COMPUTATIONAL DESIGNING OF AUDITORY ENVIRONMENTS

David Worrall

Audio Arts and Acoustics Department,
Columbia College Chicago,
600 South Michigan Avenue Chicago, IL 60605, USA.
dworrall@colum.edu

ABSTRACT

This paper is a call for sonification designers to adapt their representational practices from that of designing objects for auditory engagement to the construction of systems of formally described relationships that define the 'state space' from which streams of such objects can be drawn. This shift from the crafting individual sonic objects and streams to defining dynamical space of design possibilities we call 'computational designing'. Such sonification model spaces are inaudible, heard only through its instances, or the manifestations of particular *trajectories* through the space.

Approaching the design of auditory displays as computational tasks poses both considerable challenges and opportunities. These challenges are often understood to be technical, requiring scripting or programming skills, however the main challenge lies in computational design thinking which is not best understood as the extension of established designing processes.

The intellectual foundations of computational designing rest at the confluence of multiple fields ranging from mathematics, computer science and systems science to biology, psychophysical and cognitive perception, social science, music theory and philosophy. This paper outlines the fundamental concepts of computational design thinking based on seminal ideas from these fields and explores how they it might be applied to the construction of models for synthesized auditory environments.

1. INTRODUCTION

How sonification designers attempt to achieve an effective communication solution with a sonification design task is affected by many things, including the imagined range of possible solutions for the design (the state space), which in turn, is affected by the tools and methodologies employed, and the skills applied in using them. Attempting to evaluate the extent to which the solution is optimal, and how effective it is in achieving a stated goal, remain open questions because the connection between the individual decisions that are made in the designing process and the way they interact in the solution space are non-linear, at best open to interpretation and at worse a collection of individualized black-box heuristics. Such a description is rarely controversial, even by those calling for robust evaluation and scientific comparison of sonification methods, as it is understood that:

In the context of data exploration, what can be potentially learnt from a sonification is unknown, or at least not defined properly, and therefore it is very difficult to specify an objective performance measure. [1]

Design is a messy business and the relationship between decisions made in the process of 'tweaking' the contribution of individual parameters in the final result is rarely the sum of simple linear combinations. So, being able to evaluate the effectiveness of a sonification for a clearly defined purpose is not the same as being able to determine what aspects of the sonification are responsible for or contribute to that effectiveness. This is no different in form to being able to construct a neural-network to simulate an observable behavior without being able to understand the complexity of the individual contributions to the network itself.

2. CURRENT DATA SONIFICATION MODELS

Various methods have been developed to use synthesized sounds for the purpose of sonifying data. [1] Perhaps the simplest, or at least the most direct is *audification*, a "direct translation of a data waveform to the audible domain." [2]. Audification may be applicable as a sonification technique to many data sets that have an equally spaced metric in at least one dimension. It is most easily applied to those that exhibit oscillatory time series characteristics, although this is not a requirement.

In Model-Based Sonification [3][4], a variable of the dataset to be sonified is assigned to some structural properties of a component (elasticity, hardness etc) of the model. A user interacting with this model via 'messages'– virtual beaters, scrapers, etc.–causes it to 'resonate'. The resulting sound is thus determined by the way the data integrates through the model under excitation by the messages. By virtually beating, plucking, blowing and scraping the model (the characteristics of the dataset) are available to the listener in a way analogous to the way that the material and structural characteristics of a physical object are available to listeners who beat, pluck, blow or scrape it.

Parametric-Mapping Sonification (PMson) is the most widely used sonification technique for representing high-dimensional data as sound. Typically, data dimensions are mapped to sound parameters: either to physical (frequency, amplitude), psychophysical (pitch, loudness) or perceptually coherent complexes (timbre, rhythm). At its simplest, PMson uses homomorphic mapping in which the changes in one

---

[1] Data sonification is the acoustic representation of data for relational interpretation by listeners, for the purpose of increasing their knowledge of the source from which the data was acquired.

dimension of the auditory space tracks changes in a variable in the dataset, with as few mediating translations as are necessary for comprehension [5]. There is a kind of degenerate approach to PMson that uses sampled rather than synthesized sound objects. Such an approach limits the parameter-mapping to gross sound parameter control without the options of sophisticated modification that synthesizing sound objects provides. There are some instances where a combination of the two techniques is appropriate.

