RECOGNITION OF AUDIFIED DATA IN UNTRAINED LISTENERS

Robert L. Alexander, Sile O'Modhrain, Jason A. Gilbert, Thomas H. Zurbuchen

University of Michigan,
Ann Arbor, MI, 48104
robertalexandermusic@gmail.com

Mary Simoni

Rensselear Polytechnic University
School of Humanities Arts and Social Sciences
msimoni@rpi.edu

ABSTRACT

The effective navigation and analysis of large data sets is a persistent challenge within the scientific community. The objective of this experiment was to determine whether participants who received no training were able to identify audified data sets at a rate above chance in a forced-choice listening task. Nineteen participants with various levels of musical and scientific expertise were asked to place audio examples into one of the five following categories: Digitally Generated Sound - White Noise, Solar Wind Data, Neuron Firing Data from a Human Brain, Seismic Data (Earthquake Activity), and Digitally Generated Sound - Sinusoidal Waveform. At no time were participants made aware of the accuracy of their responses during the experiment. Participants who had never been exposed to audified data sets were able to recognize audification examples at a rate that was 23 percentage points above chance performance; however, the sample size of individuals with no previous exposure to audified data was not large enough to determine statistical significance. When controlling for previous exposure to any of the provided listening examples, all participants performed well above the statistical likelihood of chance responses in the recognition of digitally generated white noise and sinusoidal waveforms. This indicates that participants with no previous exposure to audified data were able to discriminate between audified data and digitally generated sounds.

1. INTRODUCTION

Sonification is the science that concerns the transfer of information through sound. The *Sonification Report* broadly defines this term as "the use of non-speech audio to convey information." [1] Audification is a specific form of auditory data analysis in which data samples are isomorphically mapped to audio samples. This method has proven successful in uncovering new insights that would otherwise be overlooked through traditional analysis methods [2, 3]. However, no methodological framework has been established for how this process may be successfully implemented for exploratory data analysis across a wide range of scientific disciplines. One goal of this experiment is to establish a baseline measurement for human ability to recognize audified data sets.

Formal research in the field of auditory data analysis can be traced back to the year 1946, when a volume was published on the *Principles of Underwater Sound* with the goal of advancing sonar techniques [4]. Three years later *The Mathematical Theory of Communication* laid the foundation for our modern understanding of signal processing techniques [5]. Early auditory display research demonstrating that multi-modal stimulation could greatly increase the rate of information transfer to a human operator [6-8]. This investigation was later extended to human pattern matching abilities, finding that known visual-analysis methods were often inferior to auditory analysis in the representation of multivariate data [9]. Several additional experiments utilizing multivariate data were conducted by Bly, and it was noted that "sound can indeed increase the information about multivariate data when it is presented to a human analyst." [10, 11]

Sonification techniques have been employed in a wide range of scientific studies that build upon these early foundations. In *An Illustrated Analysis of Sonification for Scientific Visualization* it was noted that, "all aspects of sonic display of information need further research." A discrete set of possible areas where sonification research could be beneficial were offered, including: data representation, interaction processes, and validation of graphical processes [12].

Modern auditory data analysis techniques are commonly taught in academic settings, though this instruction is geared towards expertise in music-production. A course at the University of entitled "Timbral Ear Training" teaches students to notice subtle changes in the spectral composition of white and pink noise fields [13]. It is possible that this type of training could also prove effective in enabling researchers to recognize subtle differences between audified data sets. The objective of this experiment is to determine a pre-training baseline rate for successful recognition of audified data sets, with a comparison against chance performance utilized as a metric. Audified data sets will be presented in conjunction with digitally manufactured noise and sinusoidal waveforms, as previous research has suggested that auditory data analysis can be beneficial in the identification of equipment-induced noise [2].

2. METHODS

2.1. Participants

Nineteen participants took part in this experiment (6 female, 13 male; age 21 to 40). A pre-test questionnaire established that four participants had received a high school diploma, ten had received a bachelor's degree, and five participants had received a Masters or PhD. Three participants had no musical training, one participant had a single year, four participants had two to three years, seven participants had four to six years, and four participants reported seven or more years of musical training.

