UCD DUBLIN

MusTWork 16

Proceedings of the 2016
Music Technology Workshop

Establishing a Partnership Between Music Technology, Business Analytics, and Industry in Ireland

Róisín Loughran and James McDermott (editors)

ISBN 978-1-910963-04-3 University College Dublin

Foreword

MusTWork, the 2016 Music Technology Workshop, was held in the leafy Blackrock campus of University College Dublin, Ireland, on Friday 10th June 2016.

The goals of the workshop were to bring together for the first time those working in academia and industry, in the intersection between music technology and business analytics. It certainly succeeded in this. It attracted 50 attendees and guests, mostly from within Ireland, but also from the UK and Canada, with a wide range of backgrounds, interests, and career paths. The workshop also highlighted the opportunities available to graduates of approximately 10 Irish higher education institutions offering technical programmes in this developing field.

The quality of research submissions to the workshop was remarkably high. In the end, all submissions were accepted for presentation – eight as full talks and five as posters. This was in keeping with the workshop's goals of being inclusive and creating a community. The submissions ranged from sonification (with potential applications in business analytics) through frameworks for web-audio apps and demos suitable for commercialisation, through to art music based on computational intelligence methods.

In addition to research talks and posters, the workshop featured an industry panel and an academic panel, and a fascinating keynote talk. This volume collects the research submissions – some abstracts, and some expanded as full papers – a short report on the panel discussions, and an abstract and short biography from the keynote speaker.

We thank all participants for helping to make the workshop a success, in particular our keynote speaker Dr Derry Fitzgerald and industry panel moderator Naoimh O'Connor.

MusTWork 2016 was funded by Science Foundation Ireland and by UCD Seed Funding. We thank UCD Michael Smurfit Graduate Business School for hosting the event and for administrative and audio-visual support, in particular Irene Ward, Felicity McGovern, and Annmarie Connolly.

Róisín Loughran
James McDermott
MusTWork 2016 Organisers
August 4, 2016.

Science
Foundation
Ireland
sfi
For what's next

Panel report

MusTWork16 included two one-hour panel discussions during the day; the first focussed on academic opportunities for students entering or already in the field while the second focussed on industry and commercial possibilities. We would once again like to thank all members of both panels for their time and contributions to the day.

Academic Panel

Participants:
Dr Dermot Furlong — TCD, MMT (`http://www.tcd.ie/eleceng/mmt/`)
Dr Niall Coghlan — DkIT (`https://www.dkit.ie/creative-arts-media-music`)
Dr Brian Bridges — UU (`http://create.ulster.ac.uk/music`) and ISSTA (`http://issta.ie`)
Dr Léon McCarthy — UL, DMARC (`http://www.dmarc.ie`)
Dr Tom Lysaght — NUIM (`https://www.cs.nuim.ie/users/mr-tom-lysaght`)
Dr Emer Cunningham — UCD Graduate Education Development Manager
Dr Róisín Loughran — UCD (Moderator)

We started with a brief introduction from each panelist in regard to their own roles in their home institution. Afterwards the discussion began with a few feeder questions and audience participation was encouraged. The main topics addressed were:

- Suitability of specific courses for prospective students
- Considering a PhD in this area?
    - Advantages
    - Pitfalls
    - First Steps
- The interdisciplinarity of the field
- Departmental awareness (across institutions) of the breadth of the field
- Industry/academic collaboration
- Ownership and intellectual property
- What are the biggest open questions?

Industry Panel

Participants:
Culann McCabe — Rocudo (`http://rocudo.com`)
Dr Amelia Kelly — Soapbox Labs (`http://soapboxlabs.com`)
Killian Magee — Xhail (`http://www.xhail.com`)
Simon Bailey — UCD Enterprise Gateway (`http://www.ucd.ie/research/workingwithus/enterprisegateway/`)
Dr Karl Quinn — NovaUCD (`www.ucd.ie/researchandinnovation`)
Dr Naoimh O'Connor — UCD Research Careers Manager (moderator)

The panel opened with each panelist giving a brief overview of their own personal background and the roles they currently play. Again audience participation was encouraged as the following topics were discussed:

- Career opportunities now and in the future
- How researchers could commercialise their work
- Industry-academia partnership opportunities
- Research opportunities and 'spin-in'
- Start-ups
- Women in tech start-ups

Keynote speaker

We were delighted to have Dr Derry Fitzgerald from Cork Institute of Technology as our keynote speaker who finished the day with a highly engaging talk on sound source separation. Dr Fitzgerald developed software that can model individual instruments and vocals from a mono recording enabling them to be isolated and re-mixed. His methods have led to the first stereo productions of a number of popular Beach Boys tracks which have been commercially released.
`http://nimbus.cit.ie/author/derry/`

Biography

Dr Derry FitzGerald is a Research Fellow at Cork Institute of Technology. Prior to this he was a Stokes Lecturer in Sound Source Separation algorithms at the Audio Research Group in DIT from 2008-2013. Previous to this he worked as a post-doctoral researcher in the Dept. of Electronic Engineering at Cork Institute of Technology, having previously completed a Ph.D. and an M.A. at Dublin Institute of Technology. He has also worked as a Chemical Engineer in the pharmaceutical industry for some years. In the field of music and audio, he has worked as a sound engineer and has written scores for theatre. He has recently utilised his sound source separation technologies to create the first ever officially released stereo mixes of several songs for the Beach Boys, including 'Good Vibrations', 'Help me Rhonda' and 'I Get Around'. His research interests are in the areas of sound source separation, tensor factorizations, automatic upmixing of music, and music information retrieval systems.

Abstract: Sound Source Separation — Real World Applications

This talk provided an overview of a number of sound source separation technologies, including factorisation based techniques, and median filtering based techniques for both vocal and percussion separation, as well as user assisted sound source separation. This was demonstrated in the context of the real-world application of these technologies to the task of upmixing classic Beach Boys tracks such as 'Good Vibrations' and 'I Get Around' from mono to stereo for official release on reissues of a number of Beach Boys albums. It also discussed some of the issues encountered when deploying sound source separation technologies in the real world.

Research papers and abstracts

A Music-generating System Inspired by the Honing Theory of Creativity

Shawn Bell
A.L.C Program, Interactive Media Arts
Dawson College, Canada

Liane Gabora
Psychology Department
University of British Columbia, Canada

Abstract This paper presents NetWorks (NW), a music-generation system that uses a hierarchically clustered scale-free network to generate music that ranges from orderly to chaotic. NW was inspired by the Honing Theory of creativity, according to which human-like creativity hinges on the capacity to (1) maintain dynamics at the 'edge of chaos' through self-organization aimed at minimizing 'psychological entropy', and (2) shift between analytic and associative processing modes. At the edge of chaos, NW generates patterns that exhibit emergent complexity through coherent development at low, mid, and high levels of musical organization, and often suggests goal-seeking behavior. The Core consists of four 16-node modules: one each for pitch, velocity, duration, and entry delay. The Core allows users to define how nodes are connected, and rules that determine when and how nodes respond to their inputs. The Mapping Layer allows users to map node output values to MIDI data that is routed to software instruments in a digital audio workstation. The scale-free architecture of NW's analytic and associative processing modes allows information to flow both bottom-up (from non-hub nodes to hubs) and top-down (from hubs to non-hub nodes).

Introduction

This paper presents NetWorks (NW), a music-generating program inspired by the view that (1) the human mind is a complex adaptive system (CAS), and thus (2) human-like computational creativity can be achieved by drawing on the science of complex systems. NW uses scale-free networks and 'edge of chaos' dynamics to generate music that is engaging and aesthetically pleasing. The approach dates back to a CD of emergent, self-organizing computer music based on cellular automata and asynchronous genetic networks titled "Voices From The Edge of Chaos" (Bell 1998), and more generally to the application of artificial life models to computer-assisted composition, generative music, and sound synthesis (Beyls 1989, 1990, 1991; Bowcott 1989; Chareyron 1990; Horner and Goldberg 1991; Horowitz 1994; Millen 1992; Miranda 1995; Todd and Loy 1991).

First, we summarize key elements of a CAS-inspired theory of creativity. Next, we outline the architecture of NW, evaluate its outputs, and highlight some of its achievements. Finally, we summarize how NW adheres to principles of honing theory and CAS, and how this contributes to the appealing musicality of its output.

The Honing Theory of Creativity

The honing theory (HT) of creativity (Gabora 2010, in press) has its roots in the question of what kind of structure could evolve novel, creative forms effectively and strategically (as opposed to at random). We now summarize the elements of the theory most relevant to NW.

Mind as a Self-Organizing Structure

Humans possess two levels of complex, adaptive, structure: an organismic level and a psychological level, i.e., a mind (Pribram 1994). Like a body, a mind is self-organizing: a new stable global organization can emerge through interactions amongst the parts (Ashby 1947; Carver and Scheier 2002; Prigogine and Nicolis 1977). The capacity to self-organize into a new patterned structure of relationships is critical for the generation of creative outcomes (Abraham 1996; Goertzel 1997; Guastello 1998). The mind's self-organizing capacity originates in a memory that is distributed, content addressable, and sufficiently densely packed that for any one item there is a reasonable probability it is similar enough to some other item to evoke a reminding of it, thereby enabling the redescription and refinement of ideas and actions in a stream of thought (Gabora, 2000, 2010). Mental representations are distributed across neural cell assemblies that encode for primitive stimulus features such as particular tones or timbres. Mental representations are both constrained and enabled by the connections between neurons they activate.

Just as a body mends itself when injured, a mind is on the lookout for 'gaps'—arenas of incompletion or inconsistency or pent-up emotion—which get explored from different perspectives until a new understanding is achieved. We use the term *self-mending* to refer to the capacity to reduce psychological entropy in response to a perturbation (Gabora, in press), i.e., it is a form of self-organization involving reprocessing of arousal-provoking material. Creative thinking induces restructuring of representations, which may involve re-encoding the task such that new elements are perceived to be relevant, or relaxing goal constraints. The transformative impact of immersion in a creative process can bring about sweeping changes to the second (psychological) level of complex, adaptive structure, that alter one's self-concept and worldview.

The Edge of Chaos

Self-organized criticality (SOC) is a phenomenon wherein, through simple local interactions, complex systems find a critical state poised at the transition between order and chaos—the proverbial edge of chaos—from which a small perturbation can exert a disproportionately large effect (Bak, Tang, and Weisenfeld 1988). It has been suggested that insight is a self-organized critical event (Gabora 1998; Schilling 2005). SOC gives rise to structure that exhibits sparse connectivity, short average path lengths, strong local clustering, long-range correlations in space and time, and rapid reconfiguration in response to external inputs. There is evidence of SOC in the human brain, e.g., with respect to phase synchronization of large-scale functional networks (Kitbiczler, Smith, Christensen, and Bullmore 2009). There is also evidence of SOC at the cognitive level; word association studies show that concepts are clustered and sparsely connected, with some having many associates and others few (Nelson, McEvoy, and Schreiber 2004). Cognitive networks exhibit the sparse connectivity, short average path lengths, and local clustering characteristic of self-organized complexity and in particular 'small world' structure (Steyvers and Tenenbaum 2005).

Like other SOC systems, a creative mind may function within a regime midway between order (systematic progression of thoughts), and chaos (everything reminds one of everything else). Much as most perturbations in SOC systems have little effect but the occasional perturbation has a dramatic effect, most thoughts have little effect on one's worldview, but occasionally one thought triggers another, which triggers another, and so forth in a chain reaction of conceptual change. This is consistent with findings that large-scale creative conceptual change often follows a series of small conceptual changes (Ward, Smith, and Vaid 1997), and with evidence that power laws and catastrophe models are applicable to the diffusion of innovations (Jacobsen and Guastello 2011).

Two Modes of Thought: Contextual Focus

Psychological theories of creativity typically involve a divergent stage that predominates during idea generation, and a convergent stage that predominates during the refinement, implementation, and testing of an idea (for a review see Runco 2010; for comparison between divergent / convergent creative processes and dual process models of cognition see Sowden, Pringle, and Gabora 2015). Divergent thought is characterized as intuitive and reflective; it involves the generation of multiple discrete, often unconventional possibilities. It is contrasted with convergent thought, which is critical and evaluative; it involves tweaking the most promising possibilities. There is empirical evidence for oscillations in convergent and divergent thinking, with a relationship between divergent thinking and chaos (Guastello 1998). It is widely believed that divergent thought involves defocused attention and associative processing, and this is consistent with the literal meaning of divergent as "spreading out" (as in divergence of a beam of light). However, the term divergent thinking has come to refer to the kind of thought that occurs during creative tasks that involve the generation of multiple solutions, which may or may not involve defocused attention and associative memory. Moreover, in divergent thought, the associative horizons simply widen generically, instead of in a way that is tailored to the situation or context (Fig. 2). Therefore, we will use the term *associative thought* to refer to creative thinking that involves defocused attention and context-sensitive associative processes, and *analytic thought* to refer to creative thinking that involves focused attention and executive processes. The capacity to shift between these modes of thought has been referred to as *contextual focus* (CF) (Gabora 2010). While some dual processing theories (e.g., Evans 2003) make the split between automatic and deliberate processes, CF makes the split between an associative mode conducive to detecting relationships of correlation and an analytic mode conducive to detecting relationships of causation. Defocusing attention facilitates associative thought by diffusely activating a broad region of memory, enabling obscure (though potentially relevant) aspects of a situation to come to mind. Focusing attention facilitates analytic thought by constraining activation such that items are considered in a compact form amenable to complex mental operations.

According to HT, because of the architecture of associative memory, creativity involves not searching and selecting amongst well-formed idea candidates, but amalgamating and honing initially ill-formed possibilities that may potentially come from multiple sources. As a creative idea is honed, its representation changes through interaction with internally or externally generated contexts, until psychological entropy is acceptably low. The unborn idea is said to be in a 'state of potentiality' because it could actualize different ways depending on the contextual cues taken into account as it takes shape.

The NetWorks System

NW consists of a music-generating system and the music it has produced. The goals of NW are to (1) generate "emergent music," i.e., self-organizing, emergent dynamics from simple rules of interaction, expressed in musical forms, and (2) through emergence, discover new musical genres.

NW is currently configured to explore the expressive potential of hierarchical scale-free networks, as the properties of such networks underlies the interesting dynamics of many real world networks, from the cell to the World Wide Web (Barabási 2002). Musical genres that appear very different on the surface exhibit an underlying scale-free structure; i.e., music composed by Bach, Chopin and Mozart, as well as Russian folk and Chinese pop music have been shown to be scale-free (Liu, Tse & Small 2009). Given the ubiquity of hierarchical scale-free topology and CAS dynamics, it is not surprising that such architectures have creative potential.

NW consists of two layers: (1) the *Core*, which allows the user to define how the nodes are connected, as well as the rules that determine when and how nodes respond to their inputs, and (2) the *Mapping Layer*, which allows the user to map node output values to MIDI data that are routed to software instruments in a Digital Audio Workstation. We now discuss these two layers in more detail.

