Scalable Auditory Data Signatures for Discovery Oriented Browsing in an Expressive Context

Joachim Goßmann, Ben Hackbarth and Ruth West
with Todd Margolis, J.P.Lewis and Iman Mostafavi

Center for Research in Computing and the Arts,
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0037
jgossmann@ucsd.edu

ABSTRACT

To be useful for browsing in vast multidimensional databases, auditory representations need to accommodate different levels of depth and detail: The ideas behind Scalable Auditory Data Signatures as a sonification strategy that accommodates these demands as well as the underlying concepts of quantitative and qualitative listening are presented. The implementation of these concepts within the interdisciplinary project "ATLAS in silico" is described.

1. INTRODUCTION

Scalable Auditory Data Signatures are entities of auditory representation that encapsulate different depths of data display, achieved through a process of temporal and spatial scaling. This is achieved by generating different versions of the entity that occupy different amounts of time and space. This is achieved through increasing and decreasing the number of sequential and simultaneous sounding elements as well as their spatial distribution and changing the polyphony of simultaneous sound streams, while conserving the perceptual context of a single entity or object.

In contrast to sonification strategies that focus on data transparency and the auditory exposition of known or expected features within a given data set, Scalable Auditory Data Signatures aim to enable an open-ended auditory data-mining process in hopes of revealing unknown or unpredicted features through the formation of emergent patterns in vast and abstract multivariate data sets, accommodating the paradigmatic shift from hypothesis-based toward discovery-based scientific research [1]. SADS are generated by a process of mapping data values onto manifold perceptual domains (see section 2.3) in a manner that allows for the emergence of distinct patterns and features to reveal themselves and dynamically transform as one traverses different levels of scale and detail.

Here we describe the development of Scalable Auditory Data Signatures and their application within immersive virtual reality environments within the scope of ATLAS in silico.

We also discuss their relation to issues of data transparency and discovery of emergent data features within complex, vast, and abstract data sets, including spatio-temporal scaling and complexity. Finally, we present our experience with theses auditory signatures in enabling expressivity in a compositional and artistic context.

1.1. Context

High-density samplings of both nature and social structures, such as genomics, high-resolution global satellite imaging and social network or surveillance data, are increasingly relied upon for knowledge of our world. These practices often generate very large and abstract data sets that present the now well-known challenges of information overwhelm, also termed "information overload." [3] In essence, we have more data about our world than ever in human history – more than what we know what to do with.

Scalable Auditory Data Signatures are a data-sonification strategy that emerged in the context of a large-scale interdisciplinary art-science collaboration entitled "ATLAS in silico" which explores the themes of digitization and data abstraction in regards to nature and culture. This work is a physically interactive virtual environment/installation and art-science collaboration spanning the disciplines of auditory display and sonification, electronic music and composition, new media arts, metagenomics, computer graphics, virtual reality and high-performance computing. It provides a unique aesthetic encounter with metagenomics data (and contextual metadata) from the world's largest known protein sequence dataset, the Global Ocean Survey (GOS) [15] - a groundbreaking snapshot of biodiversity in the world's oceans. The installation was exhibited at SIGGRAPH 2007 and is resident at the California Institute for Telecommunications and Information Technologies on the campus of University of California, San Diego, where it continues in active development. The installation utilizes a combination of infrared motion tracking, custom computer vision, multi-channel (10.1) spatialised interactive audio, 3D graphics, data sonification, audio design, networking, and the Varrier 60 tile, 100-million pixel barrier strip auto-stereoscopic display. (Additional information and video is available at http://www.atlasinsilico.net).

2. DATA REPRESENTATION

2.1. Quantitative and qualitative representations

In the process of ideation in science, people draw conclusions from the data representations they create that lead them to further inquiry. Reading specific data dimensions quantitatively from a sounding non-speech presentation however requires either special analytic listening skills or the employment of cartoonifying strategies such as auditory icons and graphs [4][5]. Quantitative displays that aim at making the data dimensionality transparent often set clear limits on the dimensionality and the variability of the data.