All but three participants self-reported average to above-average hearing. A single-frequency auditory threshold test was administered after the listening portion of the survey. This test provided 300ms bursts of a 440hz sinusoidal waveform spaced evenly at 1 second intervals. The gain of each successive waveform was reduced by 6db. Individuals who self-reported below average hearing showed no statistically significant difference in performance on this task *(P* < 0.22).

A post-test questionnaire revealed that of the nineteen participants, thirteen had previously been exposed to audified data in one form or another. All participants reported average or above average expertise with computers, and ten reported average or above average computer-music expertise (nine reported below average). All but two participants reported experience with data analysis, mathematical modeling, and/or scientific research.

2.2. Procedure

The experiment was administered within a custom software-interface built in the Max/MSP programming environment (see **Figure 1**). After completing a short pre-test questionnaire, participants were asked to listen to a series of audio examples played back over headphones. Participants were verbally informed that they could either push a button with the mouse, or press the space bar to play back audio examples. Before beginning the listening task, participants were provided with a spoken-word listening example, and asked to set their audio-playback to a comfortable level utilizing a volume-slider provided within the software interface. The participants' task was to correctly identify the source of a sound from a list of five choices. This forced-choice task included the following available responses for all listening examples: Digitally Generated Sound - White Noise, Solar Wind Data, Neuron Firing Data from a Human Brain, Seismic Data (Earthquake Activity), and Digitally Generated Sound - Sinusoidal Waveform.

On-screen instructions informed participants that audio files were either generated from scientific data or digitally manufactured. It was also made clear that multiple examples of each type could appear over the course of the experiment. A total of 8 audio files were utilized for the listening task, these included two examples of audified neuron firing data from a human brain, two examples derived from solar-wind data, two examples of audified earthquake data, and one example of both white noise and a sinusoidal waveform. Each audio example was provided 3 times: Once at full speed, once at 75% full speed, and once at half speed. Twenty-four listening examples were provided in total.

Participants were asked to make their best guess as to the source of the audio, and then press a separate button labeled "submit" to enter their selection. At no time were participants made aware of the accuracy of their selection during the experiment (the experimental coordinator was always present within the room, but did not answer any questions pertaining to accuracy of responses). Participants were provided with a number corresponding to the current question (out of 46 total questions), such that they could track their progress towards completion. An on-screen clock began counting upwards at the beginning of the pre-test questionnaire. An on-screen level-meter provided visual feedback as to the volume of the audio file at 50ms intervals.