The Core

A note has five basic attributes: pitch, loudness (usually corresponding to MIDI velocity), duration, timing (or entry delay), and timbre. The core consists of 64 nodes connected in a scale-free architecture, organized into four 16-node modules, one for each basic note attribute except timbre (Figure 1). Four, sixteen node modules allow for 16 channels of MIDI output. The timbral characteristics for each stream of notes is determined by the user by their choice of instruments.

The functions of the entry delay (ED) module are to (1) determine timing, that is, when nodes produce an output, and therefore the pattern of activation across the network, and (2) synchronize the corresponding nodes in each module so that values for pitch, velocity and duration are provided to produce a note (Figure 1). In more general terms, the ED module generates rhythmic patterns via note groupings, from motivic cells to entire movements.

In terms of connectivity, the pitch module is unique in that it includes the largest hub, which sends values to, and receives values from 40 nodes: 12 pitch nodes, and 9 nodes each from the duration, velocity, and ED modules (as well as from itself). It is important to emphasize that hubs send and receive values from hubs in other modules. In this way, note attributes can affect one another; for example, duration can influence pitch, pitch can influence entry delay, entry delay can influence velocity, and so on.

When the nodes are fully connected—that is, receiving values on all their inputs—the network architecture is scale-free; however users can prune the connectivity of the network by reducing the number of inputs to the nodes.

Nodes

Nodes send and receive integers within a range specified by the user (for example 1–13, 1–25, etc.) and function as follows:

1. nodes store the most recently received values from connected nodes (previously received values are overwritten);
2. when activated the values are summed;
3. the sum is sent to a look-up table which outputs a value;
4. the value is delayed, as determined by the corresponding node in the ED module;
5. the value is then sent to connected nodes, as well as back to the originating node.

Rules

When activated, a node sums the last values it received and sends the sum to a look-up table (LUT) that returns the value stored at the corresponding index. NetWorks has been designed to allow (1) each node to have its own LUT, (2) an LUT for each module, and (3) one LUT for all the nodes of the network. LUTs are generated using a variety of methods: random, random without repetition, ratios, etc.

Relationship between Architecture and Rules

Two observations can be made regarding the relationship of rules and network architecture. First, nodes can be thought of as “funnels,” in that the range of integers they can output is always less than the range of possible sums of their inputs. For example, if nodes can send and receive 12 values (1–13), the largest hub with 40 inputs requires an LUT with an index of 520, but can only output 13 different values, which results in a loss of “information.”

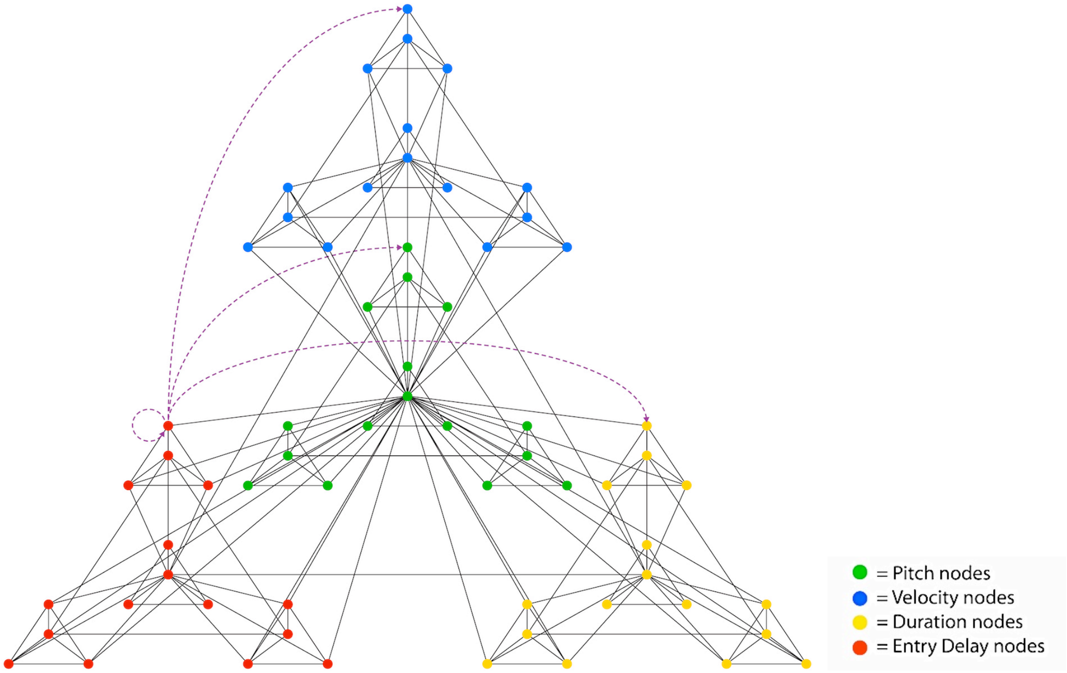


Figure 1. Schematic illustration of the different kinds of nodes and their interrelationships. Undirected edges (in black) indicate that values can be exchanged in both directions, i.e., nodes both send values to, and receive values from, nodes to which they are connected. Directed edges (purple) show the relationship between individual nodes of the Entry Delay module and the corresponding nodes of other modules. The ED module node determines when it will activate itself, and the corresponding node in the duration, velocity, and pitch modules. For clarity, only one of the 16 ED nodes and its four corresponding nodes are shown.

Second, while hubs have a wider "sphere of influence" because their output is sent to a greater number of nodes, hubs also receive input from the same nodes (which co-determine their output). Consequently, information flows both from the bottom-up and from the top-down. However, the more connected the hub, the more inputs it sums, and the less able it is to respond with unique outputs. While less well-connected nodes have a smaller "sphere of influence," their ability to respond to their inputs with unique outputs is significantly greater. Put another way, information flowing from the top-down is more "general" than information flowing from the bottom-up.

MIDI Mapping

The MIDI Mapping layer allows users to map node output values to appropriate MIDI ranges. For example, if nodes are set to output 12 values:

1. output values from pitch nodes can be mapped to a chromatic scale (e.g. C4–C5);
2. velocity node outputs can be mapped to 10, 20, 30, 40 … 120 MIDI velocity values;
3. duration node outputs can be mapped to an arbitrarily chosen fixed range (e.g., 100, 150, 200 … 650 milliseconds) or a duration based on a subdivision of the entry delay times between notes.
4. entry delays values between notes are scaled to an appropriate musical range in milliseconds.

In addition to generating the basic attributes of notes, NW provides for mapping network activity to MIDI continuous controller commands to control various synthesis parameters such as filters, and so forth, chosen the user. Currently, however, these outputs do not feed back into the network.

Since the MIDI data is computer-generated, sampled acoustic instruments are often used to give the music a "human feel," and to help the listener compare the self-organizing output patterns to known genres. When mapping patterns to sound, and during mixing, the main goal is to preserve the integrity of the generated patterns by ensuring that the changing relationships between note attributes remain audible. Instruments with complex envelopes and textures, and effects (such as echo), were avoided.

Evaluation of NetWorks Output

The dynamics of NW is controlled primarily by the choice of LUTs. Networks with nodes using LUTs with a random distribution of output values result in chaotic MIDI data sequences (and thus no element of predictability). Networks with nodes using LUTs that output the same value produce total repetition. The musicality of the output is greatest when the system is tuned to an intermediate between these extremes, i.e., the proverbial 'edge of chaos.' At this point there is a pleasing balance between familiar, repeating patterns, and novelty.

Shannon Entropy was used to (1) compare MIDI data sequences generated with rules having a random distribution of output values with MIDI data generated using LUTs that output (mostly) the same value when activated, and (2) to compare NW pieces, tuned to the edge of chaos, to other musical genres to confirm subjective comparisons.

Entropy values capture the degree of variety and repetitiveness of note sequences in MIDI data. Roughly speaking, high entropy indicates surprising or unpredictable musical patterns while low entropy indicates predictable, repeating musical patterns (Ren 2015). Differences in entropy values stem from differences of (1) the underlying possibility space size, i.e., how many different types of musical events there are, and (2) how repetitive they are. Although this does not take into account the order of events it provides a general characterization useful for comparing musical sequences (Ren 2015).

In this analysis, the entropy of a piece was calculated by counting the frequency of musical events, specifically the appearances of each note (pitch-duration pair), as well as pitch and duration separately to get the discrete distribution of those events. The information content of each note was calculated as per Equation 1.

$$H(X) = -\sum_i p(x_i) logp(x_i), i \in n = outcomes \quad (1)$$

The expectation value of the information content, defined as $-logp(x_i)$, was used to obtain the entropy. The entropy is related to the frequency of musical events in a specific range. In Figure 2, the entropy value of ten NW pieces (x-tick=3) is compared with Bach's chorales (x-tick=1) and with jazz tunes (x-tick=2). In terms of entropy, NW pieces are closer to jazz than to Bach, which confirms informal subjective evaluations of NW music. X-tick=4 shows the entropy value for three NW pieces generated using a random distribution of LUT output values and x-tick=5 shows the entropy values of three NWs pieces with near uniform LUTs. These values verify the relationship between NW MIDI outputs and the LUTs that generate them. To date, two albums have been produced using the NetWorks system: "NetWorks 1: Could-be Music" and "NetWorks 2: Phase Portraits," which can be heard online:

- https://shawnbell.bandcamp.com/album/networks-1-could-be-music
- https://shawnbell.bandcamp.com/album/networks-2-phase-portraits

The most recent experiments can be found here:

- https://soundcloud.com/zomes

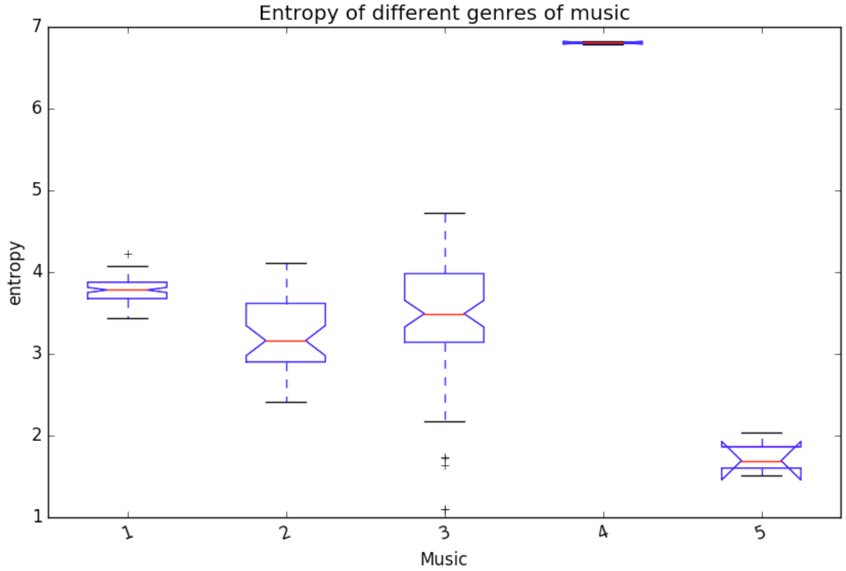


Figure 2. Comparison of entropy of ten NW pieces (x-tick=3) with Bach chorales (x-tick=1) and jazz tunes (x-tick=2).

Evaluation of NW music via social media (SoundCloud), shows an increasing interest in NW music from what is quite likely a diverse audience given the wide range of social media groups to which NW music has been posted (e.g., classical, jazz, electronic, experimental, ambient, film music, algorithmic music, creative coding, complex systems, etc.). There has been a steady growth of "followers" over the two years (2014–2016) of posting NW pieces (28 tracks). As of the writing of this paper, NW has 328 followers. 8,128 listens, 834 downloads, 368 likes, 24 reposts, and 56 comments (all of which are positive). As a search for "music-as-it-could-be," (i.e., new genres) a comment from SoundCloud indicates this goal may have been attained: "What can I say except I think I like it?" This suggests that the person has heard something they cannot categorize, but that sounds like good music.

How NetWorks Implements Honing Theory

We now summarize how the architecture and outputs adhere to and implement ideas from honing theory (HT), a theory of creativity inspired by chaos theory and the theory of complex adaptive systems (CAS).

NW as Creative, Self-Organizing Structure

NW is hardwired to exhibit the key properties of real-world complex systems through its modular, scale-free, small-world properties. NW architecture has a shallow, fractal, self-similar structure (4 node, 16 node, and 64 node modules) that allows multiple basins of attraction to form in parallel, over different timescales, and interact. NW networks are not neural networks; they do not adapt or learn by tuning weights between nodes through experience or training. Nor do they evolve; nodes simply accept input and respond. Their LUTs do not change, adapt, or self-organize over time, but their dynamics do.

Just like an experience or realization can provide the 'seed incident' that stimulates creative honing, the pseudo-randomly generated initial conditions provide 'seed incidents' that initiate processing. After NW receives its inputs it is a closed system that adapts to itself (self-organizes). Musical ideas sometimes unfold in an open-ended manner, producing novelty and surprise, both considered hallmarks of emergence. A diversity of asynchronous interactions (sometimes spread out in time) can push NW dynamics across different basins of attraction. Idea refinement occurs when users (1) generate and evaluate network architectures, LUTs and mappings, and (2) orchestrate, mix, and master the most aesthetically pleasing instances of these outputs. The role of mental representation is played by notes—their basic attributes as well as attributes formed by their relationships to other notes.

Cellular Automata-like Behavioral Classes

NW nodes have a significantly different topology from Cellular Automata (CA). While CA have a regular lattice geometry, NW has a hierarchical (modular), scale-free, small-world structure. Moreover, unlike CAs, NW is updated asynchronously. However, similar to CA, NW exhibits Wolfram's class one (homogenous), class two (periodic), class three (chaotic), and class four (complex) behaviour (Wolfram 1984), and—rather than converging to a steady state—tends to oscillate between them. This is because the nested NW architecture allow multiple basins of attraction to form in parallel and over different timescales. Pruning the scale-free architecture by reducing the inputs to hubs insulates clusters and modules from one another, thereby reducing interactions. Network dynamics within a basin of attraction can get pushed out of the basin by delayed values entering the system. In other words, because in the context of the current pattern an "old idea" can push the dynamics to a different basin, the system exhibits "self-mending" behavior. This can result in musical transitions that lead to the emergence of new patterns and textures.

Representational Redescription

The network "makes sense" of its present in terms of its past by adapting to delayed values or "old ideas" entering the current pattern of activations. Nodes hone by integrating and simplifying inputs from multiple sources, and returning a particular value. A catalyst or "catalytic value" is one that needs to be received on the inputs of one or more nodes to maintain one or more periodic structure (perhaps playing a different role in each). As NW strings notes together (often in parallel) in a stream of music, its nodes act on and react to both the nodes in their cluster, and to other clusters, via their hubs. Bottom-up and top-down feedback and time-delayed interactions are essential for an open-ended communal evolution of creative novelty.