Within certain domains however, the auditory sense is very susceptible to subtle multidimensional variations that seem to defy easy analytic quantification or verbal description and often feature an unclear number of dimensions, opening vast and complex spaces of subjectively perceived differences in perceptual quality [9].

In other words, you can listen for *quantities* of specific values in sound, or for abstract and subjective *qualities* of the sound. While the first use of sonification attempts an auditory communication of specific data values, the latter is aimed at a comparative exploration - a method to highlight differences between datasets. This dichotomy is the backdrop for the concept of SADS:

The ability to perceive subtle qualitative differences in multi-parametric sounds enables scalable auditory data signatures (SADS) to generate emergent auditory structures from the interactive process of browsing vast databases: The goal is to harness the ability of the human sensory system to generate highly differentiated and unforeseen perceptions. By this, we hope to aid in the generation useful intuitive research hypotheses for later quantitative verification.

Electronic music pioneer Pierre Schaeffer [13] promoted the role of perceptually uncategorized sound sources and *acousmatic listening*, in which sound is experienced as removed from its source, as an abstract quality, making subtle differentiations into focus. Following his ideas, we suspect that a sound source that can easily be categorized may effectively obstruct the emergence of new auditory patterns by encouraging a reduction of its sound qualities to a low dimensionality, for example pitch and instrument type. Michel Chion has named the listening mode in which the sound is perceived as an abstract quality "reduced listening" [6].

2.2. The need for scalable representations

The work of Shneiderman on data visualization [2] characterizes the process of data browsing as a repetition of the sequence *Overview first – zoom and filter – then details-on-demand*, which already hints at a scalability requirement of data representations. However, the proportions of many databases accumulating in contemporary research projects are often vast: The collection of ORFs (protein sequences) created by the GOS [15] currently yields 17.4 million entries. This puts increased demand on the process of "zooming", as structural features in the relationship and distribution of these myriads of datasets may appear on many different levels of scale.

Visualization is a mostly space-based display mode, and scaling visual representations in the visual domain means increasing their size while adding detail. Sonification per-se on the other hand is a mostly time-based: The amount of time a single data object may occupy when such a database is browsed for global overview is microscopic. The details of a single dataset however, its specific meta-data and references can be quite deep, and can take a lot of time to fully and comprehensively unfold and be absorbed by the interested participant.

We want to enable an auditory representation of the data object that can be very small and short if the database is browsed for a global overview, with patterns and structures emerging from the distribution of the values among the objects coming to the fore, while being able to scale in size up to a point at which the entire display system as well as the attention span and inner canvas of the participant with representations of a its set of descriptions and embedded references.

Thus we are able to formulate the requirements that scalable auditory data signatures need to fulfill as a tool for exploratory data browsing:

- Enable a comparative representation of similarities and differences in multidimensional datasets on multiple levels of scale and detail.
- Create overviews of the range of parameter values present and the structural properties of the underlying database
- Enable a polymorphism of audiovisual descriptions, providing for a transparent integration of qualitative as well as quantitative elements for different levels of detail.
- Enable continuous scaling

Next to these requirements for the data representation itself, the use of audio also needs to provide functional feedback for navigation and interactivity used for the browsing activity:

- Link perceptual layers of the audiovisual presentation of the data object together through the employment of spatial and temporal syncretisms and other signifiers.
- Denote changes in perspective and transitions between different levels of detail
- Clarify interactions, aiding in spatial navigation

2.3. Auditory domains

Our everyday auditory experience involves the perception of many richly differentiated auditory events. While it would be interesting to tap into our entire range of possible auditory perceptions for sonification, the auditory domains used for qualitative representation will have to be constrained to the range of concrete sound synthesis processes that allow for perceptually relevant multidimensional modulation.

For a successful qualitative representation of datasets, the dimensional axes along which subjective perceptual differentiations occur need to be made accessible for parameterization and mapping.