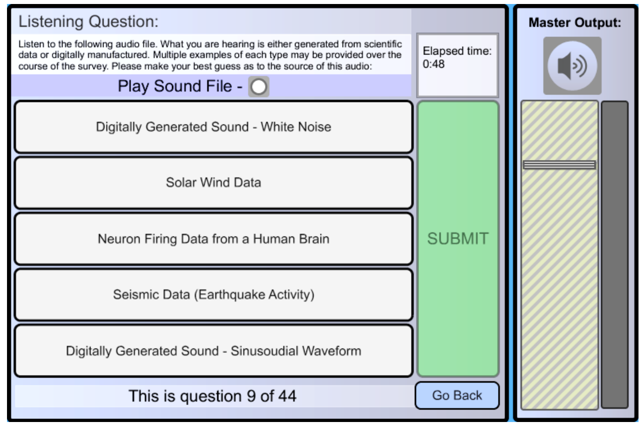


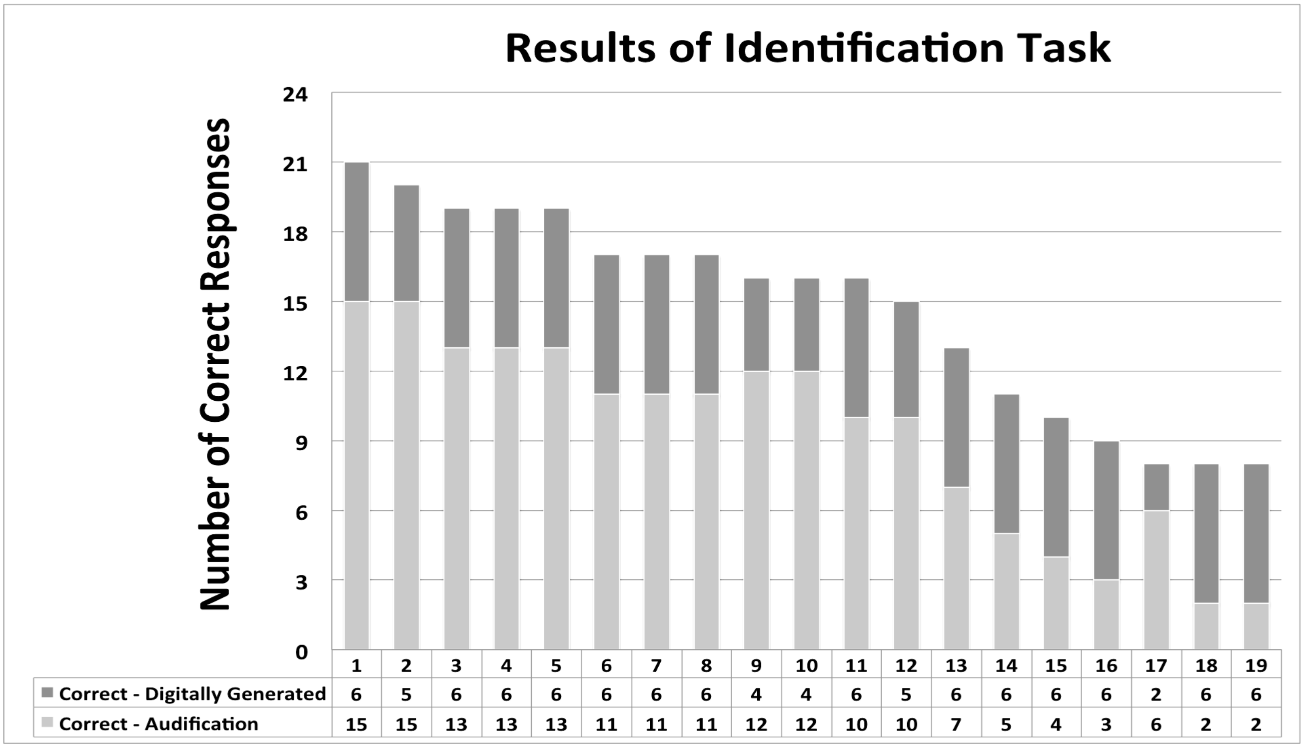
Figure 1. Participants were provided with a list of potential audio sources and asked to guess which example they were currently listening to.

2.3. Stimuli

A total of 8 audio files were utilized for the listening task, all files were 16-bit AIFF format at a sampling rate of 44.1kHz. The seismic data files were downloaded from a server in an audified data format (.wav). The solar wind and neuronal-firing examples were audified with a novel algorithm in the Matlab programming environment. This algorithm transferred the original comma-separated data files into 2-dimensional arrays, and then determined minimum and maximum values in each column of data. These values were utilized to scale the data as floating-point values between -1 and 1. These values were then sequentially mapped to 16-bit audio samples with the "wavwrite" function (all files were ultimately converted to AIFF format for playback in Max/MSP).

All audio files in this experiment were balanced to a similar playback amplitude (RMS). Examples ranged from approximately one to eight seconds in length, with a mean length of 5.3 seconds. All samples (except for the seismic data) were smoothly faded in and out over the course of approximately one to two seconds. A total of eight audio files were utilized for the listening task, each audio example was played back at total of three times: once at full speed, once at seventy-five percent of the full playback speed, and once at half speed. Changes in the rate of sound file playback were calculated in real-time utilizing the "groove~" object in the Max/MSP programming language.