Periodic structures are often disrupted (stopped or modified) by the introduction of a new (delayed) value, although sometimes this does not affect output. As interactions between nodes occur through entry delays, periodic musical structures unfold at different timescales. Slowly evolving periodic structures can be difficult to hear (due to intervening events) but can have a "guiding" effect on the output stream, i.e., they affect what Bimbot, Deruty, Sargent, and Vincent (2011) refer to as the "semiotic" or high-level structure of the music emerging from long term regularities and relationships between its successive parts. NW creates musical "ideas" that become the context for further unfolding. Asynchrony, achieved by the (dynamically changing) values of the nodes in the Entry Delay Module, allow previously calculated node values (including their own) to be outputted later in time. Outputs both *manifest* the dynamics of the network, and *generate* subsequent dynamics. As with the autopoietic structure of a creative mind, NW is a complex system composed of mutually interdependent parts.

Let us examine how a NW network could be said to take on characteristics of an autocatalytically closed creative mind. The nodes collectively act as a memory in the following sense. When a node is activated, it sums the last values received on its inputs and uses the sum to output the stored value (which is then delayed before being sent to receiving nodes). Nodes are programmed so that their individual inputs can only store or "remember" the last value received. However, because nodes have 3, 4, 5, 14 and 39 inputs (excluding their own), and the network is asynchronous, a node (as a whole) can "remember" values spread out over time. How long a node can remember depends on its own ED value and the ED values of the nodes that participate in co-determining its output. It is important to note, however, that nodes can also "forget" much of the information they receive, if, for example, it receives different values on the same inputs since only the last ones are used when the node is activated. Again, how much it forgets depends on its own ED value and those of nodes to which it is linked. These memory patterns are distributed across the network. They are self-organizing because they can recur with variation, such that the whole is constantly revising itself. NW chains items together into a stream of related notes / note attributes. As NW strings notes together in a stream of music, its nodes are acting on and reacting to (feeding-back and feeding-forward information) to and from both the nodes in their cluster and to other clusters via their hubs. It would seem that bottom-up, top-down and time-based interaction / feedback are essential for an open-ended communal evolution of creative novelty.

Contextual Focus and the Edge of Chaos

Some of NW's music sounds uninspired, i.e., does not contain surprising pattern development (e.g., a sudden transition or gradually modulated transition in texture, mood, or tempo), and/or the patterns do not elicit innovative variations. To minimize this problem, NW uses an architecture that, in its own way, implements contextual focus. Clusters of nodes that are more interlinked and share similar LUTs process in a more analytic mode. Hubs, which connect clusters into a small-world network and merge more distantly related musical ideas, process in a more associative mode. Because clusters have fewer inputs than hubs they are more discriminating than hubs. Hubs act as funnels, summarizing or simplifying the information they receive from multiple sources. Thus NW is hardwired to shift between analytic and associative modes by modulating the relative influence of top-down and bottom-up processing.

NW structures transform as they propagate in time. As mentioned above, all four behavior classes have been observed. Class one and two dynamics do not change unless disrupted. When NW processes 'associatively' the output streams exhibit class two behaviour. When NW processes 'analytically' it exhibits Class three (deterministic chaos) behavior, which does not repeat if unbounded. Class four (edge of chaos) balances change and continuity.

Network dynamics often sound chaotic at the beginning of a piece–set in motion from an arbitrary, initial configuration ('seed incident'). Repetition and development of motivic materials and/or melodic lines then moves the system toward one or more attractor(s) (or "grooves"), resulting in a more stable, organized musical texture. Nodes with different rules of interaction are apt to disturb the system, pushing it into another basin. If it returns to a basin, a similar texture returns. When tuned to midway between order and chaos the global stable dynamics are repeatedly disturbed. This pushes it either (1) into another basin, creating a transition to contrasting musical material, or (2) further from the attractor, to which it is wont to return. NW exhibits something akin to goal seeking behaviour in how it moves toward or away from an attractor by keeping within a range of "desirable" values. This is similar to the use of functional tonality in western music, where a piece departs and returns to its tonal center. Quasi-periodic dynamics provide organization through cycling musical textures, or a loose theme and variation structure. Disturbances may be caused by nodes with different rules of interaction, or by delayed values entering the stream. One factor that affects the aesthetic quality of the output is the mapping of the node output values to a specific ED scale (mapped values are used to delay node outputs). This appears to produce a balance between current events and older ones that is at the proverbial edge of chaos.

Conclusions and Future Directions

NW's unique architecture—in particular, its scale-free network and transparent relationship between rules of interaction (LUTs) and MIDI output—was inspired by the science of complex adaptive systems as advocated by the honing theory of creativity. Its dynamics can be tuned midway between order and chaos, and evolve—not through a Darwinian process (c.f., Gabora, 2005; Gabora & Kauffman, 2016)—but in the sense of generating cumulative, adaptive (in this case, aesthetically pleasing) change.

The number of possible LUTs that can generate 'edge of chaos' dynamics is extremely large. In the future we will expand the scope of NW to get a sense of to what extent the agreeable sound palette contributes to the aesthetic assessment. "By hand" rule design and "by ear" verification of the results will be augmented by evolutionary programming techniques guided by quantitative analyses. NW will also continue incorporating principles of HT. In turn, grounding the theory using NW is inspiring new developments in the understanding of creativity.

Acknowledgements

Many thanks to Iris Yuping Ren for the entropy analysis. This research was supported in part by a grant (62R06523) from the Natural Sciences and Engineering Research Council of Canada.

References

Abraham, F. D. 1996. The dynamics of creativity and the courage to be. In W. Sulis & A. Combs (Eds.), Nonlinear dynamics in human behavior. River Edge, NJ: World Scientific, 364–400.

Ashby, W. 1947. Principles of the self-organizing dynamic system. Journal of General Psychology 37:125–128.

Bak, P., Tang, C., and Weisenfeld, K. 1988. Self-organized criticality. Physical Review A 38:364.

Barábasi, A. L. 2002. Linked: The New Science of Networks. New York: Perseus.

Bell, S. 1998. Emergent Music I: Voices From The Edge of Chaos. Polycarpe Studios: EM001.

Beyls, P. 1989. The musical universe of cellular automata. In Wells, T. & Butler, D., eds., Proceedings of the International Computer Music Conference, 34–41.

Beyls, P. 1990. Musical morphologies from self-organizing systems. Interface: J. New Music Research 19(2–3): 205–218.

Beyls, P. 1991. Self-organizing control structures using multiple cellular automata. Proc Intnl Computer Music Conf. Montreal: Intnl Computer Music Assoc. 254–257.

Bimbot, F.; Deruty, E.; Sargent, G.; and Vincent, E. 2011. Methodology and resources for a structural segmentation of music pieces into autonomous and comparable blocks. Proc Intl Soc Music Information Retrieval Conf 287–292.

Bowcott, P. 1989. Cellular Automata as a means of high level control of granular synthesis. In Wells, T. & Butler, D., eds., Proc Intl Comp Music Conf. San Francisco: ICMA. 55–57.

Carver, C., and Scheier, M. 2002. Control processes and self organization as complementary principles underlying behavior. Personality and Social Psych Review 6:304–315.

Chareyron, J. 1990. Digital synthesis of self-modifying waveforms by means of linear automata. Comp Music J, 14(4):25–41.

Evans J. St. B. 2003. In two minds: dual process accounts of reasoning. Trends in Cognitive Science 7:454–59.

Gabora, L. 1998. Autocatalytic closure in a cognitive system: A tentative scenario for the origin of culture. Psycoloquy, 9(67).

Gabora, L. 2005. Creative thought as a non-Darwinian evolutionary process. Journal of Creative Behavior 39:262–283.

Gabora, L. 2000. Toward a theory of creative inklings. In (R. Ascott, ed.) Art, technology, and consciousness (pp. 159–164).

Gabora, L. 2010. Revenge of the 'neurds': Characterizing creative thought in terms of the structure and dynamics of human memory. Creativity Research Journal 22:1–13.

Gabora, L. & Kauffman, S. (2016). Toward an evolutionary-predictive foundation for creativity. Psychonomic Bulletin & Review 23:632–639.

Gabora, L. (in press). Honing theory: A complex systems framework for creativity. Nonlinear Dynamics, Psych, and Life Sci.

Goertzel, B. 1997. From complexity to creativity. New York: Plenum Press.

Guastello, S. 1998. Creative problem solving at the edge of chaos. J Creative Behavior 32:38–57.

Horner, A. & Goldberg, D. E. 1991. Computer-Assisted composition. Fourth International Conference on Genetic Algorithms. University of California, San Diego, CA.

Horowitz, D. 1994. Generating rhythms with genetic algorithms. Proc Intnl Computer Music Conference. Aarhus, Denmark.

Kitzbichler, M. G., Smith, M. L., Christensen, S. R., and Bullmore, E. 2009. Broadband criticality of human brain network synchronization. PLoS Computational Biology, 5.

Liu, X. F, Tse, C. K., and Small, M. 2010. Complex network structure of musical compositions: Algorithmic generation of appealing music. Physica A 38:126–132.

Millen, D. 1992. Generation of formal patterns for music composition by means of cellular automata. In Strange, A., ed., Proc Intl Comp Music Conf. San Francisco: ICMA, 398–399.

Miranda, E. R. 1995. Granular synthesis of sounds by means of a cellular automaton. Leonardo 28(4):297–300.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. 2004. The University of South Florida free association, rhyme, and word fragment norms. Beh Res Meth, Instr, & Comput 36(3):402–407.

Pribram K.H. 1994. Origins: brain and self-organization. Hillsdale NJ: Lawrence Erlbaum.

Prigogine, I., & Nicolis, G. 1977. Self-organization in non-equilibrium systems. New York: Wiley.

Ren, I Y. 2014. Complexity of Musical Patterns. Univ Warwick.

Ren, I Y. 2015. Using Shannon Entropy to evaluate automatic music generation systems: A Case Study of Bach's Chorales. ECE Department, University of Rochester.

Runco, M. 2010. Divergent thinking, creativity, and ideation. In J. Kaufman & R. Sternberg, (Eds.), The Cambridge handbook of creativity. Cambridge UK: Cambridge Univ Press, 414–446.

Sowden, P., Pringle, A., and Gabora, L. 2015. The shifting sands of creative thinking: Connections to dual process theory. Thinking & Reasoning 21:40–60.

Steyvers, M., and Tenenbaum, J. B. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive Science 29:4–78.

Todd, P. M. and Loy, G., eds. 1991. Music and Connectionism. MIT Press.

Wolfram, S. 1984. Universality and Complexity in Cellular Automata. Physica D 10:1–35.

A CSound and HTML5 Based Cross Platform Audio Development System

Edward Costello
The Sound and Digital Music Technology Group, Maynooth University, Ireland

Dr. Victor Lazzarini
The Sound and Digital Music Technology Group, Maynooth University, Ireland

Software based musical instruments and effects are now a ubiquitous and essential component for the composition and production of electronic music. These components may take the form of standalone applications or more commonly as plugins inside of host sequencers and audio editors.

As there exists a number of competing standards, audio plugins are commonly created using software frameworks such as WDL-OL[1]. These frameworks allow developers to focus on writing just the plugin audio DSP and user interface code which can then be easily compiled to a number of plugin standards. Having the ability to use the same software instrument or effect inside multiple audio applications allows musicians and audio engineers to easily work within multiple composition environments and platforms.

With the web emerging as a viable platform for audio applications, the ability to use individual plugin instruments and effects on this platform as well as natively would be extremely beneficial. It would enable an easier interoperability between the web and native platforms helping the web grow as a serious environment for audio production. Unfortunately there are no established audio plugin development frameworks that allow for the development of plugins which can be used across both native plugin hosts and the web. For this reason we have developed a framework based on HTML5 and the Csound language. This allows the creation of robust, cross platform audio components using a dedicated audio DSP language and a ubiquitous and well supported user interface platform.

Within this framework instruments and effects are defined entirely using the Csound language and HTML5. The audio DSP for each plugin is written using the Csound language and the user interface is created using a combination of HTML, CSS and Javascript. Other relevant information for each audio component such as adjustable parameters and presets are also defined within an additional JSON file. These assets are used by a custom software wrapper for each target platform which provides access to the Csound API and a web view for the plugin user interface. At present we have a working implementation for Apple's AudioUnit plugin framework and the web, but additional plugin wrappers for the VST and LV2 standards are also envisioned.

---

[1] https://github.com/olilarkin/wdl-ol

Creating The Illusion Of Self-Motion Through Sound

Jordan Higgins
University of Limerick

The aim of this project is to create the illusion of self-motion through the use of 3D binaural audio. Binaural audio relates to how sound is received in both ears. For this project I researched and investigated the way in which humans perceive audio in their surroundings in terms of location, distance, elevation and factors relating to motion, either of the sound source or the observer.

Once all these factors were taken into consideration I then created a 3d soundscape, utilizing these various cues and effects in order to elicit a sense of self-motion from the listener. These illusions of self-motion have so far been accomplished at a higher success rate through the use of visual stimuli, but haven't yet been explored as thoroughly through auditory means. Virtual reality has become more and more popular in recent years and has become more advanced in terms of 3d technology. It is also becoming a larger part of existing media such as 3d films and the oculus rift in gaming. There are also some 3d developments in the music industry. This project is dealing with a component of 3d virtual reality that hasn't had much exploration yet but plays a large part in immersing users into virtual environments. That is the illusion of self-motion and is one of the biggest challenges in 3d technology. In order to create this sense of self-motion I first needed to examine how we perceive sound. To do this I examined how we perceive a sounds location, elevation and cues such as head motion and sound source motion cues like the Doppler effect. Some other factors that needed to be considered are the types of sounds; will they be stationary or not, how fast will the sounds be moving if at all and any pitch changes that I'll needed to consider a result. I also needed to decide what direction the observer should be moving and also what environment they would be moving through. These considerations are important as various factors can evoke stronger senses of self-motion than others. When testing, observers are blindfolded before being lead into a room and seated. The audio is then presented on a pair of high quality headphones. Participants control their movement through the soundscape using a keyboard. By considering spatialisation cues to help create a realistic auditory environment and allowing users to control their movement through it should help elicit an illusion of self-motion.

Unlocking the Power of Conceptual Metaphor and Image Schematic Constructs in Designing An Alternate Paradigm for Loop-Based, Rhythmic Electronic Performance

J.E. King
Dundalk Institute of Technology (DkIT)
`John.King1@dkit.ie, Johnking2010@me.com`

Abstract This paper interrogates the state of the art with regard to the rhythmic interface. I discuss how aesthetic technological determinism may be seen to affect artistic practice, and present different viewpoints on reconceptualising the rhythmic interface, beyond prevalent paradigms. Contrary to Stockhausen's assertion that Daniel Pemberton (for one) should "give up this loop", as "it is too old fashioned" (Cox and Warner, 2004, p.382); I take the stance that the "loop" is fundamental to much of contemporary rhythmic-electronic music. A musical aesthetic, that embraced mechanical, machine rhythm, developed in the wake of affordable (sampling) drum machines and sequencers having been first brought to market. Many of these early electronic instruments have become paradigmatic in shaping how we conceive of electronic rhythm, and in how we interact with it. Step sequences, patterns, arrays of pads and/or buttons are common and interchangeable features within the majority of our contemporary rhythmic tools. Core to most, if not all of these tools, is the dominance of the grid in our conceptualisation of how we perform and arrange material.

