An RSS-feed Auditory Aggregator Using Earcons

Athina Bikaki
School of Science & Technology
Hellenic Open University
Parodos Aristotelous 18, 262 22 Patras, Greece
+30 697 6119757

athina.bikaki@gmail.com

Andreas Floros
Dept. of Audiovisual Arts
Ionian University
Plateia Tsirigoti 7, 49100 Corfu, Greece
+30 26610 87725

floros@ionio.gr

ABSTRACT

In this work we present a data sonification framework based on parallel / concurrent sonic earcons' representations for monitoring in real-time information related to stock market. The information under consideration is conveyed through the well-known Really Simple Syndication (RSS) feed Internet mechanism and includes both text and numeric values, converted to speech and earcons using existing speech synthesis techniques and sonic design guidelines. Due to the considered application characteristics, particular emphasis is provided on information representation concurrency, mainly achieved using sound source spatialization techniques and different timbre characteristics. Spatial positioning of sound sources is performed through typical binaural processing and reproduction. A number of systematic, subjective assessments performed have shown that the overall perceptual efficiency and sonic representation accuracy fulfills the overall application requirements, provided that the users are appropriately trained prior to using the proposed RSS-feed auditory aggregator.

Categories and Subject Descriptors

J.5 [**Arts and Humanities**]: Music – *data sonification, earcons, sonic design*

General Terms

Algorithms, Design.

Keywords

RSS-feed sonification, earcons, auditory displays.

1. INTRODUCTION

Data sonification represents a widely-employed approach for representing information originating from every-day life activities through appropriately designed auditory displays. Typical application fields include visual representation systems enhancement [1], as well as exclusively non-visual systems for data representation [2]-[3]. Sonification takes advantage of fundamental human auditory features, including (for example) the high dynamic range of human hearing and the ability to accurately localize sound source positions in the three dimensional (3D) space, allowing a significant reduction of the visual input impact to the overall human perception.

On the other hand, the rapid growth of the available information over the Internet has lead to the development of multiple means and protocols for (real or non-real time) data delivery over packet-based (IP) networks. Moreover, the enrollment of multimedia networking technologies has increased the number of potential formats for information delivery and representation. A typical, widely-employed protocol for information delivery over Internet is the so called Really Simple Syndication (RSS), a relatively simple and lightweight XML format designed for sharing Web content among targeted subscribers [4]. RSS-enabled web sites encapsulate transmitted data in information channels (termed as RSS-feeds), which are accessed by remote clients/aggregators. These clients are very frequently responsible for information representation, using mainly visual output means. Typical content transmitted through RSS include news and announcements, calendars, search results, e.tc. [5] and is formed and finally represented as text.

In this work, we aim to extend the concept of RSS-based information delivery and aggregation using sonification. More specifically, we propose an application framework, which transparently parses a number of RSS-feeds and represents them through an appropriately shaped spatial auditory display. The apparent advantage of this approach is that the perception of the RSS-feed content becomes an "eye-free", non-visual process, while, in addition, parallel and comparative web content representation is feasible. We subjectively assess the performance of the overall sonification-enabled RSS system for a typical and very common application scenario: concurrent news and stock market monitoring. The choice of this scenario is based on the fact that it incorporates multiple parallel and real-time information transmissions, allowing the evaluation of the overall functionality of the proposed system.

The rest of the paper is organized as following: the next Section provides a detailed overview of different sonification techniques already proposed in the literature, with emphasis on earcons, a sonification technique that is in particular considered in this work. Additional details regarding the implementation of the proposed sonification-enabled RSS-feed are also analyzed in Section 3, followed by the subjective assessment of the performance of the overall system presented in Section 4. The last Section concludes the work and indicates a number of issues that should be further investigated in the future.