In our current application of SADS for ATLAS in silico, two classes of sounds are used for this purpose that are both highly accessible to parametric selection or synthesis and provide seemingly infinite amounts of perceptual differentiation: Impact sounds and additive sound spectra.

To link a broad range of sound layers together with possible visual representations of the data object in order to form a single perceptual entity, specific temporal behaviors are embedded into SADS referring to participant interaction and activity. This concept of a hybrid audio-visual interactive data object is contextualized by Michel Chion's observations of audiovisual perception in films [6].

2.3.1. In milliseconds: impact sounds, micro-rhythms and syncretism

The perceptual richness of impact sounds is evident from our everyday experience. With astonishing speed and precision, this class of sounds conveys information about the nature of colliding objects - about their material, mass and spatial location. Murray Schafer writes in his famous book The Soundscape "no two raindrops sound alike" [7] and Curtis Roads underlines the perceptual importance of sonic micro-events and their combination in his book Microsound [8].

Next to the utilization of single impacts, an additional layer of sensory complexity can be afforded by grouping the impacts into micro-rhythms or micro-gestures of varying speed and density, exploiting the superior resolution and memory of the auditory sense for temporal structures.[6]

A third affordance of this class of sounds is their ability to create an audio-visual perceptual synthesis which Michel Chion calls *syncretism*: Impact sounds are presented in precise temporal coherence with changes in accompanying visuals and auditory layers during interaction with SADS, establishing a perceptual coherence between the different layers of sound and the visual aspects of the data representation. This is an auditory phenomenon akin to the characteristic attack transients of instrumental sounds and allows a contextualized scaffolding of the following sound content, helping to link the elements into a single perceptual entity.

2.3.2. *Spectral textures and harmony/tonality*

While impact sounds allow for a grouping of audiovisual display elements through synchronization with an initial impulse, sustained spectral textures can be used to prolong this context through the process of continuous scaling in the temporal domain.

In contrast to impact sounds, which tend to occupy extremely short time spans, synthetic sound spectra of static frequency composition have no inherent temporal constraints. Special attention is required to the internal temporal behaviors they may contain (e.g. through partials beating against each another). Their seemingly infinite perceptual differentiability has been conceptually demonstrated by the installation “Audio Fraktal” [9]. A spectrum can be used to provide a temporally invariant sound layer functioning as a link between all sonic events appearing during its presence and the data object that it represents.

Impact sounds can be presented in great temporal density and still retain their perceptual differentiability, additive spectra however can only be meaningfully presented one at a time [10], as the superposition of several static spectra would in most cases result not in two simultaneous spectra, but in a new, third spectrum (the spectra can however be differentiated if their partials are linked together by a common fate – such as a common amplitude or frequency modulation - a strategy that is worthwhile to explore but yields complex problems beyond the scope of this presentation). The presentation of a single discrete spectrum at any given time can thus yield information about the data object that is currently in focus and simultaneously become a representation of the current position of the participant in the traversal of the database.

Impact sounds and spectra are used in SADS as orthogonal sonic elements delineating unified data-driven auditory entities. While the impact sound and the auditory and visual changes that come in syncretism link the layers of the data representation together via interaction, the spectrum forms an underlying sustaining tonal layer that groups other sonic elements such as rhythm or percussive sounds to the entity. At the same time, both can represent different portions of the dataset qualitatively. Thus, impact sounds and spectral textures afford the shaping and contextualization of infinitely varied auditory identities, which mutually enhance explorative listening for qualitative features and enable an even broader mapping of multidimensional qualities in data sets.

2.3.3. *Spatialisation and scaling*

Positional sound is used in SADS to enhance the creation of perceptual identities as well as polyphonies. Auditory content is spatialised in coherence with the display of the visual object representing the same dataset, helping to link sonic elements and visual representation together. Furthermore, to enhance the transparency of a detailed data representation, audio is represented through several independent streams of sound coming from different directions [11]. This scalability of sound spatialisation from the positioning of a single sound source to the creation of a multi-stream sound environment is an important aspect of scalability in SADS.