The design of my own system (Axis) is based on an adaptable conceptual model for loop-based performance, closely informed by the demands of contemporary creative practice. I aim to exploit the discernible features of looped audio through considered gestures afforded by a multi-touch tablet interface. In proposing a rethink of the rhythmic interface, I suggest this system as one alternative example that facilitates meaningful, multi-parametric control over the relationships between the actions, processes and structures of rhythmic electronic music. My design framework has been guided by Lakoff and Johnson's pervasive theories of conceptual metaphor and image schema, and by experientialist perspectives of embodied cognitive perception. I examine the overlapping areas of HCI and electronic/computer music research in considering the relevance of affordances and metaphor in design, with specific regard to the creation of technologically mediated musical performance systems. Finally, this work explores the suggestion of Hurtienne et al. (2010, p.483) that the assimilation of primary metaphor and image-schematic constructs in designing our interactions with technology may foster intuitive understanding of the user interface; in addition to assisting the execution of more complex and multi-dimensional tasks, such as the creation and manipulation of electronic rhythm.

1. The State Of The Art Of The Rhythmic Interface

American composer Henry Cowell's 1932 collaboration with the renowned inventor, Leon Theremin in developing the 'Rhythmicon', served as a seminal event within the history of rhythm-based electronic music. Leland Smith (1973) perhaps erroneously[1], recognises it as the first electronic instrument for generating rhythmic material. Smith cites Cowell, commenting on his part in the design process:

---

[1] Smirnov (2013, p.66) indicates that I. Sergeev had earlier patented and developed the *Electricheskly mizikaini pribor* or *Electrical musical device* in 1925; an instrument with similar capabilities to the *Rhythmicon.*

> "My part in its invention was to invent the idea that such a rhythmic instrument was a necessity to further rhythmic development, which had more or less reached the limit of performance by hand, and needed the application of mechanical aid. That which the instrument was to accomplish, what rhythms it should do, and the pitch it should have, and the relation between the pitch and rhythm, are my ideas."
> (Cowell, 1932. In Smith, 1973)

Of a similar era, Conlon Nancarrow's intricate and complicated works for player-piano also pursued this notion of eliciting rhythmic patterns beyond human performance capability (Gann, 1995). The Rhythmicon stands as the result of developing a solution to a compositional problem conceived of by Cowell. It exists as an early example of creative collaboration between designer and performer/composer. Cowell can be seen to put the emphasis on human agency in performing rhythmic music with the instrument of his and Theremin's design. Nancarrow's compositions deftly exploit the inherent potential of the player piano's mechanisms but stand in contrast to Cowell's work in the respect that ultimately, the intent was for the *instrument* to play the (fixed-form) composition. The human interaction ends with the generation of a completed score. The piano-roll editing metaphor prevalent in today's Digital Audio Workstation (DAW) software packages conceptually reflects the compositional paradigm that Nancarrow adopted[2].

Grid, pad and step-sequence-based interfaces have long been prevalent paradigms within rhythmic electronic music. Products brought to market in the early 1980s by companies such as Roland and Akai pioneered these approaches and set the tone for future developments. Roland's TR-606, TR-808, and TR-909 drum-machines have had a deep and long-lasting impact on rhythmic electronic music. Despite their dissemblance from authentic, real drum-sounds**,** the tonal qualities and sonic signatures of Roland TR-808 and TR-909 drum-machines have become synonymous with a range of electronic music genres. Furthermore, the modes of operation engendered by these devices (step-sequencing) became *de facto* standards, the legacy of which continues to shape conceptions regarding composition and production techniques. Akai's MPC series of sampling drum-machines/sequencers offered a different perspective with regard to rhythmic production[3]. Instead of generating electronic percussion sounds, these devices facilitated a sample-based approach, enabling users to construct rhythms from raw audio materials.

2. How Aesthetic Technological Determinism May Affect Artistic Practice

The widespread availability of computer music systems in recent years has led to a surge of activity in the design of new musical instruments and interfaces. Nonetheless, there remain voices within the computer/electronic music communities who are concerned as to how to sensibly improve musical interface design within an era of such rapid technological advancement (Machover, 2002; Chadabe, 2000). Musical intent and aesthetic sensibility may influence different approaches to hardware and software setups. Conversely, different hardware and software choices may also impact upon our conceptions of musical processes and structures. The standard Digital Audio Workstation (DAW), however, with its associated multi-track mixing metaphor and piano-roll sequencing methods may be accused of engendering conceptual limitations and, hence, creative

---

[2] Nancarrow's work pushed at the boundaries of the player-piano's capabilities, foreshadowing similar activities by contemporary Japanese artists' exploring the musical area of *Black MIDI* http://impossible-music.wikia.com/wiki/Blackened_musical_notation

[3] Said (2013, p.70-71) notes the inspiration for this approach originated with US DJ and producer Marley Marl'*s* core innovations: sampling drums (or any sound) directly from vinyl and the subsequent *chopping* and *layering* of these samples These techniques significantly expanded the sonic palette of his (and subsequently others') musical works, acting as a creative template for sampling drum-machine use.

limitations.

In terms of the design and functionality of electronic instruments, it should also be stressed that the way in which artists actually use many modern electronic instruments, may be radically different from that intended by their original designers. Roger Linn (instrument designer and creator of the LM-1 drum machine and Akai MPC-60 sampling drum machine) laments the production of overly "mechanical" rhythms from his machines, as he had spent so much time in implementing features to make them expressive tools capable of subtle nuance (*From Berkeley to the Bronx: Roger Linn's Drum Machines,* 2007). Dannenberg (2007) notes the "stiff, mechanical and monotonous" qualities imparted onto various forms of music as a result of new (largely commercial) music technologies (samplers, sequencers and drum machines[4]). Magnusson (2009) posits that the music technologies themselves ultimately determine how they will be used[5], whereas Rowe suggests, "many interactive systems can be considered applied music theories" (Rowe, 2001, p.6). I contest that the adherence to the piano-roll metaphor, and other two-dimensional representations of rhythm[6] persistently dominates much of our thinking within this musical area, but provides inadequate access to relevant rhythmic structures. The use of MIDI-based physical control interfaces and input devices, in addition to (both software and hardware) synthesisers, samplers and sequencers remains widespread. Furthermore, the dominance of MIDI-based hardware may further restrict rhythmic expression. Jaron Lanier (2010) reminds us Dave Smith[7], was a keyboard player and that this ultimately influenced his conception of how to integrate his instrument with other technology. Commenting on the discrete, binary nature of MIDI events, Lanier eloquently phrases that MIDI: "...could only describe the tile mosaic world of the keyboardist, not the watercolour world of the violin." (Lanier, 2010, p.4). Lanier further argues that MIDI is guilty of severely restricting our conceptions of music[8] to the point that: "the whole of the human auditory experience has become filled with discrete notes that fit in a grid." (Lanier, 2010, p.5).

3. The Loop in Electronic Music

The loop has remained an established affordance within electronic music, having initially served as a focusing device for timbral exposition within musique concrète (via Schaeffer's locked–groove experiments) prior to the subsequent developments of the sequencer and digital sampling. The primacy of the loop within (electronic) music may be because it is not necessarily always normative, sometimes providing a recognisable structural prototype evident more in its divergence/subversion rather than conformity. This deviance from conformity is of particular relevance as rhythmic electronic music continues to explore extended textural and sonic potentials. From an embodied cognitive perspective, as the structure and affordance of the loop becomes embedded in practice, it may also be seen as providing the basis for thought: the key musical element is a cycle rather than a single note/sound or a linear progression. Treating the loop as the central element, not only of localised musical structure, but also as an extensible conceptual model,

---

[4] Typified by such devices as Roland TR-808 and TR-909 drum machines and Akai's MPC series of sampler-sequencers.

[5] Magnusson speaks of "technologies inherent with ideological content" as "...never neutral, they contain scripts that we subscribe to or reject according to our ideological constitution" (Magnusson, 2009, p. 170). Andy Clark provides the foundation for Magnusson's claim, his theory of the *Extended Mind* suggests that "objects and artefacts serve as an external playground for thinking" (Clark in Magnusson, 2009, p170). This idea further resonates with Norman's concept of our using "Knowledge in the world" (Norman, 1988) as a mechanism to reduce our cognitive load.

[6] Trevor Wishart (1996) describes these as "two-dimensional lattice" structures.

[7] The inventor of MIDI.

[8] This sentiment was expressed nearly a century earlier in Russia by Arseny Avraamov, who articulated his extreme desire for "a project to burn all pianos – symbols of the despised twelve-tone, octave based, 'well tempered' scale" (Smirnov, A. 2013. p.31). A firm advocate of microtonal music, Avraamov was of the belief that human hearing had been altered to some degree by the canon of music reliant upon the twelve-tone, octave based system (*ibid*.).

is a key tenet of developing a new performance interface. Representations derived from the use of cyclic musical motifs are not in themselves new ideas. London (2012) cites Anku (2000, in London, 2012), referencing his approach to notating West African rhythmic patterns via circular diagrams. Sethares (2007) presents further varied examples of cyclic representations facilitating the transcription and display of rhythmic patterns via *lace notation*. McLachlan (2000) also notes the cyclical mapping of rhythm as commonplace within African and Indonesian musical structures, indicating that this approach also adheres to gestalt psychological principles of good continuation and that of simplified representation.

4. How To Reconceptualise The Rhythmic Interface

I propose a shift in how we, as a community of musicians, technologists, performers and designers think about approaching rhythmic looped material. The key motivations for this work are to engender easier and more intuitive access to the processes and structures necessary for creating such music, whilst simultaneously facilitating highly expressive gestural control over multiple inter-related rhythmic elements. In offering interesting new functionality that exceeds the performative capabilities of the typical rhythmic electronic interface, my research questions the dominance of many archetypes of rhythmic electronic tool. It further attempts to overcome the limitations of simplistic, linear, track-based approaches to loop-based material, enabling access to all relevant control parameters through an integrated interface design. In developing such a system, my work offers the potential for rich engagement with affordance-laden structures provided through the approach taken towards interface design. This rhythmic electronic interface strives to provide congruence between the control structure of the device, and the perceptual nature of the musical task intended, as suggested by Jacob et al.:

> "Current input device taxonomies and other frameworks typically emphasise the mechanical structure of input devices. We suggest that selecting an appropriate input device for an interactive task requires looking beyond the physical structure of devices to the deeper perceptual structure of the task, the device, and the interrelationship between the perceptual structure of the task and the control properties of the device."
> (Jacob et al., 1994, p.3)

5. Design Framework

5.1 Affordances and Usability

The theoretical basis for the system proposed by this research is partly inspired by the work of Donald A. Norman, and in particular by his advocacy of usability as a primary concern for designers. His work has made a deep impact upon the overall field of design, remaining influential in the area of HCI, and also in the development of digital musical instruments (DMIs)[9].

---

[9] Atau Tanaka (2010) provides an account detailing the impact of the theory of affordances within the computer music community, pointing to work conducted by a number of researchers from related disciplines including: Magnusson (2006), Gurevich (2006), Cook and Pullin (2007), Godøy (2009), Brassch (2009), and Dillon and Brown (2010), (cited in Tanaka, 2010, p.2).

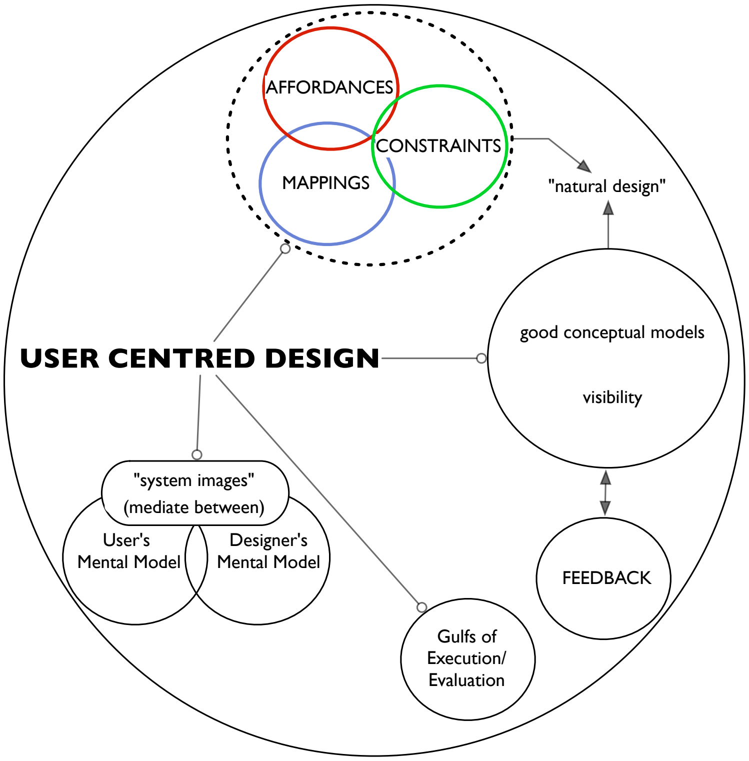
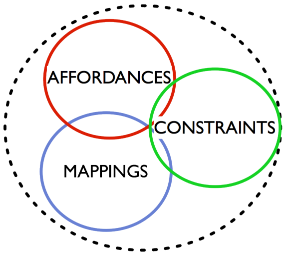


Figure 1 – Overview of core components within Norman's User Centred Design.

Norman maintains that the strong inter-relationships between *affordances*, *constraints*, and *mappings* contribute toward our forming conceptual models of devices/objects in the world around us.



"Affordances provide strong clues as to the operation of things.
Plates are for pushing.
Knobs are for turning.
Slots are for inserting things into.
Balls are for throwing or bouncing.
When affordances are taken advantage of, the user knows what to do just by looking..."
(Norman, 1988, p.9)

Figure 2 – Affordances, constraints and mappings diagram

Regarding the current use of the term affordances, a number of different perspectives are examined in depth by Greneo (1994), and by McGrenere and Ho (2000). The theory of affordances initially arose from James J. Gibson's research in the field of perception, with emphasis on the visual, which he initially explained as:

> "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill." (Gibson, 1986/1979. p.127)

Affordances can be said to exist as the product of the bi-directional relationship between a particular animal and its environment: they "point both ways" (Gibson, 1979/1986, p.129). Greneo (1994) succinctly comments on the dualistic aspect of affordances, stating that:

> "...people and animals are attuned to variables and invariants of information in their activities as they interact as participants with other systems in the world that we inhabit." (Greneo, 1994, p.337)

Gibson's perspective has been said to ignore the experience, knowledge, culture and perceptual abilities of the actor, focusing on their action capabilities; whereas in Norman's view, "the mental and perceptual capabilities of the actor" are central to our ability to recognise and make use of affordances. (McGrenere and Ho, 2000) In their analysis, McGrenere and Ho dismiss the (Gibsonian) binary view of affordances as being an oversimplification. I share the view of affordances as being "preconditions for activity" (Greneo, 1994, p.340), treating them as "conditions for constraints"(*ibid.*) that take into account the intended actions and motivations of an agent in a particular situation. In applying the metaphor: A DIAL IS A LOOP OF AUDIO to guide this design process, the conceptual blend of visual and sonic domains was reinforced by adherence to user-centred design principles, closely guided by Jakob Nielsen's usability heuristics (see Figure 3).