2. SONIFICATION THROUGH EARCONS: AN OVERVIEW

In general, the term sonification defines the process of using non-speech audio to convey information [6]. It is different in concept compared to audification, which particularly refers to the process of considering the data values under representation as the amplitude values of the sonic waveform. The fundamental implementation aspect of sonification is the auditory display [7], which incorporates all the necessary sound signal components for achieving the desired acoustic perception. With the continuous evolution of three dimensional (3D) sound coding and reproduction formats, the information volume that can be conveyed through an auditory display is significantly increased, since it can be reproduced and perceived in parallel / concurrently [8].

A number of different sonic design approaches for data sonification has been already appeared in the literature and is widely employed in many application fields, including non-visual human-machine interfaces for blind and visually impaired users. Starting from simple alarm signals, one can additionally identify parameter mapping techniques [9], where the monitored data values are directly mapped to specific sonic attributes (such as pitch, vibrato strength or speed, brightness, e.tc), producing a kind of "sonic" scatter plot. Musical sonification extends the above concept by providing additional mapping parameters of organized music structure [10], such as tempo, rhythm or even orchestration, hence adding aesthetic characteristics to the final sonic representation. Moreover, significant enhancements can be achieved when using Model-Based Sonification (MBS) [11], which aims to introduce dynamic characteristics for the connection between the monitored data and the acoustic representation, allowing the production of sounds driven and controlled specifically by the user actions.

Auditory icons [12] also represent a common sonification technique, widely employed in every-day applications. An auditory icon is the direct equivalent of a visual icon. It mainly relies on everyday sounds and is targeted to provide a recognizable sonification mapping based on the specific type of data rather than mapping to sound the data values themselves. Earcons on the other hand are defined as "nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation or interaction" [13]. Under this perspective, earcons are constructed using specific sonic motives employed as fundamental building sound structures combined to form the complete sonic event. Sonic motives variable characteristics may include rhythm, pitch, intensity, timbre and register [14]. An exception to this earcons structure principle are spearcons, which are derived from speech signals sufficiently depressed in time, in order to sound not as speech, but as an earcon [15].

Hence, due to their variable and controlled structure, earcons are generally flexible and can be additionally organized in hierarchical families, a significant feature for representing information organized in a tree-format. Moreover, concurrent (or parallel) earcons may extend the presentation bandwidth of the auditory display using sonic reproduction in a temporally overlapping fashion, provided that specific design guidelines will be followed [16]. However, despite their flexibility, earcons suffer from the lack of a consistent and meaningful relation with their referent context. In order to overcome this, targeted users have to learn this relation in advance and train themselves.

3. IMPLEMENTATION ASPECTS
3.1 Design requirements

As mentioned in the introductory Section, aim of this work is to define, develop and assess a framework for RSS-feed sonification. Provided that the web content distributed though RSS refers to a variety of information types, the following initial requirements were considered during the RSS-feed auditory aggregator design phase:

- The complete system should be able to support multiple types of information conveyed through corresponding RSS-feeds.

- Data transmissions through an RSS-feed are asynchronous. Hence, many multiple data-streams can arrive simultaneously to the RSS aggregator. The probability of concurrent reception increases with the generation rate of the transmitted data.

- In order to avoid any information losses imposed by auditory masking phenomena during concurrent RSS-feeds transmissions, a sonification strategy that supports parallel sonic events should be selected and designed.

- Periodic or very frequent data transmissions impose short durations of sonic events used to create the complete auditory display.

- Sonification concurrency also resolves the very-first requirement of supporting multiple types and formats of RSS transmitted information.

Given the above initial assumptions, the earcons sonification strategy was selected for the purposes of this work, since it fulfills the above list of initial requirements. Moreover, in order to develop a realistic test application scenario, we have considered the following use case: The subscribed RSS-feeds included two types of information: a) stock market data values, delivered as text but translated to numeric values by the RSS-aggregator system and b) announcements and news related to the stock market transactions, received as plain text.