Using the three strategies as combined interactive device, SADS are able to represent multidimensional data values qualitatively through multiple simultaneous layers of auditory and visual representation while preserving the impression of a single audio-visual perceptual entity.

2.3.4. *Micro-types, Macro-types and temporal scaling*

Spatialised impact sounds and spectral textures belong to a class of auditory *micro-types:* Highly differentiable sound elements that are devoid of inherent symbolic value, requiring juxtaposition and subjective comparison to other elements of similar type in order to carry meaning. They can be described as two categories of qualitative perceptual value.

On the other hand, many forms of auditory representation are characterized by their temporal evolution and do formulate interior grammars. Among the known examples of this category are speech, rhythm, harmony and melody, that each rely on a specific temporal sequence of micro-types. We therefore call these auditory types *macro-types*.

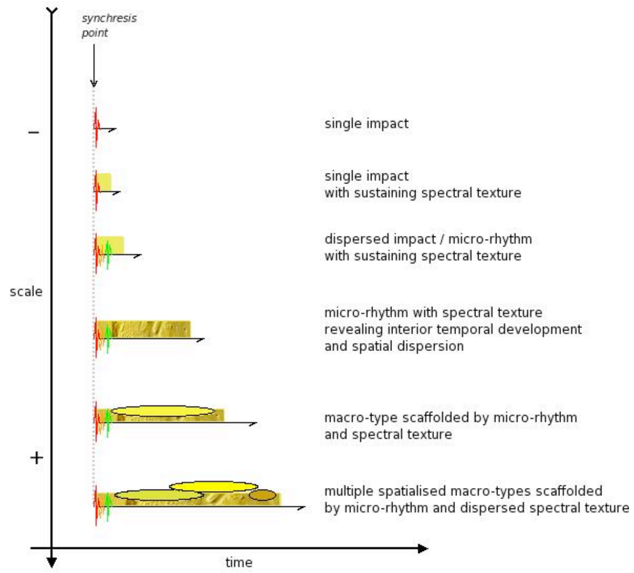

Figure 1. *Scalability*

3. USE CASE: ATLAS IN SILICO

3.1. Installation Concept

ATLAS in silico enables participants to explore relationships within data that spans from the molecular (nano-scale of protein molecules) to the global (geo-referenced sampling site, socio-economic and environmental data) by interacting with luminous and colorful 3D graphics and a responsive, data-driven sonic microworld within an immersive virtual environment constructed from contextual metadata. This context animates the virtual world as a driving force, much like natural ocean currents, revealing internal structure within the data and metadata - bridging out from the nano scale to the global and back again creating a multi-scale, multimodal multi-resolution experience. Underlying this experience are the 17.4 million protein sequences comprising the Global Ocean Survey (GOS) metaganomics dataset, the metadata describing the entire data collection, and associated metadata describing socio-economic and environmental factors of geographic regions where biological samples for the GOS were collected. The massive scope and multi-resolution nature of this data impelled our approach for auditory representation and provided us with a use case in which to explore the applicability of scalable auditory data signatures in an expressive and artistic context.

3.2. A multi-scale experience

The data- and interaction-driven motion and transformation of objects within the virtual environment responds to different accessible viewing resolutions and scales: While the global view presents all objects in the database simultaneously in a particle system simulation, subsequent zooming reveals selections of objects and their interior structure. The sound responds to the particular level of detail, revealing more and more features of single objects in scaled perspectives.
The scaling parallels a transition from qualitative to quantitative data representation, revealing more and more of the meta-data values and contexts explicitly.