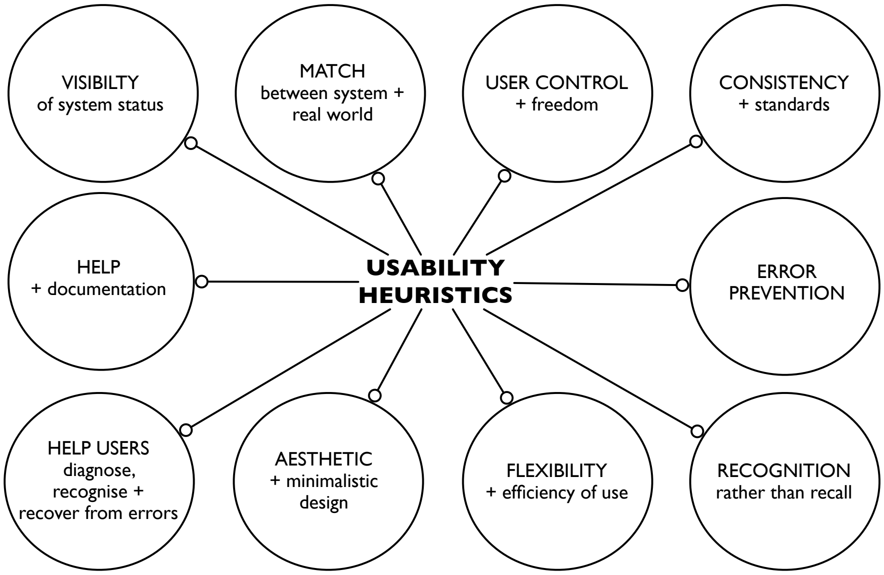


Figure 3 – Ten Usability Heuristics derived from Nielsen (1994).

Nielsen's heuristics are intended as guidelines for the user-centred design process, drawing influence from the work of Norman in attempting to construct and exploit natural mappings within an interface intended for the performance of rhythmic electronic music. Many of the mappings within the proposed interface design are themselves rooted in physical/spatial analogies and cultural standards, in efforts to engender "immediate understanding" (Norman, 1988. p.23). The idea that an object can impart knowledge to a user by its structure or appearance, in conjunction with the cultural and experiential knowledge of the user is not new[10], yet it remains a fascinating prospect for design to pursue, regarding digital technology.

5.2 Image Schemas and Metaphor

Core to my research is the ability of metaphor to facilitate the transfer/transformation of knowledge:

---

[10] Gestalt psychologist Koffka claimed "each thing says what it is" (Koffka in Gibson, 1986/1979. p.138), furthermore, that things "tell us what to do with them" (*ibid.*).

more specifically, how this may be employed to our advantage in designing and creating technologically-mediated musical tools. A number of authors (including Lakoff and Johnson, 1980, p.153; Modell, 2009, p.6; Gibbs, 2011, p.113) state that metaphor is a core feature of our ability to structure concepts, merely reflected in our use of language. Goatly (2007, p.14) further suggests that it facilitates abstract thought, and Modell (2009) highlights the huge potential of metaphor to both transfer and (more importantly) transform meaning between domains of our human experience. These perspectives support the theory that the human conceptual system is itself metaphorical in nature and based on our massive embodied experience of interacting with our physical and cultural environments (Lakoff and Johnson, 1980, p.3). Lakoff and Johnson's stark opening to *Philosophy in the Flesh* (1999) demonstrates one interpretation of the experientialist view of embodied cognitive psychology, opposing the classical objectivist perspective, reliant upon Cartesian notions of mind/body dualism:

> "The mind is inherently embodied. Thought is mostly unconscious. Abstract concepts are largely metaphorical." (Lakoff and Johnson, 1999, p.3)

This embodied perspective stresses that an organism and its environment are engaged in a two-way dialogue, what philosopher Mark Johnson calls "interactive coordination" (Johnson, 2007, p.136). Lakoff and Johnson's theory of image schema further explains cross-domain mapping as a function of conceptual metaphor, in assisting understanding between a variety of *target* and *source* domains. Johnson (2007) defines image schema as referring to:

> "these basic structures of sensori-motor experience by which we encounter a world that we can understand and act within. An image schema is a dynamic, recurring pattern of organism-environment interactions"
> (Johnson 2007, p.136)

Image schemas facilitate our mental representations of concepts via our particular (human) corporeal experience in the world enabling cross-modal perceptual understanding at a pre-conceptual, non-conscious level, binding body and mind (Zbikowski, 1997; Johnson, 2007, p.145).

5.3 Embodied Music Cognition

Johnson details the sensori-motor basis of our relationship with music, highlighting the connection between image-schematic structures and embodied musical meaning. He states that human understanding of the "logic of physical motion" (Johnson, 2007, p.57) (developed through lived bodily experience in the world) facilitates our experience and comprehension of motion within musical structures. Further research indicates that bodily image schemas involving force dynamics[11] and path structures are tightly bound to our processes of cognitively engaging with music (Brower, 2000, p.324; Johnson, 2007, p.57). Zbikowski indicates that images schemas act as the basis for conceptual models and that those fundamental to music arise from cross-domain mappings originating in other concrete domains of experience (Zbikowski, 1997, p.217). Candace Brower elaborates on this relationship, positing that musical meaning arises out of mapping heard patterns of music onto what she terms *intra-opus patterns*, *musical schemas* and *image schemas*; noting that only image- schematic pattern matching facilitates cross-domain or metaphorical mappings

---

[11] Neurolinguistic psychotherapists Tompkins and Lawley (2009) discuss the idea of superimposition of multiple image schemas (compositional blending), noting that schemas related to force, often operate in what they term *clusters*, engaging different sensory modalities.

(Brower, 2000, p.324). Brower states that it is via these cross-domain mappings, that
“the everyday metaphors of tonal music can be represented in the form of music-metaphorical schemas.” (Brower, 2000, p.325). She identifies the primary embodied image schemas that contribute to musical understanding as those of: VERTICALITY (UP-DOWN), CONTAINER, SOURCE-PATH-GOAL, BALANCE, CENTRE-PERIPHERY and CYCLE (Brower, 2000, p.326).

5.4 Image Schemas And Design

Of explicit relevance to my design framework are a number of research findings outlined by Hurtienne et al. (2010). Their work suggests that the assimilation of primary metaphor and image-schematic constructs in designing our interactions with technology may foster intuitive understanding of the user interface (Hurtienne et al. 2010, p.483). They suggest that further investigation is needed in this area, specifically pointing to how their approach may assist in advancing the design of interfaces intended for increasingly complex tasks (*ibid*.). The generation and manipulation of rhythmic electronic material has already been shown to exist as a complex and multi-dimensional task, one that I argue is well suited to the above detailed methods of innovative interface design. Wilkie et al. (2009) provide a comprehensive account detailing their own approach to image-schematic analysis of conceptual models within the context of musical interface design. In their work they describe successfully extending basic notions of existing image schemas to include dependent (CONTAINER has CONTENT) or related schema (NEAR-FAR, PART-WHOLE) (*ibid*.). Their research indicates that we may improve musical performance systems through selecting more appropriate conceptual metaphors in the design process. My own interface design draws inspiration from Wilkie et al.’s discussion, and from Brower’s idea of image schemas yielding more complex structures through combination (of shared features). Figure 4 illustrates some of the core schemas at play in my design.

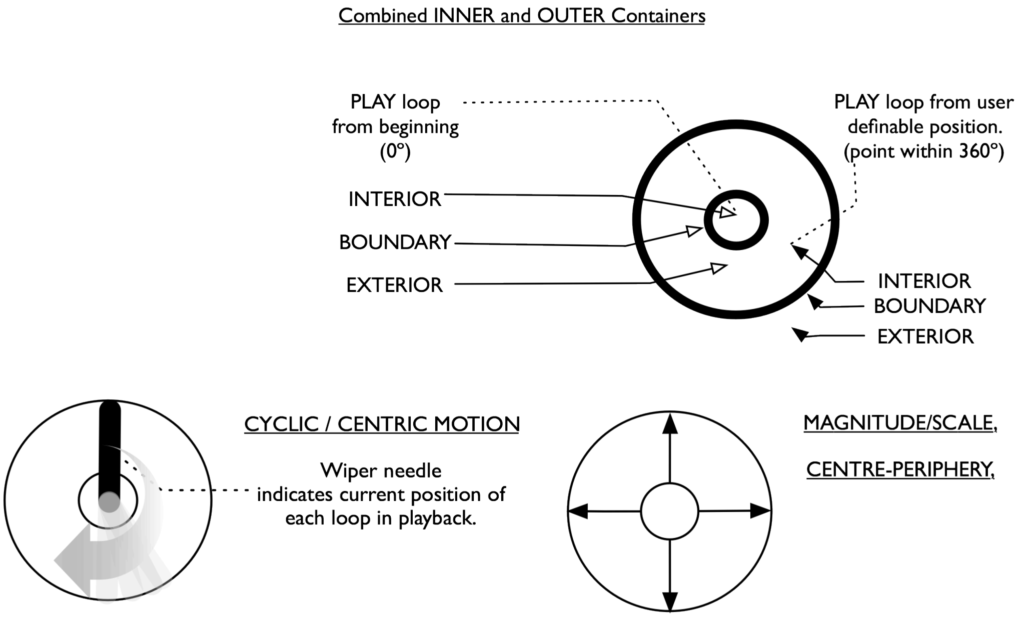


Figure 4 – Combined image schemas in use: NESTED CONTAINER, CYCLIC MOTION, CENTRE-PERIPHERY/MAGNITUDE/SCALE

6. Presenting the design of my own rhythmic electronic performance system: Axis

The rationale for splitting the functionality of this performance system between two parts (10” Android tablet device as input control interface and computer as audio engine) derives from the desire to keep this system modular, in order to allow for wide ranging experimentation during the design process, and potentially by others in the future. Through the use of a mature audio synthesis/sampling software package (Supercollider) with a professional soundcard (RME Fireface 800), running on a stable platform (OSX), my work overcomes certain limitations otherwise inherent in developing solely for mobile devices[12]. In effect, the approach taken by my work exploits the best features from both platforms: the processing power and sonic clarity of computer-based sound generation, and the intuitive creative potential offered from carefully considered multi-touch control. Figure 5 illustrates the different interaction spaces within the Axis performance interface.

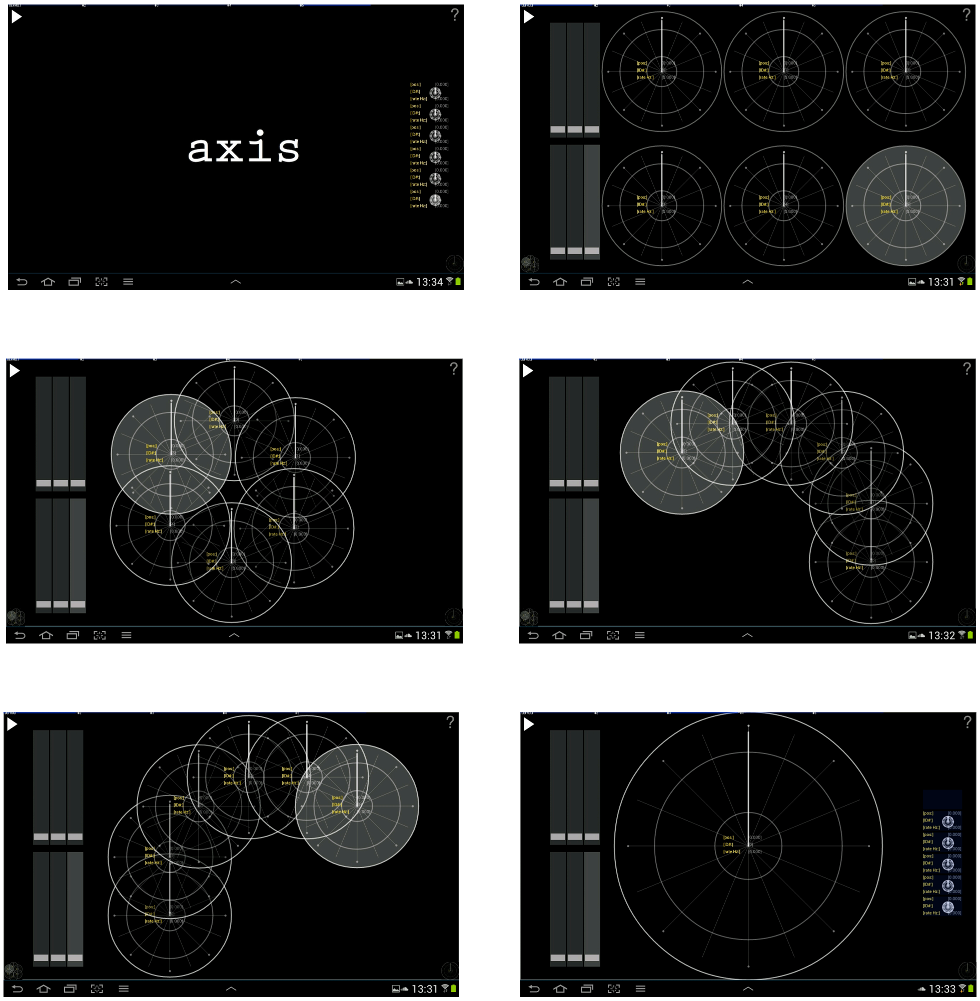


Figure 5 – Overview of the interaction spaces within Axis

---

[12] Allison, Oh and Taylor (2013) note that centralised audio engines allow for a more adequate approach to sound reproduction, in place of the “small and typically impotent sound system on mobile devices” (*ibid*).

My key contention is that the relevance of cyclic rhythmic structures within this musical area extends beyond the simple playback of looped audio, pertaining also to many manipulative techniques. My research applies performance strategies that facilitate the creation of complex rhythmic patterns, through the use of multiple cyclic elements that utilise short audio loops as their fundamental raw material. The minimalistic design presented incorporates image-schematic constructs, primary metaphor and gestalt phenomena in efforts to facilitate more intuitive use of technology (as encouraged by Hurtienne et al. 2010; Modler 1997; and Leman, 2008) through "embodied interaction" (Dourish, 2004) with a perceptually unified user interface. In summary, the system I have developed is highly extensible and may be tailored to suit the needs of users of different skill and experience levels. The system promotes practice-based learning of nested control possibilities within a cyclic interaction space, focused on rhythm. My approach to providing intuitive interactions with multiple loop-based musical structures is one of the core innovations of this work laying, firm foundations for a number of engaging future developments for this system.