3.2 System Architecture

The RSS-feed auditory aggregator general architecture is illustrated in Figure 1. We hereby assume that the proposed sonification-enabled RSS-feed aggregator is subscribed to N different RSS-feeds. Depending on their type, the parsed information is organized into M information streams (where M≤N, a value that exclusively depends on the type of the received data). Currently, this information categorization is performed manually by the user during the initial subscription to a specific feed. However, more advanced automated methods could be employed in the future, taking into account the RSS-feed identity, as well as the transmitted content itself.

Data-stream clustering information is transmitted to the Earcon Design Engine, which is responsible for producing the appropriate earcons in real-time, taking into account the type of each information category (a process that is illustrated in Figure 1 as a dotted arrow). Real-time earcons construction is preferable compared to alternative approaches (i.e. pre-recorded set of earcons), due to the high amount of storage capacity required to support a large variety of transmitted RSS information.
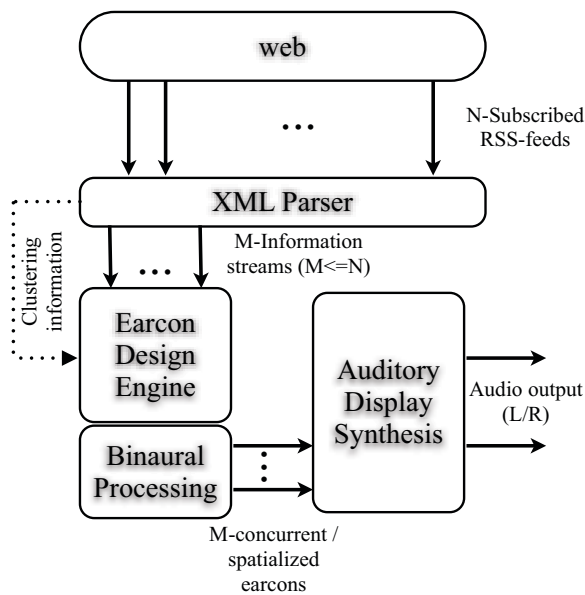
Figure 1. Diagram of the RSS-feed auditory aggregator general architecture

The binaural processing module is used to introduce spatialization in the final auditory display and provide a robust and efficient (in terms of acoustic perception) mean of concurrency during sonification. The choice of binaural technology was based a) on the localization accuracy it achieves (especially when using equalized headphones; however, audio reproduction through a loudspeaker pair is feasible if cross-talk cancellation pre-processing techniques are applied) b) on the simplicity of the reproduction system setup, and c) on the portability that can be supported, a factor that can address mobility issues frequently required by the end user. Finally, the derived spatialized earcon signals are forwarded to the Auditory Display Synthesis module, which is responsible for mixing the corresponding binaural signals and reproducing the complete auditory display.

Figure 2 illustrates the particular implementation details and data flow followed within the scope of the application scenario described previously. Both the announcements and the stock values data were received from RSS-feed subscriptions of the www.nasdaq.com web site. The stock market news feeds were converted to speech, using text-to-speech (TTS) techniques. The synthesized speech signal was processed by the binaural processing sub-system to allow the creation of a virtual speech source in the 3D space of the final auditory display. The control parameters of the speech synthesis and the voice virtual source binaural processing are appeared in Figure 3. On the other hand, the stock values data were converted to earcons, following the procedure that is described in detail in the next Section.

Binaural signal processing was performed using the CSound audio synthesis programming framework [17]. The selection of CSound was based on the spatialization simplicity required by the developed application: for the purposes of this work, virtual sound sources placement was allowed in discrete, pre-defined positions at different elevation and azimuth values. Moreover, the relatively short duration of the earcons and speech signals relaxed the requirements for real time processing and binaural rendering, especially for sampling rates equal to 44.1kHz. Higher sampling rate values were found to induce unacceptable reproduction delays