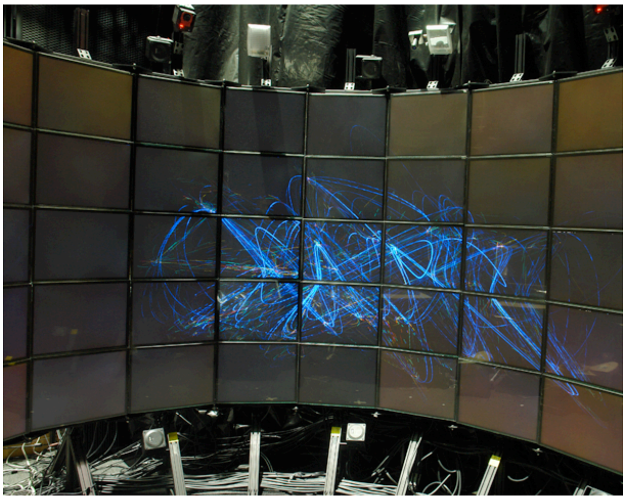

Figure 2. Display system and global view of particle simulation

3.3. Scaling Data-Signatures in ATLAS in silico

In order to afford open-ended exploratory listening, SADS in ATLAS generate perceptual identifiers driven by the data values of database entry. These identifiers are expressed and unified by a combination the two auditory domains described: Impact sounds as well as additive spectra or harmonies that afford an auditory scaffolding for more complex types of audio, such as speech, melody or spectral transformations.
In ATLAS, three scaling levels are currently implemented:

3.3.1. Scale 1 – global browsing with micro-signatures

In a global overview, the user is immersed in a simulated particle system that can be splashed through by hand. The hand, which reaches into the virtual world by optical tracking, repels the particles when moving and attracts them when standing still. As particles collide with the hand reaching into the virtual environment, the micro-signatures sound.

SADS are represented on their smallest scale, consisting of a single data-made or data-selected impact sound. The impact sound, albeit extremely short, is specific to the represented data object and allows an intuitive interpretation about some of its data features. A syncretic link between auditory and visual representation for individual objects does not exist on this scale due to the possibly microscopic time span between two consecutive events and the comparatively slow visual frame rate.

3.3.2. Scale 2 – a selection of objects

The second scale surrounds the user with a scaled view of selected data objects. The data-made visual objects can be touched with the participant's tracked hand, causing the touched object to display a data driven auditory and visual cue that emphasize the object perceptually, form a quasi-haptic feedback and at the same time represent a qualitative display of the dataset.
The SADSs on this scaling level start out with the same impact sound, which is syncretically linked to the start of its signature spectrum as well as the changes in its visual representation. The sound spatialisation aids to localize the sound pertaining to a specific dataset to the location of its visual representation. The initial impact sound is however expanded into a micro-sequence of percussive events that presents additional information about the data object. In the current implementation of ATLAS, the impact sequence represents the amino acid sequence of the underlying protein sample, while the signature spectrum is composed from the mass distribution of the different amino acids within the molecule.

3.3.3. Scale 3 and beyond – toward a full audio-visual representation of the data object

On a third scale of the SADS, the entire display system is filled with representations pertaining to a single dataset, revealing its interior structure and data values. The initial impact is followed by a spectrum, that in turn breaks up to form melodic or pitched sequences, containing additional sequences of impact sound.
On this level, which displays only data representations pertaining to a single dataset, the entire space of the display system is used to immerse the participant in polymorphic data representations for that data object. The spatial localization of the sounds is no longer used to bind the sounds to a data object, but to increase the number of simultaneous auditory streams that are presented to the user. Voice and additional audio-visual narrative devices such as animated graphic elements are used to provide insight to the contents of all data fields of the selected single data object. The context and original characteristic identity of the data object is sustained however through the use of its spectral signature as tonal material, ringing through all aspects of auditory representation of the data object.

3.3.4. *Mapping and scale*

The impact sequences and spectral textures were derived from the structural and chemical description of the ORF-sequence stored in the database. Contextual meta-data are exposed at higher scaling levels.
On the smallest scale (3.3.1), the sound of the impact representing a single dataset is determined by the mass and an averaged index value representing the chemical composition of the molecule.
At the next scale level (3.3.2), the spectral texture is composed by a frequency mapping of the value of molecule mass per amino-acid within the protein, while a micro-rhythm of impacts is generated from the property and placement of structure-defining elements in the protein-sequence, thereby generating a rough temporal impression of the molecule's sequential structure.