PMson is the closest data sonification method to the traditional musical model of an abstract score (the data) rendered into sound. Largely because of the number of data dimensions it can potentially carry, PMson remains the most flexible and potent method for embedding higher-order hierarchical percepts such as timbre and meter, which is necessary to overcome some of its limitations, such as the parameter-mapping problem which arises from the non-orthogonal nature of the hearing system [6][7][8][9]. In order for this embedding to occur, we need to develop tools and techniques to compute user navigable auditory environments (sonifications) from a conceptual infrastructure in which are embedded cognitive and psychoacoustic supports for the basic techniques.

1. SONIFICATION DESIGNING AND MUSIC COMPOSITION

Many sonification designers are also music composers who have spent hundreds of hours, not unusually from a very early age, learning to listen intently and experiment intelligently. They have learnt to compose sound combinations that either crystalize into harmonically coherent complexes (harmonies) or individual streams of sound (musical lines) that maintain various degrees of independence from each other when sounded simultaneously (counterpoint). Many composers also undertake studies in orchestration, which involves both learning how, over hundreds of years, composers of different of styles of music have scored harmonic and contrapuntal ideas to different effect, as well learning how to score abstract musical structures for performance by a wide variety musical instruments for their own compositions.

Composers thus bring to the task of data sonification a plethora of skills and experience that have been black-boxed into their individually developed Heuristic Auditory Environment Synthesis (HAES)[2] techniques. Many of the techniques they have acquired have their foundations in psychoacoustic principles even though they were developed, often over centuries, by composers who listen, rather than by psycho-acousticians or cognitive psychologists. There are a large number of such "rules-of-thumb': those for maintaining linear integrity by avoiding parallel fifths and octaves, for example, or for where in the tessitura of various instruments of different timbres the attack transient onset-times are similar, thus affording the smooth dove-tailing of multiple-octave scalar sweeps, or how to adjust relative onset times to ensure simultaneities (chords) are heard as such.

It has become increasingly recognized that design is not free expression and is not scientific engineering; it is in-

---

[2] A term I invent here to distinguish the process from Computational Auditory Environment Synthesis (CAES), discussed later. The term Auditory Environment Synthesis is preferred to Auditory Scene Synthesis (ASS) as it emphasizes the three-dimensional surrounding nature of sonically being in the world.

between (a 'third culture'), sometimes fluxing more towards one than the other; the application of scientific *and other organized knowledge* to practical tasks. Today, these increasingly artificial or designed worlds are also digital. In fact, they rely on digitization. Yet, just as the designed world is rapidly replacing the natural world as the principal mode of (humans) being in the world, it is widely recognized that many design problems are ill-defined.

> *They are not the same as the "puzzles" that scientists, mathematicians, and other scholars set themselves. They are not problems for which all the necessary information is, or ever can be, available to the problem-solver. They are therefore not susceptible to exhaustive analysis, and there can never be a guarantee that "correct" solution-focused strategy is clearly preferable to go on analyzing "the problem," but the designer's task is to produce "the solution."* [10]

There are some efforts to make sonification a well-defined scientific method, i.e. the "data-dependent generation of sound that uses systematic, objective and reproducible transforms" [11]. However, the act of designerly producing a sonification obscures its methods of production, leaving an object that is resistant to analysis, or at best not conducive to it. So, this 'covering up' of the design process does not assist the development of better sonification practices. Nor does it promote the reproducibility requirement which has been declared by some as necessary for sonifications to be called scientific:

> *"… given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical."*
>
> *The definition claims reproducibility. This may not strictly be achieved for several reasons: … The use of the term "structurally identical" in the definition aims to weaken the stronger claim of sample-based identity. Sample-based identity is not necessary, yet all possible psychophysical tests should come to identical conclusions.* [1]

The above discussion suggests a need for sonification design decisions to be derived from more explicit criteria than the educated responses of designers with black-box skills. It is not just a question of the importance of user-centered interactive design replacing fixed-format audio files. What is needed is a deeper and more encompassing codification of the design process itself, including all the relevant black-box contents that go into heuristic decision-making processes, the translation of historical and current knowledge into computer-actionable knowledge banks and the ability for such systems to both suggest novel solutions and to learn and adapt to individual designers' styles. This is a non-trivial task but one that is being attempted in allied disciplines such as audio game engine design.