caused by the more intensive calculations; however, this does not impose any quality issues to the overall implementation.
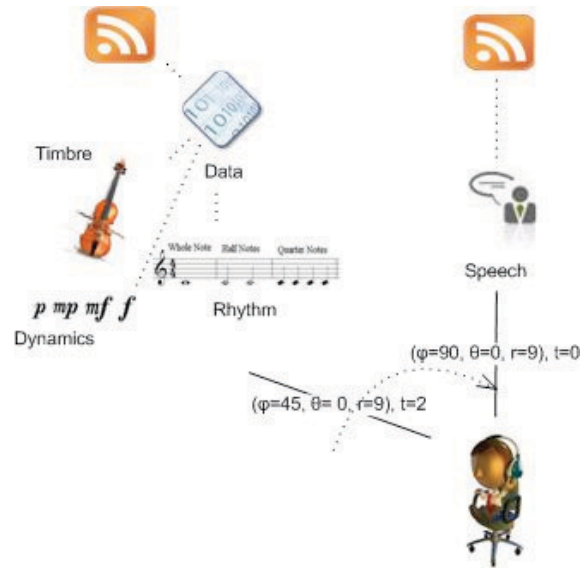

Figure 2. Schematic description of the proposed implementation

Regarding the RSS-feed subscription handling, a categorized list of different RSS feeds was offered by the system. By selecting a category, the list was filtered accordingly and the corresponding RSS feeds were displayed to the user. The user was also able to define RSS-feed clusters (i.e. stock quotes), a process that corresponds to the information categorization procedure described previously.
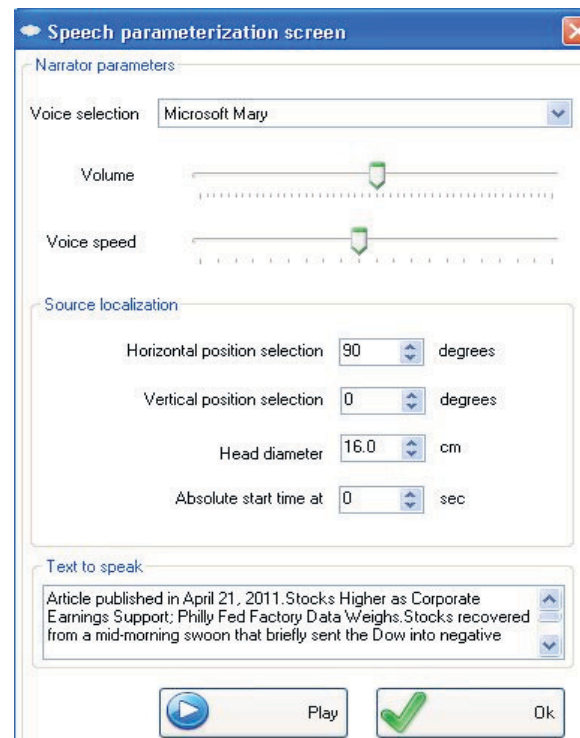

Figure 3. Speech synthesis parameterization options

3.3 Earcons Design

As mentioned in the previous Sections, stock market values were sonified using earcons. Towards a robust and efficient earcons sonic design, we followed basic design principles introduced in [14] and [16]. Hence, we mapped different, numeric RSS data feeds to different musical instruments (timbres). Additionally, numeric and text data representation concurrency was implemented using sound source spatialization (realized again through earcons binaural processing). In [16], this approach was found to achieve better concurrent / parallel earcons identification scores, compared to earcons derived using non-spatial information only.

The musical instruments employed were selected to be members of a symphonic orchestra, providing an adequate variety of available timbres. Moreover, this approach was selected for defining their spatial positions, which followed a typical symphonic orchestra layout. More specifically, we considered instruments from different categories i.e. strings, woodwind, brass etc, in order to allow a maximum amount of perceptual timbre discrimination. The different combinations of instruments in the orchestra allowed the investigation of how to produce a well-sounded chord of certain tone-quality, acquire a uniformity of structure and requisite power. Using this information, we tried to balance and correct the "colour" of the final sound. Additionally, during sound source spatialization, we placed stock information categorized in the same stock cluster (defined by the user, see previous Section) in the same orchestra spatial section i.e. left, right, front-left, front-right etc, so that grouping of stock market information was supported.