On the level of greatest zoom, the spectral texture generated from the molecule masses per amino-acid is treated like a musical mode or pitch repository defining the auditory identity, the melodies and intervals that come forward function as quantitative display of properties of the sampling site at which the molecule was found during the expedition of the global ocean survey, such as the sample depth/water depth ratio, the type of habitat, the water temperature and salinity, etc. [15]. Additional data is explicitly added via a spatialised chorus of spatially distributed synthesized voices.

In the current implementation of ATLAS in silico, this data-mapping strategy is static, and cannot be changed by the participant. We are aware that this mapping contains many arbitrary choices and weightings. Other publications rightfully emphasize the necessity of an interactive loop in the process of data mapping, for example to highlight specific features for a more specific comparisons. Especially the smaller scales, which necessarily have to limit the data fields can influence the mapping in a meaningful way represented could use the possibility to choose which fields specifically influence the generation of the sounds. An interactive mapping process was beyond the current scope of this work as its target application is a public installation affording immediate and intuitive interaction even without specialist knowledge.

3.4. Data display and acousmatic composition

Next to the purpose of presenting the prototype of an intuitive browsing environment for one of the vast multi-scale, multivariate databases of our time, ATLAS in silico is an installation that immerses participants in an audiovisual environment that dramatically evolves with the participant's actions. A continuous consideration in the design process that accompanied the optimization of data transparency and the emergence of auditory structures were sensibilities applied to the use and context of the sound in terms of its physicality and kinesis – in other words, criteria derived from compositional thought processes. This lead to the implementation of specific sound behaviors and layers that contextualize the current scale perspective with its specific afforded interactions, the enrichment of transient behaviors between the different scales with kinetic and quasi-haptic sonic attributes. Each of the three scaled perspectives employs a distinct continuum sound and composed transitions, inviting to perceive each subsequently smaller scale as a decomposition of the previous one, lending physical allusion and material associations to the disembodied world of abstract data. The result can be described as a composed vessel filled by data-controlled sounds: The transient and interactive behaviors are not static, but reference the auditory display of the interactively browsed data objects or the dynamic properties of the interaction, thereby simultaneously generating an overview and a reference point and interaction feedback for the browsing activity.

3.5. Generating and rendering sounds – aesthetics vs. parameterization

Another important impact of the ambition to make this installation work not only as a data display but also an interactive acousmatic composition was the choice of the sound synthesis strategy used for sonification. The original plan to base the sound material on *data-made* sounds using multi-parametric sound-synthesis as in [9] was changed towards the creation of vast sample banks that allow ATLAS to achieve a similar result through *data-selected* sounds.
While sound generated from arithmetic lends itself to easy multidimensional parametric manipulation, it is not trivial to produce sounds that reach the perceptual richness and aesthetic quality of sampled real-world sounds. After employing different strategies to synthesize impact sounds from granular synthesis and physical modeling such as has been described for W.Gaver's partially related work on auditory icons [5], we were still unsatisfied with the sensory qualities of the resulting sounds and decided to use an existing sample collection of 3000 grains of rice dropped on 40 different materials ranging from metal to stone. The events were normalized and categorized along axis of different spectral sound features, allowing for data spaces to be mapped to the repository of rice sounds.
The spectral textures were first generated by additive synthesis (superposition of sine waves). Even though the thus manufactured spectra yield an immense perceptual differentiability, their sound characteristic is peculiarly electronic and does not blend well with the sampled impact sounds and other sample-based elements of the presentation. Since the spectra generated were not deeply complex, consisting of 32 frequencies mostly, the sine waves were replaced with samples of bowed vibraphone. The vibraphone samples are sufficiently sinusoidal and constant in pitch to retain most of the perceptual differentiation, but feature subtle spectral modulations and contain bowing noise, making them more pleasant to the ear and achieve a better blend with the elements of composed kinesis and physical allusion.