Seismic data was downloaded as audio files from a publicly accessible website maintained by the United States Geological Survey (USGS) science program. The first example contained data from a magnitude 5.1 event that was recorded in Parkfield, California (1994). The second example contained data from a magnitude 6.5 event that was recorded in Petrolia, California (1992). Researchers from U.C. Berkeley recorded both seismograms [14].



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Correct - Digitally Generated	6	5	6	6	6	6	6	6	4	4	6	5	6	6	6	6	2	6	6
Correct - Audification	15	15	13	13	13	11	11	11	12	12	10	10	7	5	4	3	6	2	2

Figure 2. Individual performance on the identification task sorted by number of correct responses (highest to lowest). This stacked bar graph provides the number of correctly identified audification examples (bottom) as well as the number of correctly identified digitally manufactured examples (top).

Two examples of audified solar wind data were created for this experiment. The first example was downloaded from the NASA's Coordinated Data Analysis Web (CDAWeb) as a comma-separated text file. This type of data, which merges records from multiple satellites, is referred to as OMNI data and is available to the general public publically available. This specific file contained solar wind hourly averaged bulk proton flow speed (km/s) measurements spanning the years 1963 to 2010 inclusively, and was 421,057 entries in length.

The second solar wind example was generated with data collected by the Solar Wind Ionic Composition Spectrometer (SWICS) instrument on the Advanced Composition Explorer (ACE) satellite. This data measured the variance of the solar magnetic field at 16-second intervals, and was gathered over the course of the year 1997. The source file was downloaded from a publically accessible data repository [15], this file was 112,104 data samples in length.

The neuronal firing data was collected from a probe during a Deep-Brain Stimulation (DBS) surgical procedure. The probe, measuring approximately 40-microns in circumference circumference, recorded micro-voltage fluctuations at a rate of 30,000 samples per second. This audio was converted to a sampling rate of 44,100 for playback within the experimental interface. The two neuronal firing examples were taken from separate files; one file measured 83.3 megabytes in size, while the other measured 1.69 gigabytes. After audifiaction, a sub-section was chosen from each file that contained prominent firings from a single-neuron (as identified by a researcher experienced in close-listening to audio from DBS procedures).

The white noise and sinusoidal listening examples were generated with the Max/MSP computer-music programming language, utilizing the "noise~" and "cyle~" objects respectively. The frequency of the sine wave example was 440hz.

2.4. Apparatus

The experiment was conducted utilizing a 15-inch MacBook Pro running the Mac OS X operating system (Version 10.6.7). All participants used Sony MDR-7509HD Dynamic stereo headphones for all listening examples. The software interface was designed and constructed within the Max/MSP computer-music programming environment (Version 5.1.8). A standalone application was created, which saved experimental data as files in ".txt" format. Before beginning the experiment, participants were prompted to provide their first name, middle initial, and last name; this data was parsed and the resulting initials were used to create unique file names.

3. RESULTS

3.1. Overview

In this forced-choice listening task with 5 possible responses, a 20% success rate across all 24 listening examples would result in an average of 4.8 correct responses. This success rate would indicate chance-performance. Results from this identification task have been summarized in **figure 2**. The average number of correct responses across all participants (and all listening examples) was 14.68, with a standard deviation of 4.41. This finding is considered to be extremely statistically significant when compared against chance performance ($P < 0.0001$). Measures of statistical significance in all cases were calculated utilizing a t-test, unless otherwise noted the statistical mean was measured against chance performance. The highest number of correct responses was 21 (1 participant) and the lowest number of correct responses 8 (recorded by 3 participants). Information regarding the number of correct responses for each listening example has been provided in **figure 3**.