3. COMPUTATIONAL DESIGNING[3]

The dominant mode of utilizing computers for audio design and production today is computerization or compilation: sonic objects or processes that have been conceptualized in

---

[3] This section relies heavily on ideas developed through the synthesis of multiple authors in [12].

the designer's mind are recorded, manipulated and/or stored in a computer using a Digital Audio Workstation (DAW). Typically, a DAW consists of a computer with sufficient storage and processing speed to be able to mix multiple simultaneous channels of audio to which pre-programmed processes are applied (often in the form of 'plugins') by software (such as *Protools*, *Abelton Live, Logic* and *Reaper),* which are used to edit, mix, transform, and store or simultaneously play back those audio channels for direct audition.

From a design perspective, this computer-aided, compilation approach does not allow the designer to take advantage of the *computational* power of the computer in the design process itself.

> *The manifest form - that which appears - is the result of a computational interaction between internal rules and external (morphogenetic) pressures that, themselves, originate in other adjacent forms (ecology). The (pre-concrete) internal rules comprise, in their activity, an embedded form, what is today clearly understood and described by the term algorithm.* [12]

Expressed simply, computational designing employs computation in the design process itself to *deduce* and place elements, for example, real-time synthesized sounds including microsound responsiveness to situational criteria, user input and the like. Computational design is best understood when compared to computerized (or computer-aided) design, in which the computer is used to *compile* and arrange fixed design elements such as transforming pre-recorded sound samples to better fit the specific situation in which they are to be used.

Computerized design is based on a data model, whereas computational design relies on a procedural model. Computational design involves the processing of information and interactions between elements that constitute a specific environment. The final appearance of these elements, whether they be game objects or sonified information derived from data, is rendered from the processing of intrinsic properties, such as the specific values of data points, important information beacons, or extrinsic properties such as the positional rendering of the object in the acoustic environment in which it is being placed, taking into account the effect (salience, occlusion[4] etc) of other objects that have already been or will be placed there. Computation provides a conceptual framework for highlighting the data being rendered according to the importance placed on it at the time by the designer and the interacting user. As a design methodology, whereas computer-aided design begins with *objects* (such as sound samples) and adapt them to specific situations, computational designs start with *elemental properties* of objects (as synthesis parameters) and environmental influences, and use generative rules to formulate (or proceduralize[5]) the specific objects in the specific environment into which they are placed.

---

[4] The salience of a sound is its attention-grabbing or distinctiveness[17]; occlusion refers to the virtual hiding or masking of a sound by others. These characteristics can be altered using signal processing techniques such as filtering and reverberation.

[5] The terms Procedural Design and Computational Design are frequently used interchangeably. Procedural Design is the more general term. When they are computed, Procedural Designs become Computational.

Most of the work of the computational designer involves explicitly defining and editing the definition of sets of variables such as psychoacoustic parameters and constraints. In generating specific solutions, logical operations on these sets and their (often dynamically generated) subsets are performed without the designer necessarily being able to conceptualize the full formal implications of their relationships. This can be a positive consequence of such abstraction as it can produce state-space solutions that might not have been intuited, considered or imagined using non-computational approaches. Because the designer is freed from the requirement to produce a single 'masterful' solution, many instantiations can be produced and then evaluated for their effectiveness, in-keeping with the goals of the SonEx project [1].

3.1. Most music is composed procedurally

Adapting Schön and Wiggins' three types of seeing for architectural designing [13] to listening, we have

1. Literal aural apprehension of auditory objects–sensorially-led appreciation.
2. Appreciative judgements of quality (tone, texture, timbre, pitch and, duration), discovered by more intellectually–led reduced listening, and 'deep' pattern recognition.
3. Apprehension of auditory state-space gestalts which are not instantaneously heard or imagined. Such higher-order designs are well known to be important in music and include things like beat, pulse, meter, swing, scales, modes, chords, etc;–higher level perceptual groupings that afford memory retention, comparison and contrast.