For earcons construction, we kept the rhythm and tempo fixed, but we varied the notes' durations within one measure and consequently the number of notes in the same time interval. Small note lengths are better to be avoided as suggested in [14], since they might not be noticed. Thus we considered note lengths always greater than 0.4 seconds. We defined a numerical scale of stock value changes (SVCs) and assigned different ranges to different note values in the measure as well as to the number of notes, as displayed in Table 1. The numerical scale was built using historical data and based on the price changes of a specific stock category i.e. oil companies' stock data. Moreover, we applied sound amplitude variations to control the volatility of each stock quote. For example, if a stock value was increased, the corresponding SVC was positive and it was mapped to a gradually increasing sound gain and vice versa.

Table 1. Stock value changes numerical scale and earcon structure

Note Values in the Measure	Stock Value Change (%)
♩ + ♩ + ♩ + ♩	$|SVC| \geq 1.5$
♩ + ♩ + ♩	$0.8 < |SVC| < 1.5$
♩ + ♩	$0.2 < |SVC| \leq 0.8$
o	$|SVC| \leq 0.2$

We generally tried to keep the earcons' structure as simple as possible, so that it would be easier for a human user to understand, learn and memorize it. To improve the identification of concurrently presented earcons, we also followed Brewster's guidelines [14] to add an extra silent time interval between them. Additionally, the user was allowed to re-define a number of the default earcons design parameters (see Figure 4), in order to create some optimized personal earcons profiles. However, only specific parameters could be assessed by the user, in order to ensure that the basic design principles were not ignored.
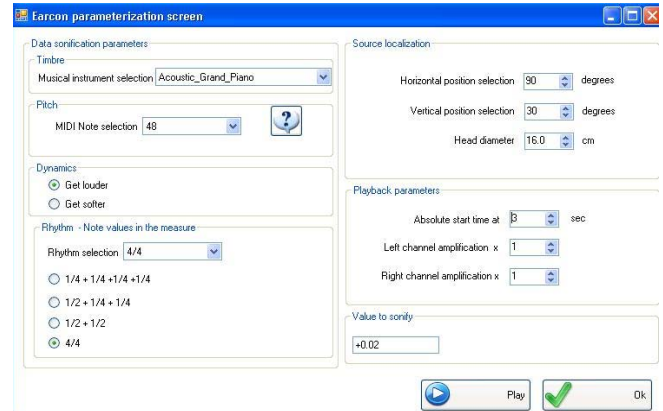

Figure 4. Earcons design parameterization options

4. ASSESMENT AND RESULTS

The assessment of the performance of the RSS-feed auditory aggregator was performed through a sequence of subjective tests. During these tests, ten participants (six men and four women aged between 32 and 75) used the system under six different initial test setups. These test sequences included different numbers and types of active/received RSS-feeds, as well as earcons design and voice parameters. The third column of Table 2 summarizes these applicable test variable conditions.

In the case of speech reproduction we have used two different voices (one male and one female). The participants were requested to fill a questionnaire suitable for measuring the average perceptual score during each test case. Prior to that, detailed guidelines explaining the earcons design rules were provided to all human subjects, which were allowed to train themselves in practice by listening to pre-known stock values sonification cases. It should be also noted that the majority of the participants responded that they didn't have any kind of systematic music education and that they were not music experts.

The score values range was defined within the interval 0 (low perception accuracy) and 5 (high perceptual accuracy). With the term perceptual accuracy, we hereby define the difference measure between the real information parsed by the RSS aggregator and the actual one perceived by the user. The smaller this difference is (measured in percentage units), the higher the obtained perceptual accuracy.

A summary of the results/average scores obtained is displayed in Table 2. In general, the following conclusions were drawn:

- Components that have similar attributes are more likely to be grouped together. Hence, similar news and data are better to be positioned in the same direction (proximity), so that objects close to each other are grouped together. This trend was successfully identified in the sequence of the experiments performed.