References

Allison, J., Oh, Y. and Taylor, B. (2013) Nexus: Collaborative Performance for the Masses, Handling Instrument Interface Distribution through the Web. In *Proceedings of the International Conference on New Interfaces for Musical Expression Conference (NIME)*. KAIST, Daejeon, South Korea, 27th-30th May 2013. p.1-6. [Online] Available from: http://aimlab.kaist.ac.kr/nime2013/program/papers/day1/paper1/287/287_Paper.pdf [Accessed: 30th March 2015]

Brower, C. (2000) A Cognitive Theory of Musical Meaning. *Journal of Music Theory*. 44 (2). p.323- 379.

Chadabe, J. (2000) Remarks on Computer Music Culture. *Computer Music Journal*. 24 (4).

Cox, C. and Warner, D. (2004) (eds.) *Audio Culture: Readings in Modern Music*. New York and London: Continuum.

Dannenberg, R. B. (2007) New Interfaces for Popular Music Performance. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression (NIME)*. NYU, New York, 6th- 10th June. p.130-135. [Online] Available from: http://www.nime.org/proceedings/2007/nime2007_130.pdf [Accessed: 14th June 2011]

Dourish, P. (2004) *Where the Action Is: The Foundations of Embodied Interaction,* Cambridge, MA: The MIT Press.

From Berkeley to the Bronx: Roger Linn's Drum Machines. (2007) BBC Radio 4. 29th May.

Gann, K. (1995) *The Music of Conlon Nancarrow (Music in the Twentieth Century)*. Cambridge: Cambridge University Press.

Gibbs, R.W. Jr. (2011) Metaphors. In Runco, M.A. and Pritzker, S. A. (eds.) *Encyclopedia of Creativity*. 2nd Edition. Burlington, MA: Elsevier.

Gibson, J.J. (1986/1979) *The Ecological Approach to Visual Perception: Resources for Ecological Psychology*. New York: Psychology Press.

Goatly, A. (2007) *Washing the Brain – Metaphor and Hidden Ideology*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Greneo, J.G. (1994) Gibson's Affordances. *Psychological Review*. 101(2). p.336-342.

Hurtienne, J., Stößel, C., Sturm, C., Maus, A., Rötting, M., Langdon, P.M. and Clarkson, P.J. (2010) Physical Gestures for Abstract Concepts: Inclusive Design with Primary Metaphors. *Interacting with Computers*. [Online] 22 (6). p.475-484. Available from: doi:10.1016/j.intcom.2010.08.009 [Accessed: 20th June 2012]

Jacob, R.J.K., Sibert, L.E., McFarlane, D.C. and Mullen Jr., M.P. (1994) Integrality and Separability of Input Devices. *ACM Trans. Human Computer Interaction*. 1 (1). p.3-26.

Johnson, M. (2007) *The Meaning of The Body: Aesthetics of Human Understanding*. Chicago: The University of Chicago Press.

Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*. Chicago: The University of Chicago Press.

Lakoff, G. and Johnson, M. (1999) *Philosophy in The Flesh*. New York: Basic Books.

Lanier, J. (2010) *You Are Not a Gadget*. New York: Alfred A. Knopf.

Leman, M. (2008) *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: The MIT Press.

London, J. (2012) *Hearing in Time*. New York: Oxford University Press.

McGrenere, J. and Ho, W. (2000) Affordances: Clarifying and Evolving a Concept. In *Graphics Interface*. p.179-186. [Online] Available from: http://teaching.polishedsolid.com/spring2006/iti/read/affordances.pdf [Accessed: 7th July 2014]

McLachlan, N. (2000) A Spatial Theory of Rhythmic Resolution. *Leonardo Music Journal*. 10. p.61- 67.

Machover, T. (2002) Instruments, Interactivity and Inevitability. In *Proceedings of the International Conference on New Interfaces for Musical Expression, (NIME)*. Media Lab Europe, Dublin. 24th-26th May. p.115. [Online] Available from: http://www.nime.org/proceedings/2002/nime2002_115.pdf [Accessed: 13th May 2010]

Magnusson, T. (2009) Of Epistemic Tools: Musical Instruments as Cognitive Extensions. *Organised Sound* 14 (2).

p.168-176.

Modell, A. H. (2009) Metaphor – The Bridge Between Feelings and Knowledge. *Psychoanalytic Enquiry*. 29 (1). p.6-11. [Online] Available from: http://www.tandfonline.com/doi/abs/10.1080/07351690802246890 [Accessed: 29th January 2014]

Modler, P. (1997) Interactive Music Systems and Concepts of Gestalt. In Leman, M. (ed.) *Music, Gestalt and Computing: Studies in Cognitive and Systematic Musicology*. Berlin: Springer-Verlag.

Nielsen, J. (1994) Heuristic Evaluation. In Nielsen, J., and Mack, R.L. (eds.) *Usability Inspection Methods*. New York: John Wiley & Sons. *Usability Inspection Methods*, [Online 1995] Available from: http://www.nngroup.com/articles/ten-usability-heuristics/

Norman, D.A. (1988) *The Psychology of Everyday Things / The Design of Everyday Things*. 2002 edition. New York: Basic Books.

Rowe, R. (2001) *Interactive Music Systems: Machine Listening and Composing*. Cambridge, MA: The MIT Press.

Said, A. (2013) *The BeatTips Manual: The Art of Beatmaking, the Hip Hop/Rap Music Tradition, and the Common Composer*. 6th Ed., New York: Superchamp Books.

Sethares, W.A. (2007) *Rhythm and Transforms*. London, UK: Springer-Verlag.

Smirnov, A. (2013) *Sound in Z: Experiments in Sound and Electronic Music in Early 20th Century Russia*. London: Koenig Books.

Smith, L. (1973) Henry Cowell's Rhythmicana. *Anuario Interamericano de Investigacion Musical*. 9. p.134-147.

Tanaka, A. (2010) Mapping Out Instruments, Affordances, and Mobiles. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. University of Technology, Sydney, Australia, 15th-18th June 2010. [Online] Available from: http://research.gold.ac.uk/6834 [Accessed: 7th June 2010]

Tompkins, P. and Lawley, J. (2009) Embodied Schema: The Basis of Embodied Cognition. First presented at The Developing Group, 6th June 2009. [Online] Available from: http://www.cleanlanguage.co.uk/articles/articles/245/1/Embodied-Schema-The-basis-of-Embodied-Cognition/Page1.html [Accessed: 7th September 2012]

Wilkie, K., Holland, S. and Mulholland, P. (2009) Evaluating Musical Software Using Conceptual Metaphors. In Blackwell, A. (ed.) *Proceedings of BCS Human Computer Interaction*. p. 232-237.

Wishart, T. (1996) *On Sonic Art. Contemporary Music Studies. 12*. Emmerson, S. (ed.) 2nd edition. Amsterdam: Harwood Academic Publishers.

Zbikowksi, L. M. (1997) Conceptual Models and Cross-Domain Mapping: New Perspectives on Theories of Music and Hierarchy. *Journal of Music Theory*. 41(2). p.193-225. [Online] Available from: http://www.jstor.org/stable/843958 [Accessed: 6th January 2014]

The role of fitness measures in evolutionary composition

Róisín Loughran and Michael O'Neill
Natural Computing Research and Applications Group,
University College Dublin, Ireland.
`roisin.loughran@ucd.ie`

Abstract One of the biggest issues in applying evolutionary computational methods to creative applications such as music composition lies in the design of an appropriate fitness measure. Evolutionary Computational (EC) methods are driven by their fitness function; it determines how likely an individual is to reproduce or die out. If the individual is a piece of music, how can we create a fitness function that can reliably and fairly determine how fit this is? We present a number of versions of a system that composes melodies using Grammatical Evolution (GE). We give an overview of the system including the grammar and representation used. We describe three distinct variations of the system that incorporate alternative fitness measures to those typically used in EC. We provide a number of results including composed melodies and discuss how such systems may be developed from being purely generative to being able to demonstrate true creativity.

1. Introduction

Evolutionary Computation (EC) is a branch of natural computing based on Darwin's theory of survival of the fittest. In EC algorithms, a good solution is *evolved* from a population of solutions that is continuously improved over a number of generations. Individual solutions survive according to their measured *fitness*. This fitness is defined at the beginning of the experiment, generally by the programmer, as a measure of how well the given solution solves the proposed problem. In typical EC experiments such as classification or regression, this measure is easy to define. If however, we apply EC to a more aesthetic or subjective task, defining this fitness measure becomes an interesting problem in itself; how would one measure the fitness of a piece of music?

Generally, this problem is addressed using a pre-determined numerical measure, a random choice or a human observer. Using a pre-determined metric is not particularly creative, as taking an objective, numerical measure of how good one melody is over another is not likely to genuinely describe a subjective quality. Likewise, pure random choice is not creative; selecting one piece over another at random does not acknowledge any measure of merit, musical or otherwise, between them. The employment of a human observer, known as Interactive EC, is often used but must be considered less computational than autonomous methods as this is ultimately being driven by human choice. Hence, systems that employ any of these choices in their fitness functions are either less computational or less creative than a computationally creative system should be. We would like to examine the question: what kind of measures can one use within an EC system that do not rely on predefined musical knowledge, music theory rules, similarity to given style of music or a human observer? What could a system learn to like if we did not tell it what to like?

In this paper, we review a number of variations of an algorithmic compositional system based on Grammatical Evolution (GE), a grammar based form of EC. GE evolves an individual by using a context-free grammar to transpose an integer genome into the problem domain. In the proposed systems we transform the integer-array genome into a series of MIDI notes that can be played and listened to by the user as music. The differences in the systems lie in the manner in which the fitness function is defined for each system. Our initial system employs a user-defined measure of the tonality of the music. Our second system attempts to abstract conscious control of the melody away from the user by employing a distance metric that maps the distance between segments of music to the distance between a set of points in a two-dimensional plane defined by the user. Our third system attempts to take the level of control further from the user by creating a system of critics that evolve their own 'opinion' of what is popular and evolve music accordingly. This paper outlines each of these methods. We then compare the merits and limitations of each method and consider ways in which to bring aesthetic applications of EC further towards computational creativity or creative AI.

2. Previous Work

In recent years a number of studies have been undertaken applying EC techniques to the problem of algorithmic composition. Details of the EC algorithms described below can be found in Brabazon et al. (2015). GenJam was one of the first systems to use a Genetic Algorithm (GA) in music, to evolve jazz solos that have been used in live performances in mainstream venues (Biles, 2013). GAs were further modified to create GeNotator to manipulate a musical composition using a hierarchical grammar (Thywissen, 1999). More recently, GAs were used to create four part harmonies without user-interaction or initial material according to rules from music theory (Donnelly and Sheppard, 2011). Dahlstedt developed a Genetic Programming (GP) system that implements recursively described binary trees as genetic representation for the evolution of musical scores. The recursive mechanism of this representation allowed the generation of expressive performances and gestures along with musical notation (Dahlstedt, 2007). Grammars were used with Grammatical Evolution (GE) for composing short melodies in (Reddin et al., 2009). From four experimental setups of varying fitness functions and grammars they determined that users preferred melodies created with a structured grammar.

The various attributes used in the evaluation of melodies based on pitch and rhythm measurements were discussed in de Freitas et al. (2012). It was concluded that previous approaches to formalise a fitness function have not comprehensively incorporated all measures. Some studies addressed the problematic issue of determining musical subjective fitness by removing it from the evolutionary process entirely. GenDash was an early developed autonomous composition system that used random selection to drive the evolution (Waschka II, 2007). Others used only highly fit individuals and used the whole population to create melodies (Biles, 2013; Eigenfeldt and Pasquier, 2012).

3. Method: 'The Composing Pony'

GE is a grammar based algorithm based on Darwin's theory of evolution. As with other evolutionary algorithms, the benefit of GE as a search process results from its operation on a population of solutions rather than a single solution. From an initial population of random genotypes, GE performs a series of operations such as selection, mutation and crossover over a number of generations to search for the optimal solution to a given problem. A grammar is used to map each genotype to a phenotype that can represent the specified problem. The success or 'fitness' of each individual can be assessed as a measure of how well this phenotype solves the problem. Successful or highly fit individuals reproduce and survive to successive generations while weaker individuals are weaned out. Such grammar-based generative methods can be particularly suitable to generating music as it is an integer genome that is being manipulated rather than the music itself. This allows the method to generate an output with a level of complexity far greater than the original input. This added complexity generation is helpful in creating interesting and diverse pieces of music. In the system proposed, the grammar defines the search domain — the allowed notes and musical events in each composition. Successful melodies are then chosen by traversing this search space according to the defined fitness function.

We exploit the representational capabilities of GE resulting from the design of a grammar that defines the given search domain. GE maps the genotype to a phenotype — typically some form of program code. This phenotype can then be interpreted by the user in a predetermined manner. In these experiments, the programs created are written in a command language based on integer strings to represent sequences of MIDI notes. We design a grammar to create this command language which is in turn used to play music. The grammar used results in a series of musical sequences which may be single notes, chords, turns or arpeggios. An overview of the GE process including the mapping of the grammar to MIDI notes is shown in Figure 1. The melodies discussed in each of the methods described below can be found at `http://ncra.ucd.ie/Site/loughranr/music.html`.

4. System 1: Pseudo-tonality

The first system is of the traditional GE form with a grammar mapping and a pre-defined fitness measure to drive the evolution. This method bases fitness on a measure of the tonality within the given melody. An initial fitness measure is taken from the length of the individual:

$$fitness_{initial} = (Len - 200)^2 + 1 \tag{1}$$

where Len is the length of the current phenotype.

For an emergent tonality one pitch should be the most frequently played within the melody, with an unequal distribution of the remaining pitches. We define the *primary* as the pitch value with most instances and the *secondary* as that with the second highest number of instances. The number of instances of the seven most frequently played notes is defined as Top7 and the number of instances of the top nine notes as Top9. We evolve towards a good (low) fitness by controlling these relationships. For each of the following inequalities that hold:

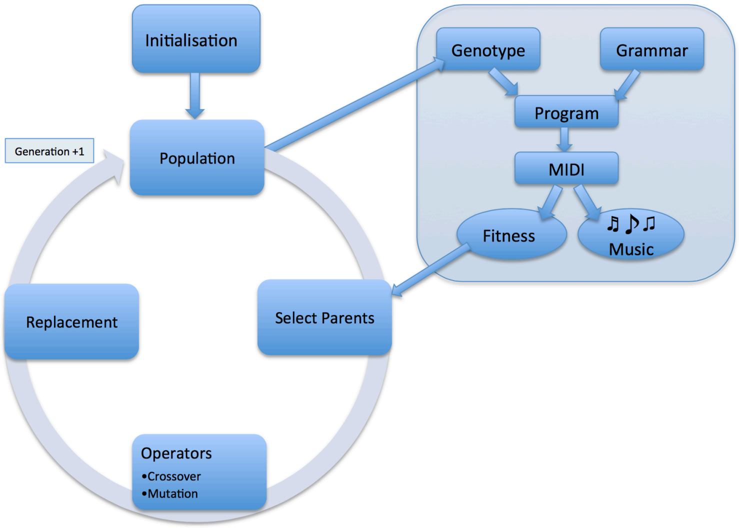


Figure 1: Overview of Grammatical Evolution.

$$\frac{\#\text{ instances of primary}}{\#\text{ instances of secondary}} < 1.3 \tag{2}$$

$$\frac{\text{Top7}}{\text{Total number of played notes}} < 0.75 \tag{3}$$

$$\frac{\text{Top9}}{\text{Total number of played notes}} < 0.95 \tag{4}$$

the fitness is multiplied by 1.3. This enforces the primary tone to have significantly more instances than the secondary and encourages most of the notes played to be within the top seven or top nine notes. These limits of 0.75 and 0.95 enforce more tonality than 12 tone serialism but will unlikely create a melody with typical Western tonality. For these experiments, the top four melodies in the final population are concatenated together to encourage the emergence of themes within the final compositions. A full description of this system can be found in Loughran et al. (2015).