Most composing undertakes these steps in reverse: General conceptual ideas and gestalts like melodies and harmonies precede thoughts about instrumentation and scoring. We can observe this in music that can be performed on multiple instruments: it is the structural organization of the sounds which becomes primary, not the sensation of the tones used to "carry" the organization. This principle is also exemplified by the fact that, historically, many large-scale instrumental compositions were composed first into a keyboard score, with specific instrumental orchestration of these ideas being undertaken as separate and later processes.

Generative music composition has a long tradition, pre-dating the invention of the digital computer by several hundred years. With the advent of the computer, it quickly became formalized [14]. Equipping machines with the ability to play a vitally important role in achieving musical tasks, such as composition, improvisation, and accompaniment, is now an important development in innovative musical practices.

3.2. Digital sound synthesis

In addition to their use in compiling audio, computers can also be used to digitally generate or synthesize new sounds. Such computation has had a profound impact on our ability to produce conceptually simple sounds such as sine tones as well as sounds of great complexity such as stochastically controlled microsound-grains as well as complex sounds using simulations of physical resonators such as acoustic musical instruments. Early work by pioneering composers and engineers such as Xenakis, Risset, Chowning, Mathews, and Moorer, was seminal in establishing both tools and techniques for exploring and understanding the dynamic

nature of natural sounds and thus the design of 'lively' and interesting computer-generated tones and the multi-reverberant environments in which they are formed.

The synthesis of individual sounds is an intensely computational process and algorithms for doing so using granular, physical models, fixed-waveform and other hybrid techniques are available today in a number of mature, well-honed applications such as *Csound*, *Supercollider* and *Pure Data*. Though not sufficient in themselves to design sophisticated auditory environments, such programs already contain the necessary sound-synthesis tools for supporting a computational approach.

3.3. Some advantages of using a computational approach for designing auditory environments.

While there may be some circumstances in which auditory icons[6] and earcons[7] might be sufficient for creating a useful sonic environment, having to rely on them, with some general sonic smudging to attempt to smooth over the cutting-and-pasting, is not a recipe for the development of more sophisticated and responsive sonifications which are clearly needed as the understanding of the psychoacoustic and cognitive correlates of data sonification increases. As the complexity of an environment increases and/or the number of object in it increases, computational load of rendering all auditory objects becomes a critical defining feature of how many such objects can be rendered. In a sample-based model, either the library of different samples for an increasing number of sonic objects has to be the generated and compiled, or the processing requirements for modifying a smaller subset of each object has to be increased.

In a computational design model, the rendered form of both individual auditory objects and general environmental factors–and the ways in which they interact–can be dynamic and highly flexible. This increases variety and reduces the reliance on the modification of decisions that need to be made in advance when using a samples model. For example, whereas sound samples might need to be modified according to situational salience requirements, in a computational model, the salience of objects can be made a feature of the synthesis of the objects themselves. In a sonic environment which is responsive to user-directed interests, for instance, this affords the production, of a better balance between local and global reverberation requirements, resulting in a deeper sense of environmental continuity and sound-object integration.

Another advantage of using a procedural approach is that, as the level of detail needed and the number of objects increases in the rendered auditory environment, the overall computation time involved relative to that required by manipulating samples, significantly reduces.

4. A CASE STUDY: AUDIO IN COMPUTER GAMES

A game world is composed of discrete entities (objects) that have a set of properties (name, visual appearance, location, sound signature etc) and a set of behaviors and sounds associated with those behaviors as they interact in the world,

---

[6] Auditory icons are "sounds which mimic everyday non-speech sounds that we might be familiar with from our everyday experience of the real world" [15]

[7] Earcons are "…short, structured musical phrases that can be parameterized to communicate information in an Auditory Display." [16]

and/or the world in which they exist interacts with them. Early games used generative techniques to produce sound-effects and musical accompaniment before computers became powerful enough to use sampling technology to deliver 'realistic' recorded sounds. Because a sound recording captures the digital signal of a single instance of a sound and not its dynamic behavior, many clever tricks have been developed to blend, layer, filter, interpolate and time-warp these singular sound 'instances'. What this has meant for game audio over the last two decades is that overwhelmingly, the auditory design approach has remained *event based* (i.e using sound samples) despite Farnell's prediction that "procedural sound is set to make a huge comeback":

> *Traditional game audio binds each action to an event which triggers a sound sample. Some real-time modifications can be applied such as distance damping, or combining alternative samples in a random or granular way to get more variation. But none of these techniques is able to map the underlying physical behavior an object and its sound. ...an equivalent event-based graphical game would only be a series of static photographs. …*
> *In software engineering terms, games audio is badly coupled and inconclusive ...* [18].