- The subject's ability to simultaneously comprehend a speech stream and understand other related sound data decreases as the number of transmitted RSS-channels increase.

- The use of male–female voices alternatively increases the ability of the user to comprehend and easily pay attention to different text RSS-streams/announcements.

- The use of different musical instruments for earcon sonic construction resulted into better perceptual segregation compared to the simplistic use of different pitch.

- Variations in tempo and envelope/amplitude dynamics additionally and significantly contribute in optimizing the overall perception measured.

Table 2. Subjective tests results summary

Test ID	RSS Feeds	Setup Properties	Average perceptual score
1	3	Speech-speech- speech same gain and rate, different voices	2.5
2	3	Earcon-earcon-earcon same pitch, different timbre, gain, rhythm	4.0
3	3	Earcon-earcon-earcon same timbre, different pitch, gain, rhythm	3.9
4	4	Earcon-speech-earcon-speech same pitch, volume and rate, different timbre, gain, rhythm and voices	5.0
5	6	Earcon-earcon-speech-earcon-earcon-speech partially same pitch, same gain and rate, partially different timbre, different gain, rhythm and voices	3.9
6	6	Earcon-speech-earcon-speech-earcon-speech same pitch, volume and rate, different timbre, gain, rhythm and voices	4.1

We additionally performed two more accurate subjective tests to evaluate the system efficiency. All the tests represented daily stock value changes as well as stock news information. In each earcon we encoded the stock quote name, the stock price trend and a range of its value change, while in each speech we encoded the stock news text. We carried out a training session followed by an evaluation one for each test, for seven different participants. In the short training session we ensured that the participants had learned the sounds and we proceed to the evaluation of the system. During this evaluation, the participants had read the instruction sheet, which described the system setup, and were asked to indicate the trend and the daily value change range of each stock quote, considering the ranges appeared in Table 1. The score values range was also defined within the interval 0 (low evaluation accuracy) and 5 (high evaluation accuracy). A summary of the evaluation tests results is presented in Table 3.

Table 3. Evaluation tests results summary

Test #	Feeds	Setup Properties	Average score
1	4	Earcon-speech-earcon-speech	5
2	6	Earcon-speech-earcon-speech-earcon-speech	3.8

The results have confirmed that the grouping of information in the layout described in the previous section helped the participants to correctly identify stock price information. From the evaluation experiment we verified, that the layout of the four different RSS feeds grouped as a pair of earcon-speech for each stock, was the most robust. An increase in the number of stocks, showed a slight decrease in the identification performance.

5. CONCLUSIONS

Data sonification represents an alternative mean for eye-free information representation in a wide range of every-day life applications. Starting from auditory displays targeted to blind or visual-impaired users and extending its' potential employment in cases where visual focus cannot be achieved, there is a large number of alternative sonification techniques with different characteristics and effective perceptual performance.

In this work we aim to introduce the concept for a novel RSS-feed auditory aggregator, based on one of the most interesting realizations of sonification: earcons. RSS protocol represents a widely-employed framework for information delivery over Internet. Multiple information formats can be supported, limited only by the state-of-the-art in multimedia technologies applied in packet-based networks. The proposed RSS-feed auditory aggregator takes advantage of the fundamental properties of earcons (such as the parallel / concurrent reproduction support using 3D audio spatialization and timbre variation) for achieving an efficient representation of multiple, concurrent RSS streams with different content (i.e. organized as simple text or as discrete data values). The blending of these techniques enables a wider range of innovative applications for both mobile and stationary home use. A few examples of interesting applications that can be envisioned for the RSS-feed auditory aggregator are: real-time monitoring of different stock's information data (open/close, high/low) and trade volume, real-time monitoring of different financial instruments (currencies, bonds, mutual funds, etc.), real-time monitoring of a selected stock portfolio, sonifying and monitoring a stock market index composition, etc.

