge the funding assistance of RDECOM.

REFERENCES

Baars, Bernard J., Banks, William P. & Newman, James B. (2003). Essential Sources in The Scientific Study Of Consciousness. The MIT Press.

Bly, S. (1982). Sound And Computer Information Presentation - Doctoral Dissertation. University Of California, Davis, California.

Brewster, Stephen A. (1996). A Sonically Enhanced User Interface Toolkit. Proceedings of the International Conference On Auditory Display.

Chambers, J. M., Matthews, M. V., & Monroe F. R. (1974). Auditory Data Inspection. Technical Report. AT&T Bell Laboratories, Murray Hill, New Jersey.

Chang, Irene & Boulanger, Pierre (2005). A 3D Perceptual Metric Using Just-Noticeable Difference. Eurographics. 97-100.

Hermann, Thomas (2002). Sonification for Exploratory Data Analysis - Doctoral Dissertation. Neuroinformatics Group, Bielefeld University, Germany.

Hermann, T. & Ritter, H. (1999). Listen To Your Data: Model-Based Sonification For Data Analysis. Int. Inst. For Advanced Studies In System Research And Cybernetics. 189-194.

Kramer, Gregory (1994). An Introduction to Auditory Display. Auditory Display, G. Kramer Ed., International Conference on Auditory Display. pp. 1-79.

Kramer, Gregory et. al. (1997). Sonification Report: Status of The Field And Research Agenda. National Science Foundation.

Levintin, Daniel P. & Cook, Perry R. (1996). Memory For Musical Tempo: Additional Evidence That Auditory Memory Is Absolute. Jour. Perception and Psychophysics 58. pp. 927-935.

Lodha, Suresh K., Wilson, Catherine M., & Sheehan, Robert E. (1996). Listen: Sounding Uncertainty Visualization. Proceedings Of The 7Th Conference On Visualization. pp. 189-195.

Minghim, R., & Forrest, A.R. (1995). An Illustrated Analysis Of Sonification For Scientific Visualization. Proceedings of The 6th IEEE Visualization Conference (Visualization '95). pp. 110-117.

Nesbitt, Keith V. (2001). Modeling The Multi-Sensory Design Space. Proceedings of The 2001 Asia-Pacific Symposium On Information Visualization, Volume 9. pp.27-36.

Pollack, I. & Ficks, L. (1954). Information Of Elementary Multidimensional Auditory Display. Journal Of Acoustic Soc Amer 36. pp. 155-158.

Song, Hong Jun (2008). Aesthetic And Auditory Enhancements For Multi-Stream Information Sonification. 3Rd International Conference On Digital Interactive Media In Entertainment And Arts. pp. 224-231.

Speeth, Sheridan (1961). Seismometer Sounds. Journal Of Acoust Soc Amer 33. pp. 909-916.

Thomas, Kim (May 2007). Just Noticeable Difference and Tempo Change. Jour. Of Scientific Psychology. pp. 14-20.

Walker, Bruce N. (Sept. 2000). Magnitude Estimation Of Conceptual Data Dimensions For Use In Sonification - Doctoral Dissertation. Rice University, Houston, Texas.

Walker, Bruce N., Lindsay, Jeffrey, & Godfrey, Justin (2004). The Audio Abacus: Representing Numerical Values With Nonspeech Sound For The Visually Impaired. The Tenth International Conference On Auditory Display.

Walker, Bruce N., & Kramer, Gregory (Oct. 2005). Mappings And Metaphors In Auditory Displays: An Experimental Assessment. ACM Trans. On Applied Percep., Vol 2 No. 4. pp. 407-412.