This event-based approach causes timing problems such as the aligning sample loops with game movements, thus limiting visual and/or audio actions to a predetermined duration and raises the need to devise new methods, such as randomly selecting samples and sub-samples, to alleviate the repetitive quality of a limited data set. What is somewhat surprising is that although the use of computational generation of dynamic sound objects is being introduced to game design, the approach has not (yet?) made the huge comeback Farnell predicted. The most common usages seeem to be in continuous and enviromental sounds:

> *Our procedural audio synthesis technologies use minimal amount of samples to generate a great variety of high-quality sound events. More importantly, our synthesis engine comes with specially-designed controls for intuitive and instant sound manipulation. Currently available technologies include wind, rain, fire, electric sparks and space ambience synthesizers. Some more are coming.* [19]

Joe Cancellaro, composer and game design educator informally described the current state in this way:

> (procedural audio) *did not make a huge comeback. Sampled audio is pervasive in the game industry. What has changed significantly are the audio engines that deliver the samples, processing them (samples) in real-time, and in many cases altering the original sample. This could be seen as procedural but is it not really. There are examples where procedural sound is used in games, but it turns out it is less effective, emotionally, than samples. It lacks a "human" touch. For big pads and sustained environments procedural could work, but even then, I would go with a composed sample and let the non-linear audio engine chew it up and make something contained but random. It's all Stravinsky at that point. Shackles!* [20]

Overtly emotive sounds have a more function functional role in developing and enhancing satisfying game experiences than they do in data sonification tasks. So, it will be interesting to observe if, as game worlds increase in size and complexity, the time critical interactivity requirements that gamers demand will eventually force designers to rely more on computational sound design. Perhaps the increasing computer power of the platforms on which games are delivered will be sufficient to permit game sound designers to continue using sound samples as a means of delivering satisfying user experiences in the foreseeable future.

5. NEW DIRECTIONS IN COMPUTATIONAL DESIGNING OF AUDITORY ENVIRONMENTS

This paper has its origins in thinking about The Mapping-Problem and a developing sense that, although there is a general understanding of the Problem there seemed be no solutions to circumventing it that don't involve the compilation and integration of a large body of psychoacoustic and cognitive knowledge in a sonification design framework.

Incorporating task-based analysis into design criteria is increasingly understood to be an important step in developing effective data sonifications. At the same time, there is a need to develop larger solution state-spaces with an increased number of dimensions, incorporating both microsound and gestural levels and the intelligence to form multiple mapping solutions from correspondences between them which produce both highly dynamic and situationally responsive individual sound objects and higher-order extra-objective perceptual experiences such as swing, which is produced by systematically modulating the temporal flow.

The ability for designers, and ultimately uses, to adapt a data sonification to their aural developing skills is also important. This is supported by Jean Piaget's observation of an interesting aspect of the relationship between representational and perceptual space. During the development of an understanding of a representational space through experimentation (or 'play'), representational activity is 'reflected' or 'projected back' on to perceptual activity [21] as exemplified in the way understanding of musical structures affects the way one perceives musical affect. This observation supports the hypothesis that for a listener, there is dynamic relationship between their critical listening skills, and the sonic complexity and variability of a sound mapping that can be understood. This emphasizes both the need to have auditory designs which are responsive and/or able to accommodate different and developing listening skills, but also the need to develop the integration of the development of listening and sound mapping experiences in early childhood education curricula.

If perception is not solely a bottom-up process but incorporates inherited or acquired response biases related to stimulus characteristics and sensory systems that have evolved for multiple, integrated representations of certain features which are meaningful to the perceiver rather than for just single one-to-one reproductions of the physical world, it makes sense to generate multiple sonifications of any particular data set according to the user's developing sense of meaningfulness. In order to accomplish this, I suggest it is necessary for sound designers interested in creating dynamic auditory environments to shift their design representation and thinking efforts from creating bespoke hand-crafted solutions using just their own black-boxes, to a dynamic system model approach in which design activity is supported by extended

state spaces that incorporate psychoacoustic and cognitively-informed transforms.