4.1. Results

To consider the success of the system we track how the best fitness evolves in relation to the average fitness of the population. To examine this, we ran our system 30 times and took the mean of the best and average fitnesses over the 50 generations. A plot of the log of the results is given in Figure 2. It is clearly evident from these runs that although the best fitness converges quite quickly to the target, at the end of the run the average fitness is still very high. This implies that after 50 generations the population is still very diverse. Generally in EC tasks we would like the population to converge on the correct solution. This diversity indicates that the fitness measure currently used is not allowing the population to converge and that the individual melodies in the final generation are very different from one another. In contrast to traditional EC problems, for a

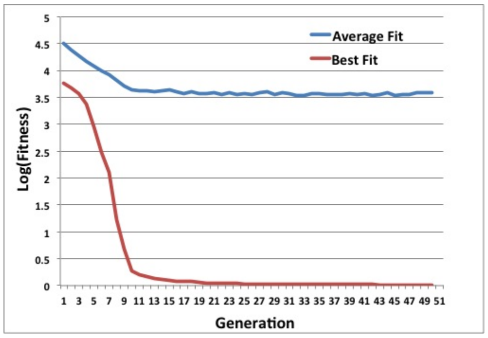


Figure 2: Average and Best Fitnesses over 30 runs. Fitness values reported as $\log_{10}(Fitness)$.

creative compositional application a diverse population may be acceptable or even desirable; we cannot assume there is one global best melody within the population.

One of the best aspects of using EC methods to evolve melodies is that it results in a population of solutions rather than a single result. Because EC causes convergence of solutions towards an ideal, this means that by the end of an evolutionary run we have a number of individuals with high fitness. In this creative domain, this results in a number of similar but not identical melodies. With this system we can take advantage of this to create longer melodies by concatenating the top four individuals together. This joining of similar but different melodies together causes the emergence of variations on a theme. These emergent themes or motifs are clearly audible in a number of the melodies that accompany these experiments.

5. System 2: Shape Matching

The emergence of themes from similar but different individuals in the first system inspired us to create a fitness function based on this ‘dissimilarity measure’. System 2 uses a hill-climbing methodology to create a melody composed of a number of segments. The system is evolved towards a given level of dissimilarity between all segments in a population. This dissimilarity is defined at the beginning of a run by the user who places a series of dots in a square using a web-based GUI. The distance between each of these dots is measured and the melodic segments are evolved towards forming the same distances in their own musical metric space.

To define the distances in musical space, each melody segment is expanded to give a pitch value at each time step (duration of one demisemiquaver) and normalised to start at 0 resulting in a pitch contour. The distance between two contours is taken as the sum of the absolute distance between their pitch contours at each time step plus five times any length difference between them. These distances are then compared to the distances defined by the shape of dots depicted by the user at the start of the experiment. For these experiments, the four shapes shown in Figure 3 are considered.

The algorithm used in these experiment is a hill-climbing variation of GE. At each generation five operators are implemented in sequence: Switch, Permute, Crossover, Mutate and Copy. Once an improvement is found, the population is updated and the system moves on to the next generation resulting in a Variable Neighbourhood Search. Four variation of the system are implement:

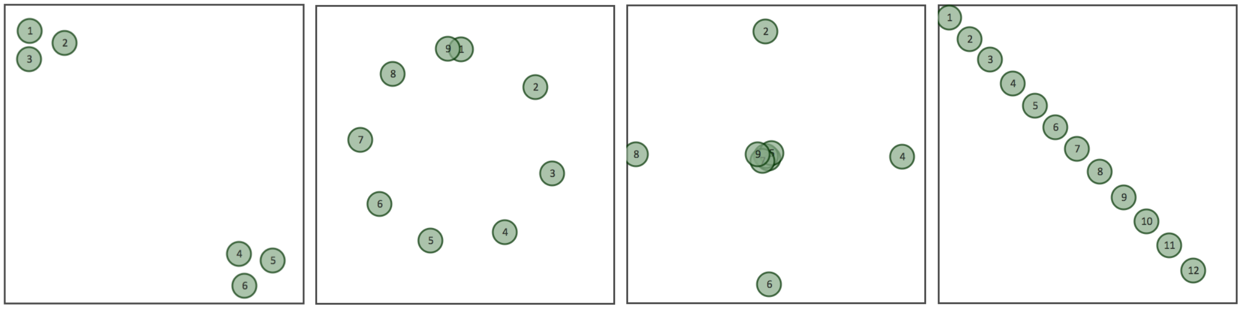


Figure 3: Shape Targets for Cluster, Circular, Cross and Line melodies created by the Music Geometry GUI

Table 1: Average (over 30 runs) best Fitness after 1000 generation achieved with each method for each pattern with System 2. Standard deviation is shown in parenthesis.

Method	AllMu01	AllMu1	NCMu01	NCMu1	Random
Cluster	5.37(1.9)	**4.49**(1.8)	28.79(11.4)	21.15(15.0)	252.52(26.6)
Circle	13.23(4.5)	**12.78**(3.15)	35.88(14.5)	23.48(8.26)	535.861(54.7)
Cross	10.54(5.7)	**7.55**(3.4)	40.08(18.7)	17.89(11.3)	556.21(54.3)
Line	19.14(5.9)	**15.44**(3.7)	50.45(18.8)	32.73(13.3)	864.48(65.7)

AllMu01 uses all operators with a Mutation Coefficient (μ) of 0.01, AllMu1 uses a μ value of 0.1, NCMu01 does not include the Copy operator with a μ of 0.01, NCMu1 does not use Copy with a μ of 0.1. All experiments are compared with Random Search. Full details of this experiment can be found in Loughran et al. (2016).

5.1. Results

Results for the 'Cross' shape experiments averaged across 30 independent runs are shown in Figure 4. It is clear that each version of the system achieves better fitness than random search. It is also evident that the AllMu01 and AllMu1 experiments converge faster than the methods that do not include the Copy operator demonstrating that Copy is important for quick convergence. NCMu01 performs worst, implying that without the Copy operator a high μ is very important in traversing the search space to find a good solution. Similar evolution trends were found for the alternative shape experiments. From the fitness curves, it appears that AllMu01 and AllMu1 display a very similar fitness performance. To confirm their performances, the best final fitnesses achieved after 1,000 generations were examined. The average best fitness achieved at the end of each run is shown in Table 1. From this table it is evident that AllMu1 is the highest performer in each version of the system. This indicates that unlike in standard GE experiments, a higher μ of 0.1 is more beneficial to the system than the typical 0.01.

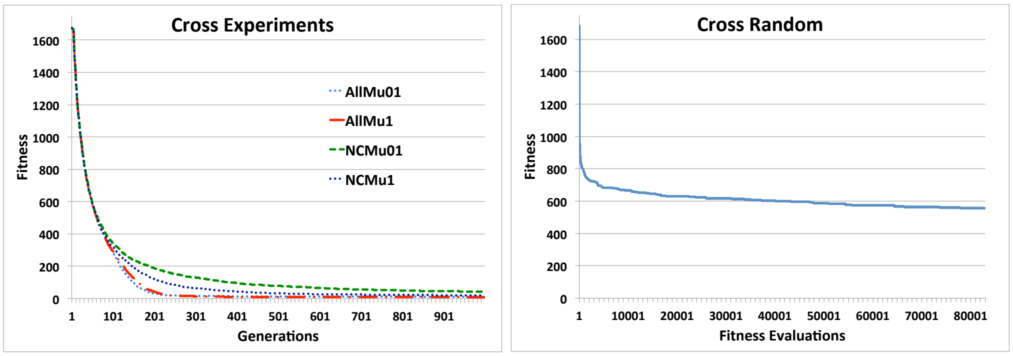


Figure 4: Average best fitness for each experimental run and Random Initialisation for Cross melody shape.

6. System 3: The Popular Critic

The third system we propose is called the 'Popular Critic'. This is a cyclical system that, once initialised, uses an internal self-referential method to determine the fitness of a piece of music. There are three distinct stages to the system. In Stage 1, 40 melodies are created and stored in a corpus using the pseudo-tonality method described above in Section 4. In Stage 2, a population of critics are evolved that result in a numerical judgement of these melodies. They are evolved with GE using a simple grammar that results in a linear combination of the number of distinct pitch and duration values in each melody. It is important to note that these fitnesses have *no aesthetic value* but are merely a way to numerically distinguish the melodies. The 40 melodies are then ranked by each Critic according to this numerical output. A general 'opinion' of the population of Critics can then be found by averaging the numerical results of each of the 50 Critics for each of the 40 melodies. This overall ranking of all 40 melodies is taken as the popularity consensus of the population. The fitness of each individual Critic is then calculated according to how closely it correlates with this overall popularity, hence the fitness of the individual Critic is aligned with how much it conforms to the consensus of the population of Critics. The Kendall-Rank Correlation is used to calculate this fitness. Selection, Crossover and Mutation are then performed over successive generations to evolve one best 'Popular Critic' as with typical EC methods. The best evolved Popular Critic is saved to be used to evolve new music in the final Stage 3 of the system. Although each stage of the system uses evolutionary methods, the self-adapting nature of this system proposes a move away from pure EC and towards a complex adaptive system. An overview of the system is shown in Figure 5. Full details of this system can be found in Loughran and O'Neill (2016b).

6.1. Results

The fitness improvement of the Critic Evolution (Stage 2) from the Popular Critic systems is shown in Figure 6(a). As can be seen, both the best and average fitness display a dramatic improvement in the first 10 generations. This improvement gradually tapers off in the following 10 generations and remains approximately stable thereafter. As described earlier, the fitness of the individual is taken as a measure of the correlation between the individual's numerical output and the most popular

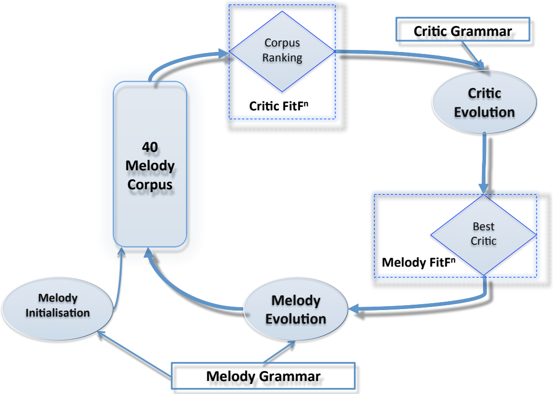


Figure 5: Flow diagram of the Popular Critic System.

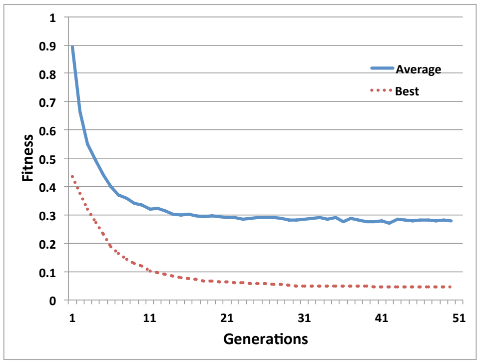


Fig 6(a): Fitness over 50 generations

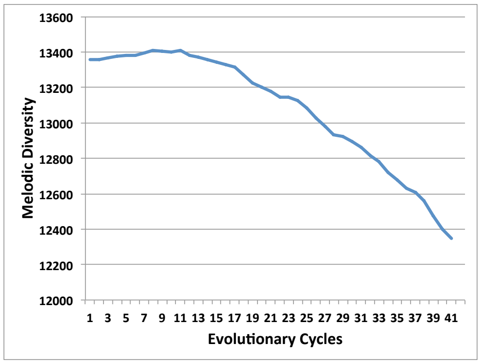


Fig 6(b): Melody Diversity, 40 cycles

opinion of the overall population. Over successive generations, we would expect the best fitness to improve as the population converges on a 'most popular' vote and one individual manages to approximate it. Hence, as expected the best and average fitness is seen to improve, but as crossover and mutation are used until the final generation the population does not converge completely.

Once the best Critic has been evolved, it is used as a minimising fitness function in a further evolutionary run to create new music. The grammar used is the same as that used to create the original corpus. A population of 100 melodies is evolved, the best four of which are combined to create the resultant melody. At the end of each run, this resultant best melody replaces a melody in the original corpus and the cycle starts again. When this cycle is repeated 40 times, the initial corpus of melodies has been completely replaced by melodies created by the system. To consider the change in the corpus over the course of the run, we compared the diversity of the melodies present in the corpus as the corpus was re-filled by evolved melodies. The diversity of the given corpus was measured as the sum of the Levenshtein distances between the representation of each pair of melodies in the corpus. The change in this diversity across 40 runs is shown in Figure 6(b). This shows that for approximately 10 evolutionary cycles, the diversity does not change dramatically from that of the initial corpus, but after 15 cycles, as the corpus is filled by newly evolved melodies there is a steady decrease in this measured diversity. The decrease is small, but nevertheless displays a definite trend. This shows that the process is having a directed effect on the melodies being produced. The original corpus was created without any preference from a Critic. This diversity reduction shows a move towards similarity in melodies created by the Critics.

7. Discussion

The most notable point to make from these experiments is that, on the surface, the output of each system can sound very similar. Melodies created by System 2 may be identifiable from the amount of segments used but, apart from that, even the author would find it hard (if not impossible) to discern purely from the musical output which one of these systems created a given melody. This is because a similar grammar and similar high level considerations are used for each experiment. Regardless of the system, the grammar results in a series of notes, turns and arpeggios, with a number of similar yet slightly different melodies played in succession. Furthermore, these are all played though a MIDI piano sound, resulting in similar sounding musical output. This highlights that when using any EC system to create music, it is not merely the fitness measure that is important; the representation used defines the space of possible music while the fitness measure is used to traverse this search space. Therefore, systems that use the same representation will result in similar sounding music. It is of course possible to transform the output afterwards — such as using a different MIDI instrument, or not using MIDI at all, but this is a higher level choice of the user and is not controlled within the system.