The realization of a typical application scenario and a sequence of subjective tests performed have shown that the proposed RSS-feed sonification approach achieves adequate performance in terms of perceptual accuracy of the transmitted content. Within the context of this work, the overall system lacks support of an automated procedure for an arbitrary number of RSS-feed subscriptions and their direct mapping to earcons design cues. However, it is the authors' future intention to investigate this topic, towards a robust and optimized system for RSS-feeds and real-time information sonification.

6. REFERENCES

[1] Brewster, S.A. 2002. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6, 3, 88–205.

[2] Edwards, K., Mynatt, E.D. and Stockton, E.D. 1995. Access to graphical interfaces for blind users. *Interactions* 2, 1, 56–67.

[3] Brown, L. and Brewster, S.A. 2003. Drawing by ear: Interpreting sonified line graphs. In *Proceedings of the International Conference on Auditory Display*, (Boston, Massachusetts, July 6-9, 2003). ICAD'03, 152–156.

[4] The RSS 2.0 specification, RSS advisory board, 2009. http://www.rssboard.org/rss-specification

[5] RSS Tutorial for Content Publishers and Webmasters. http://www.mnot.net/rss/tutorial/

[6] Kramer, G. 1998. Sonification Report: Status of the Field and Research Agenda. NSF Sonification White Paper.

[7] Kramer, G. 1994. An introduction to auditory display. In *KRAMER, G. (Ed.) Auditory Display : sonification, audification, and auditory interfaces.* Reading, MA, Addison-Wesley.

[8] Frauenberger, C., Putz, V. and Holdrich, R. 2004. Spatial Auditory Displays A Study on the Use of Virtual Audio Environments as Interfaces for Users with Visual Disabilities. In *Proceedings of the 7th International Conference on Digital Audio Effects*, (Naples, Italy, October 5-8, 2004). DAFx'04.

[9] Ben-Tal, O., Berger, J., Cook, B., Daniels, M., Scavone, Gary.2002. SonART : The Sonification Application Research Toolbox. In *Proceedings of the 8th International Conference on Auditory Display*, (Kyoto, Japan, 2002). ICAD'02.

[10] Lodha, S.K., Beahan, J., Heppe, T., Joseph, A. and Zane-Ulman, B. 1997. MUSE: A Musical Data Sonification Toolkit. In *Proceedings of the 4th International Conference on Auditory Display,* (Palo Alto, California, November 2 – 5, 1997)*. ICAD'97.

[11] Hermannand, T. and Ritter, H. 1999. Listen to your data: Model-based sonification for data analysis. *Advances in intelligent computing and multimedia systems*, Baden-Baden, Germany, G. E. Lasker, Ed., 189–194.

[12] Gaver, W. 1986. Auditory Icons: Using sound in computer interfaces. *Human Computer Interaction* 2, 2, 167-177.

[13] Blattner, M.M., Sumikawa, D.A. and Greenberg, R.M. 1989. Earcons and Icons: Their Structure and Common Design Principles. *ACM SIGCHI Bulletin* 21, 1 (July 1989), 123-124. DOI = http://dx.doi.org/10.1145/67880.1046599.

[14] Brewster, A.S., Wright, P.C. and Edwards, A.D.N. 1995. Experimentally Derived Guidelines for the Creation of Earcons. In *Proceedings of the HCI'95 Conference*, (29 August - 1 September 1995).

[15] Dingler, T., Lindsay, J. and Walker, B.N. 2008. Learnability of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons and Speech. In *Proceedings of the 14th International Conference on Auditory Display*, (Paris, France, June 24 – 28, 2008). ICAD'08.

[16] McGookin, D.K. and Brewster, S.A. 2004. Empirically Derived Guidelines for the Presentation of Concurrent Earcons. In Proceedings of the HCI2004, (Leeds, UK, September 6-10, 2004).

[17] The CSound acoustic compiler official web site: http://www.csounds.com/