Because of the breadth and depth of detail required, sets of such transforms need to be developed collaboratively using agreed-upon standards. A community-based approach to developing these resources should go some way towards increasing the computational power of the computer as a design tool to *deduce* a wider variety of more effective auditory designs.

6. REFERENCES

[1] Degara, N., F.Nagel, and T.Hermann, 2013. SonEX: An evaluation exchange framework for reproducible sonification, in *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 167-174.

[2] Kramer, G., 1994. Some organizing principles for representing data with sound. In G. Kramer (ed.). 1994. *Auditory display: Sonification, Audification, and Auditory Interfaces*. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, Volume XVIII. Reading, MA: Addison Wesley Publishing Company, p186.

[3] Hermann, T., & H. Ritter, 1999. Listen to your data: Model-Based sonification for data analysis. *Lasker, G. E. (Ed.), Advances in intelligent computing and multimedia systems*, 189-194.

[4] Hermann, T., 2002, *Sonification for exploratory data analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, Germany.

[5] Kramer, G., 1994. An introduction to auditory display. In G. Kramer (ed.). 1994. *Auditory display: Sonification, audification, and auditory interfaces*. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, Volume XVIII. Reading, MA: Addison Wesley Publishing Company, p26.

[6] Worrall, D., 2014. Can micro-gestural inflections be used to improve the soniculatory effectiveness of parameter papping sonifications? *Organised Sound, Volume 19, Special Issue 01*, Cambridge University Press, pp 52-59.

[7] Worrall, D., 2013. Understanding the need for micro-gestural inflections in parameter-papping sonification. *Proceedings of the 19th International Conference on Auditory Display*, Łódź, Poland July 6-10, pp 197-204.

[8] Worrall, D., 2011. A method for developing an improved mapping model for data sonification. *Proceedings of the 17th International Conference on Auditory Display*, Budapest, Hungary, June 20-24.

[9] Worrall, D., 2010. Parameter mapping sonic articulation and the perceiving body. *Proceedings of the 16th International Conference on Auditory Display*, Washington, D.C, USA, June 9-15.

[10] Cross, N., 1982. Designerly ways of knowing. *Design Issues*, 3(4), 221-7.

[11] Hermann, T., 2008. Taxonomy and definitions for sonification and auditory display, in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, P. Susini and O. Warusfel, Eds. IRCAM, Paris, France.

[12] Kwinter, S., 2008. Far from equilibrium: Essays on technology and design culture, Actar, Barcelona, p 10. as quoted in A. Menges and S. Ahlquist, *Computational design thinking*. Wiley and Sons, 2011, p 147.

[13] Schön, D. and G. Wiggins, 1992. Kinds of seeing and their functions in designing, in *Design Studies*. pp 135-56.

[14] Xenakis, I., 1971. *Formalized music: Thought and mathematics in music*. Indiana: Indiana University Press.

[15] Brazil, E. and M. Fernström, 2011. Auditory icons, in T. Hermann, T., A. Hunt and J. Neuhoff (Eds.) (Eds.), The *Sonification Handbook*. Berlin: Logos Verlag, Ch 13, p325-338.

[16] McGookin, D and S Brewster, 2011. Earcons, in T. Hermann, T., A. Hunt and J. Neuhoff (Eds.), *The Sonification Handbook*, Berlin, Germany: Logos Publishing House, Ch14, pp. 339-361.

[17] Delmotte, V., 2012. *Computational auditory saliency*, Ph.D. Thesis, Georgia Institute of Technology, Atlanta.

[18] Farnell, A. 2010. Sound Design. The MIT Press., p 318.

[19] https://lesound.io/technologies/

[20] Cancellaro, J., 2017. (Chair, Interactive Arts and Media Department, School of Media Arts, Columbia College Chicago.) Personal communication.

[21] Piaget, J and B. Inhelder, 1956. trans FJ Langdon and JL Lunzer, *The Child's Conception of Space*, Routledge & Kegan Paul (London; New York), p 4.