	Audio Example	Playback Speed	Correct Response
1.	Solar-Wind Data (Proton Flow Speed)	1/2	63.2%
2.	Brainwave Data (Example 1)	Original (1)	31.6%
3.	Seismic Data (Parkfield 1994, Mag 5.1)	1/2	63.2%
4.	Brainwave Data (Example 2)	3/4	47.4%
5.	Solar-Wind Data (Proton Flow Speed)	3/4	57.9%
6.	Sine Tone (440hz)	3/4	100.0%
7.	Seismic Data (Petrolia 1992, Mag 6.5)	Original (1)	26.3%
8.	Solar Wind Data 2 (magnetometer)	3/4	26.3%
9.	White Noise	3/4	84.2%
10.	Brainwave Data (Example 2)	Original (1)	63.2%
11.	Sine Tone (440hz)	1/2	94.7%
12.	Solar-Wind Data 2 (Magnetometer)	Original (1)	57.9%
13.	Seismic Data (Parkfield 1994, Mag 5.1)	3/4	63.2%
14.	Brainwave Data (Example 2)	1/2	47.4%
15.	White Noise	Original (1)	84.2%
16.	Seismic Data (Petrolia 1992, Mag 6.5)	3/4	47.4%
17.	Sine tone (440hz)	Original (1)	94.7%
18.	Brainwave Data (Example 1)	1/2	47.4%
19.	White Noise	1/2	89.5%
20.	Brainwave Data (Example 1)	3/4	42.1%
21.	Seismic Data (Petrolia 1992, Mag 6.5)	1/2	52.6%
22.	Solar Wind Data 2 (magnetometer)	1/2	73.7%
23.	Solar Wind Data (Proton Flow Speed)	Original (1)	57.9%
24.	Seismic Data (Parkfield 1994, Mag 5.1)	Original (1)	52.6%

Figure 3. Complete list of audio examples, relative playback speed, and percentage of correct responses. This is the order in which the listening examples were provided to all participants.

3.2. Correlative Evaluation

Participants completed the pre-test questionnaire, listening task, and post-test questionnaire in an average of eleven minutes. No correlation was found between above-average and below-average completion time and the number of correct responses ($P < 0.7524$). Similarly, no significant correlation was found between gender and recognition ability ($P < 0.73$). There was no significant difference between the performance of participants aged 24 and younger, or 25 and older ($P < 0.8541$). Participants with Masters or PhD degree correctly identified an average of 16 examples (out of a total 24), while participants who completed high school or a received a bachelor's degree correctly identified an average of 14.21 ($P < 0.45$). Participants with 7 or more years of musical training successfully identified an average of 17.75 examples, while participants with zero to six years of musical training identified an average of 13.87 ($P < 0.12$). These results were determined to be statistically insignificant based on the small sample size (see **figure 4**).

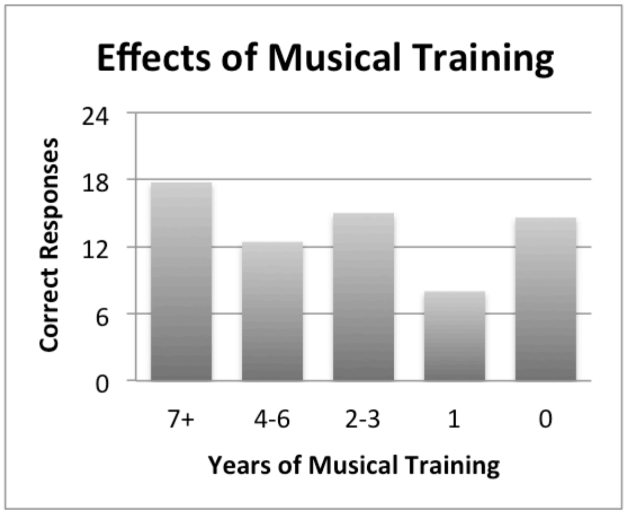


Figure 4. No statistically significant correlation was found between a participant's level of musical training and performance on the identification task.