3.6. Some technical details

All sound rendering was implemented in Miller Puckette's open source PD [14] environment. The multi-channel sound rendering uses an amplitude and delay cues in different combination on a 10.1 channel system consisting of Meyersound MM-4 and MD-1 loudspeakers. (The detailed description of the spatial rendering mechanisms will be topic of a separate publication.).
The database handling and query is achieved via network and implemented externally. (Technical detail is also available on the project website [16]).

3.7. Future work

While the presented concepts have been confirmed and tested within the ATLAS in silico team and are inspired by audio-visual media theory, an evaluation in form of a user study has not taken place. A traveling version of the piece is currently under development that will be on display at various media arts festivals globally (details can be found on the project website

[16]). The possibility to generate additional projects from this browsing strategy for other vast databases is currently investigated. We hope to implement the possibility of dynamic data mapping in future projects that will allow scientists to browse for specific data features.

4. CONCLUSION

Sonification is a highly interdisciplinary field – scientists and artists are attempting to collaborate in environments of complex technology, often unable to understand each other's language. While the goal of a composer is to work with structures immanent to the perception of sound, or sounding *qualities,* for their own sake, the goal of discovery-oriented scientific work is the generation of intuitions and the discovery of patterns with a final result that can be formally *quantified*. The concept of *qualitative listening* for sonification alienates both parties: While the composer criticizes a possible prostitution of the sound for non-musical purposes and the loss of a sensibility for exactly these qualities, the scientist may not be willing to accept the idea of non-quantified, abstract intuition. The most difficult problem inherent in this seems to be to generate successful descriptions of the inherent structures and delineations of sound elements – this is what we have tried to work toward with Scalable Auditory Data Signatures.

We hope that this approach will help to inspire the design of data representations of the many vast and multidimensional interrelated databases currently accumulating around the world, enabling a successful adaptation of information presentation strategies to the multi-faceted possibilities of human perception.

5. REFERENCES

[1] Peter Medawar, *Pluto's Republic: Incorporating The Art of the Soluble and Induction and Intuition in Scientific Thought*, New Ed (Oxford University Press, USA, 1984).

[2] B. Shneiderman, *The eyes have it: a task by data type taxonomy for information visualizations*. In Proceedings 1996 IEEE Symposium on Visual Languages VL-96.

[3] R. S. Wurman, *Information Anxiety*, Doubleday, New York, 1989

[4] Bruce N. Walker and Joshua T. Cothran, *Sonification Sandbox: A Graphical Toolkit for Auditory Graphs*, in Proc. 2008 Int. Conf. On Auditory Display (ICAD), Boston, MA, U.S.A., July 2008

[5] William W. Gaver, *Synthesizing Auditory Icons*, in *Proc. INTTERCHI 93*, Amsterdam, Netherlands, April 1993

[6] M. Chion *Audio-Vision*. Columbia University Press, New York , U.S.A., 1994.

[7] R. M. Schafer *The Soundscape*. Destiny Books, Rochester, Vermont , U.S.A., 1977.

[8] C. Roads *Microsound*. MIT Press, Cambridge, MA , U.S.A., 2001.

[9] J. Gossmann, "Toward an Auditory Representation of Complexity" in *Proc. 2005 Int. Conf. On Auditory Display (ICAD)*, Limerick, Ireland, July 2005

[10] P. R. Cook (ed) *Music, Cognition and Computerised Sound*. The MIT Press, Cambridge, MA , U.S.A., 2001

[11] A. S. Bregman A*uditory Scene Analysis*. MIT Press, Cambridge, MA , U.S.A., 1990

[12] T. Hermann *Sonification for Exploratory Data Analysis,* PhD Thesis, Tu-Bielefeld, Germany, 2002

[13] P. Schaeffer *Solfege de l'Objet Sonore,* INA-GRM, Paris France, 1966

[14] http://www.crca.ucsd.edu/~msp/software.html

[15] http://collections.plos.org/plosbiology/gos-2007.php

[16] http://www.atlasinsilico.net