3.3. Controlling For Pre-Exposure

The following section independently evaluates recognition ability for the six digitally generated sounds (three white noise and three sinusoidal) as opposed to the eighteen audified examples (six solar wind, six neuronal and six seismic). All participants correctly identified an average 9.21 of the 18 audified examples, which indicates a performance significantly better than chance ($P < .0001$). The 13 participants who had previously been exposed to at least one of the listening examples were able to correctly identify an average of 9.85 of the 18 audified data examples, this is considered to be extremely statistically significant when compared to chance ($P < 0.0001$). The 6 participants who had not previously been exposed to any of the listening examples were able to correctly identify an average of 7.83 of the 18 audified data sets. This performance is 23 percentage points higher than chance, however, the sample size is not large enough to determine statistical significance ($P < 0.09$).

All participants correctly identified an average of 5.47 of the 6 digitally manufactured sounds, which indicates a performance significantly better than chance ($P < .0001$). Participants who had never previously been exposed to any of the audio examples correctly identified an average of 5.67 of the 6 digitally manufactured sounds ($P < .0001$), while participants who had been previously exposed to some of the audio examples correctly identified an average of 5.38 out of 6 of digitally generated sounds ($P < .0001$). The performance difference between the two groups in the identification of the digitally generated sounds was statistically negligible ($P < 0.6$).

4. DISCUSSION

The objective of this experiment was to determine whether participants who received no training were able to identify audified data sets at a rate above chance. One notable outcomes of this experiment was that participants who had never been exposed to audified data sets (6 of the total 19) were able to recognize the audification examples at a rate of 43.5%, which was 23 percentage points above chance. However, the sample size of individuals with no exposure was not large enough to determine statistical significance ($P < 0.0907$). A future experiment should pre-select individuals with no exposure to audified data of any kind in order to determine recognition ability for individuals with no previous exposure.

The success rate for identifying audified data sets was found to be 11% higher for participants with pre-exposure to audified data sets (54.7%) than individuals without pre-exposure (43.7%,), and this success rate was found to be statistically well above chance. This finding indicates that exposure to audified data could greatly assist in the future recognition of audified data sets, which supports the previous finding that individuals can improve recognition of non-musical auditory stimuli with training [13].

When controlling for previous exposure to any of the provided listening examples, all participants statistically performed well above chance in the recognition of white noise and sinusoidal waveforms. This indicates that participants with no previous exposure to audified data were able to discriminate between audified data and these digitally manufactured sounds without training. This provides strong support for the previous assertion that auditory data analysis can be beneficial in the identification of equipment-induced noise, particularly in the training of non-experts [2].

Many steps could have been taken to improve upon the design of this experiment. Several participants, when prompted for additional feedback in the post-test questionnaire, mentioned that they recognized repeated audio examples, despite the fact that recurring examples were always played back at different speeds. It was noted that this could be a "confounding element" as participants may try to "match… answers to the pervious ones to be as consistent as possible." As suggested by Levitin, the order of examples could have been randomized in order to minimize any bias imposed by potential "order effects" [16]. All participants correctly identified the sinusoidal waveform upon first listening, while the identification rate dropped slightly the second and third time it was presented. A randomization of presentation order across participants would be necessary in order to determine whether the playback rate of this specific sample had any impact on the number of correct responses.

One participant provided the following additional feedback: "Sometimes I wanted to put none of these I felt like the noise presented didn't sound like any of the 5 categories." This points to potential priming effects induced by the limited forced-choice selection. Participants may have responded significantly differently had they been provided with an option for "Other – This sounds like a type of audified data which is not included in this list." If the purpose of a future study were to examine the benefits of audification in exploratory data analysis, a forced choice paradigm might include an "other" option with space provided for free response. In this way the experiment could extract some ideas as to what untrained listeners believing they are hearing when they are free to craft novel responses in their own words.

In addition to these improvements, a multi-frequency auditory threshold test could have been administered to establish the presence of a healthy audiometric threshold in all participants. A single-band threshold test was not found to be sufficient in this task

5. CONCLUSION

Audification has proven successful in uncovering new insights that would otherwise be overlooked through traditional analysis methods [2, 3]. However, no methodological framework has been established for how this process may be successfully implemented for exploratory data analysis across a wide range of scientific disciplines. The objective of this experiment was to determine whether participants who received no training were able to identify audified data sets at a rate above chance in a forced-choice listening task. Participants who had never been exposed to audified data sets were able to recognize the audified examples at a rate that was 23 percentage points above chance performance; however, the sample size of individuals with no exposure was not large enough to determine statistical significance. When controlling for previous exposure to any of the provided listening examples, all participants statistically performed well above chance in the recognition of digitally generated sounds (White Noise and Sinusoidal waveforms). This indicates that participants with no previous exposure to audified data were able to discriminate between audified data and digitally generated sounds.

Upon repeated listening, pattern-recognition processes within the brain rapidly begin to enhance deeply embedded structural details of extremely noisy signals [17]. Exposure to audified data could greatly assist in the future recognition of audified data sets, which supports the previous finding that individuals can improve recognition of non-musical auditory stimuli with training [13]. A future experiment should pre-select individuals with no exposure to audified data of any kind in order to determine recognition ability for individuals with no previous exposure. Another future study should examine the benefits of audification in exploratory data analysis through a forced choice paradigm with an "other" option. This free-response space would allow participants to craft novel responses in their own words, which could provide valuable insight.

References

[1] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N.; Neuhoff, J., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M., Tipei, S. *Sonification report: status of the field and research agenda*. International Community for Auditory Display, 1999.

[2] Hayward, C. Listening to the Earth Sing. in Kramer, G. ed. *Auditory display: Sonification, Audification, and Auditory Interfaces*, Addison-Wesley, Reading, Mass., 1994, 369-404.

[3] AUTHOR.

[4] Eckart, C. *Principles of Underwater Sound*. Division of War Research. Sonar Data Division, United States. Office of Scientific Research and Development. National Defense Research Committee., California. University., 1946.

[5] Shannon, C.E. and Weaver, W. *The Mathematical Theory of Communication* University of Illinois Press, Urbana, 1949.

[6] Pollack, I. and Ficks, L. *Information of Elementary Multidimensional Auditory Displays*. ASA, 1954.

[7] Fidell, S. *Sensory Function in Multimodal Signal Detection*. ASA, 1970.

[8] Loveless, N.E., Brebner, J. and Hamilton, P. Bisensory presentation of information. *Psychological Bulletin*, *73* (3). 161-199.

[9] Yeung, E.S. Pattern recognition by audio representation of multivariate analytical data. *Analytical Chemistry*, *52* (7). 1120-1123.

[10] Bly, S. Presenting information in sound *Proceedings of the 1982 conference on Human factors in computing systems*, ACM, Gaithersburg, Maryland, United States, 1982, 371-375.

[11] Bly, S. Sound and computer information presentation *Other Information: Thesis. Portions of document are illegible*, 1982, 124.

[12] Minghim, R. and Forrest, A.R. An Illustrated Analysis of Sonification for Scientific Visualisation *Proceedings of the 6th conference on Visualization '95*, IEEE Computer Society, 1995, 110.

[13] Corey, J. *Audio Production and Critical Listening: Technical Ear Training*. Elsevier/Focal Press, 2010.

[14] United States Geological Survey: Listen for Fun., http://earthquake.usgs.gov/learn/listen/allsounds.php, 2011.

[15] (ASC), A.S.C. ACE SWICS-SWIMS-V3 Level 2 Data, http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_SWICS-SWIMS.html, 2011.

[16] Levitin, D.J. Experimental design in psychological research. in Levitin, D.J. ed. *Foundations of cognitive psychology: Core readings*, MIT Press, Cambridge, MA, 2002, 115-130.

[17] Kaernbach, C. Temporal and spectral basis of the features perceived in repeated noise. *J. Acoust. Soc. Am.* 91-97