We intentionally used the same representation for the systems above as it is the fitness measure we wish to consider. Regardless of the representation, the fitness function must somehow make a judgement of the given music. We tried to move away from typical systems that judge purely on a 'musical' measure based on conformity to Western tonality or a given style and instead tried to develop functions that abstract the musical control away from the programmer and the user. In this way, we hope to move towards satisfying the Lovelace Test for Creativity (Bringsjord et al., 2003) which states the programmer should not be able to explain the creative output of their system. Evaluation of such systems involves more than merely judging the output; we have described above three very conceptually different systems that produce very similar music. Therefore it is imperative that when judging such systems we consider the context, reasoning and motivation for the methods undertaken in each step of their creation and avoid the pitfalls of simply judging a system by its output (Loughran and O'Neill, 2016a). We feel there is an opportunity to use EC methods to learn more about how an autonomous system may learn to judge aesthetic artefacts. Only by looking at systems that use autonomous, abstracted yet reproducible and defensible fitness measures can we progress EC algorithms into this new domain of computational creativity.

8. Conclusion

We have presented three specific systems that use variations of GE methods to create autonomous MIDI melodies. No key or time signatures were specified and the systems were not required to evolve towards any specific style or genre. We have shown that the representation used is highly influential on the output produced and yet it is the fitness measure that drives the evolution and hence makes the aesthetic judgments throughout the process. We are interested in using such systems to investigate how EC methods may be employed with more creativity in an aesthetic domain such as music. For future work, we are looking to develop more autonomous fitness measures that may help us investigate aesthetic evaluation and computational creativity.

Acknowledgments

This work is part of the App'Ed project funded by SFI under grant 13/IA/1850.

References

John A Biles. Straight-ahead jazz with GenJam: A quick demonstration. In *MUME 2013 Workshop*, 2013.

Anthony Brabazon, Michael O'Neill, and Seán McGarraghy. *Natural Computing Algorithms*. Springer, 2015.

Selmer Bringsjord, Paul Bello, and David Ferrucci. Creativity, the Turing test, and the (better) Lovelace test. In *The Turing Test*, pages 215–239. Springer, 2003.

Palle Dahlstedt. Autonomous evolution of complete piano pieces and performances. In *Proceedings of Music AL Workshop*. Citeseer, 2007.

Alan RR de Freitas, Frederico G Guimaraes, and Rogério V Barbosa. Ideas in automatic evaluation methods for melodies in algorithmic composition. In *Sound and Music Computing Conference*, 2012.

Patrick Donnelly and John Sheppard. Evolving four-part harmony using genetic algorithms. In *Applications of Evolutionary Computation*, pages 273–282. Springer, 2011.

Arne Eigenfeldt and Philippe Pasquier. Populations of populations: composing with multiple evolutionary algorithms. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design*, pages 72–83. Springer, 2012.

Róisín Loughran and Michael O'Neill. Generative music evaluation: Why do we limit to 'human' ? In *Computer Simulation of Musical Creativity (CSMC)*, Huddersfield, UK, 2016a.

Róisın Loughran and Michael O'Neill. The popular critic: Evolving melodies with popularity driven fitness. In *Musical Metacreation (MuMe), Paris*. 2016b.

Roisin Loughran, James McDermott, and Michael O'Neill. Tonality driven piano compositions with grammatical evolution. In *Evolutionary Computation (CEC), 2015 IEEE Congress on*, pages 2168–2175. IEEE, 2015.

Róisın Loughran, James McDermott, and Michael O'Neill. Grammatical music composition with dissimilarity driven hill climbing. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design*. Springer, 2016.

John Reddin, James McDermott, and Michael O'Neill. Elevated Pitch: Automated grammatical evolution of short compositions. In *Applications of Evolutionary Computing*, pages 579–584. Springer, 2009.

Jianhua Shao, James McDermott, Michael O'Neill, and Anthony Brabazon. Jive: A generative, interactive, virtual, evolutionary music system. In *Applications of Evolutionary Computation*, pages 341–350. Springer, 2010.

Kurt Thywissen. GeNotator: an environment for exploring the application of evolutionary techniques in computer-assisted composition. *Organised Sound*, 4(02):127–133, 1999.

Rodney Waschka II. Composing with genetic algorithms: GenDash. In *Evolutionary Computer Music*, pages 117–136. Springer, 2007.

Applying a Group Delay technique to parameter resynthesis of Brazilian Cuica sounds decomposed using the Modal Distribution

Thomas Lysaght
National University of Ireland, Maynooth
tlysaght@cs.nuim.ie

Joseph Timoney
National University of Ireland, Maynooth
jtimoney@cs.nuim.ie

The Modal distribution is a time-frequency distribution specifically designed to model the quasi-harmonic, multi-sinusoidal, nature of music signals and belongs to the Cohen general class of time-frequency distributions[1]. Using an improved modal smoothing kernel to help reduce cross term and sidelobe interference in automatic partial extraction, we have explored modal distribution synthesis of orchestral instrumental such as the clarinet [2]. One drawback of the Modal distribution is that it is real-valued and so no phase information is present allowing a synthesis from estimates of instantaneous amplitude and frequency only. However, it does provide accurate estimates of group delay computed as a moment of the distribution. In [3] we demonstrated signal synthesis from modal distribution partial tracks using the group delay along each partial track applied to synthetic signals.

In this paper we further develop the work in [3] by focusing on quasi-harmonic synthesised signals and then explore synthesis from the Modal distribution of recordings of the Brazilian cuica examining the effect of correcting the magnitude of partials using group delay. We used 23 recordings of a small Brazilian Cuica that exhibited characteristics of chirp-like harmonic sounds. We specified the smoothing windows in time and frequency based on the fundamental pitch of each instrument. This provided a sufficient separation between partials for cross-term suppression and showed the superior localisation in time and frequency of tracks using the Modal distribution. For each track we calculated the group delay as a local moment at each hop step about the main lobe width of the smoothing kernel. We then applied an alignment in time to each track magnitude using the group delay similar to the method of reassignment[4]. We then evaluate the effectiveness of this method in correcting the signal shape compared with the original synthesised MD tracks from instantaneous magnitude and frequency and draw conclusions from listening tests comparing the timbre of the synthesised sounds from both methods. The results show that in some cases the resynthesis is improved in terms of signal shape reconstruction, but in others it is not so convincing. Conclusions are drawn regarding these results and suggestions for improvement are given.

References

[1] Cohen, L. *Time Frequency Analysis*. Prentice-Hall, New Jersey, 1995.

[2] Lysaght, Thomas and imoney, Joseph and Lazzarini, Victor (2012) '*Automatic Partial Extraction from the Modal Distribution'*. In: DAFX-12 the 15th Int. Conference on Digital Audio Effects, September 17-21 2012, York, Uk

[3] Thomas Lysaght, Joseph Timoney '*Modal Distribution Phase Synthesis using Group Delay*' ISSTC 2015 Irish Sound Science Technology Convocation (2015)

[4] F. Auger and P. Flandrin. '*The why and how of time-frequency reassignment.'* IEEE International Symposium on Time-Frequency and Time-Scale Analysis, pages 197–200, 1994. Philadelphia

Mixed methods case studies in the assessment of impaired capabilities

Dr Brendan McCloskey
Ulster University School of Creative Arts and Technologies

Dr Brian Bridges
Ulster University School of Creative Arts and Technologies

Prof Frank Lyons
Ulster University School of Creative Arts and Technologies

Introduction

The authors present the results of a series of case studies designed to qualitatively assess the degree of upper limb motor control amongst a small group of digital musicians with quadriplegic cerebral palsy. These case studies formed part of a larger project: the design of a customised digital musical instrument for enhanced performative independence and real-time dynamic control.

Project background

The project participants are eminently familiar with a range of consumer and specialised digital musical instruments (figure 1), through their activities in inclusive music settings. The vast majority of such tools, designed for the average user, employ a variety of buttons, sliders, menu layers, and dials – artefact multiplication – and are consequently deemed non-optimal for independent dynamic control. Inclusive music practices rely on facilitator intervention, to set up and maintain software and hardware settings, a process that undermines real-time creative independence.

Figure 1. Two examples of control interfaces familiar to the participants.

In pursuit of an accessible yet sophisticated digital musical instrument (DMI) the authors first designed a series of case study methods ad formats – qualitative, indicative, game-based – to allow a gradual and organic revelation of actual capability. The studies were explicitly designed to progress from broad control parameters (long, short, loud, soft) to finer details (duration, dynamic force, dyads).

Study data

Of the three case studies it is the data from the third that is of particular interest (for a more detailed discussion of the study data, see [1]). The studies employed a simple customised pressure pad, with embedded force sensor and dynamic LED (figure 2). In the first study the sensor data was mapped to a simple sound synthesiser, with force data mapped directly onto sound loudness and duration; the participants demonstrated intentional control over long, short, soft and strong targeting gestures.

Figure 2. The customised sensor pad; a silicone rubber pad with embedded force sensor and LED.

In the second study control data was gathered *discretely,* providing indicative data on the individuals' responses to those subjective descriptors. The third, game-based study is of particular interest as it provided an unexpected outcome: the participants discovered a competitive element to the study, and challenged each other to 'do better', both personally and within the group. The control data from the sensor was mapped onto a virtual on-screen paddle and ball; the stationary paddle is moved using the sensor's force data, and the ball is assigned physical properties of mass and restitution. Striking the sensor pad causes the paddle to rise and ball to bounce, before slowly coming to rest.

The participants responded well to subjective descriptors of targeting force and duration employed in the previous studies. However, the competitive element arose when they discovered the challenge of raising the paddle *and thereafter maintaining constant pressure until the ball came to rest*. Quadriplegic cerebral palsy causes permanent and profound impairment to upper limb motor skills; designing a case study to *explicitly test* capability in this domain is ethically questionable (such studies are, however, important in clinical and rehabilitation sciences), and the study methods described herein allow the participants to independently explore the

---

[1] McCloskey, B., *et al* (2015). *Accessibility and dimensionality: enhanced real time creative independence for digital musicians with quadriplegic cerebral palsy*, in Proceedings of NIME'15, May 31-June 3, 2015, Louisiana State University, Baton Rouge, LA.

affordances and constraints of the system, without being challenged or stigmatised, giving them a greater sense of comfort, and ownership in both the process and the design prototype.

Future work

The authors produced a prototype instrument (figure 3), based on data from a variety of study methods and formats. It provides the performers with enhanced creative independence and dynamic control of a sophisticated yet transparent instrument model. The authors now intend to develop and reiterate the case study in more inclusive settings – larger study groups, and other types of disability. Furthermore, the technical elements of the prototype interface can be refined, increasing the range of creative control while reducing the size and shape of the interface.

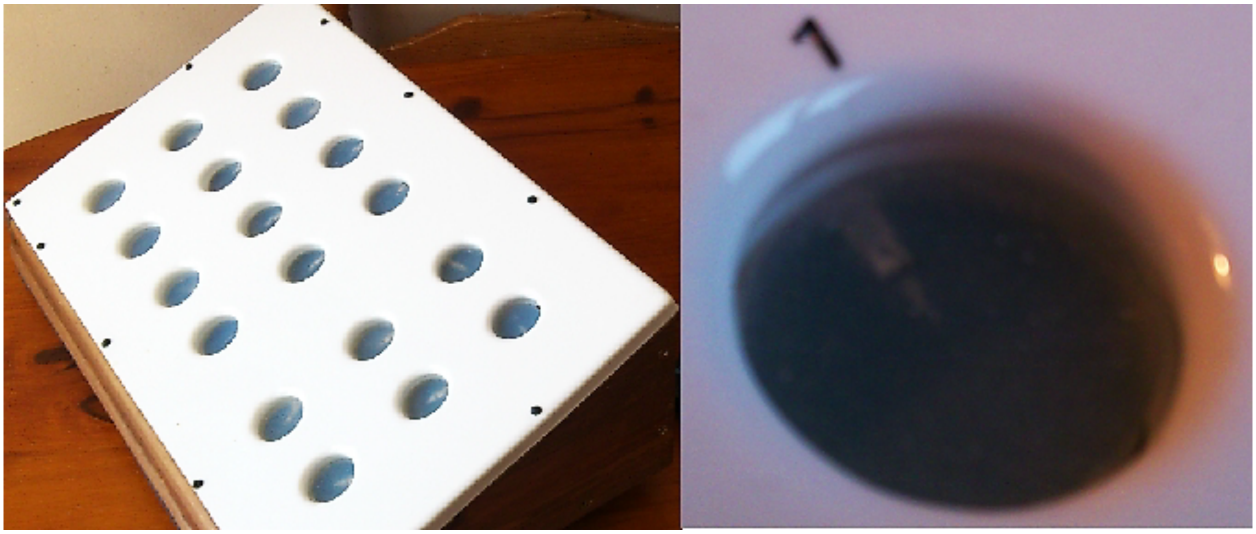


Figure 3. The final prototype instrument.

Comparing visualisations and sonifications of a large song dataset

James McDermott
Management Information Systems *and*
Natural Computing Research and Applications Group,
University College Dublin, Ireland.
`jmmcd@jmmcd.net`

Abstract Data visualisation is ubiquitous in science and business, but sonification occupies a much smaller niche. In order to explore the boundaries of this niche, we discuss the different purposes which visualisation and sonification can serve. We compare several visualisations and sonifications of a large popular music dataset, in order to explore their pros and cons. We compare the insight achieved using the various methods.

1. Introduction

Data visualisation is ubiquitous in modern science and business. It helps us to make abstract and complex data concrete and understandable, and it is easy to see why. As physical beings, our intelligence is *embodied* (Anderson, 2003). In particular, our visual sense is intertwined with our thinking, so deeply intertwined that it is difficult to carry on any conversation without using visual metaphors. We see the other person's point of view; we look into the past; we see things in our minds' eye; we see a way out of a problem. As a result, the very important problem of understanding data is deeply connected with *seeing* it. When we use data visualisation in this way, we are "using vision to think" (Card et al., 1999).

It is natural to think about generalising from vision to other senses. Sonification is the auditory analogue of visualisation. It is described in the *Sonification Handbook* as "rendering sound in response to data and interactions", or "the use of [non-speech] sound to convey information" (Hermann et al., 2011).

In recent years, sonification has become a recognised area of research (Hermann and Hunt, 2005; Kramer et al., 2010; Hermann et al., 2011). There is a dedicated conference (the International Conference on Auditory Displays), edited books (Hermann and Hunt, 2005; Hermann et al., 2011), and special issues of journals including *Artificial Intelligence and Society*, *Organised Sound* and *Displays*.

1.1. Between functionality and aesthetics

In visualisation, different works put different emphasis on two goals which in a sense lie at opposite ends of a spectrum – *functionality* and *aesthetics*. In visualisation, the two ends of this spectrum correspond to the charts and graphs typical of science and business whose goal is to convey information, and (much less common) the type of data-driven art in which patterns arising in data are enjoyed for their own sake. Of course, even when the real goal is a functional one, aesthetic considerations are rarely ignored: a beautiful piece of work will likely achieve more clarity and more attention than an ugly one, all else being equal.

An important special case is the *infographic*, which lies somewhere in the middle of this spectrum. Typically, a punchy, journalistic style is used in both text and graphics, and the result should be both informative and aesthetically pleasing, though rarely would it be called a work of art.

The same categories can be applied to work in sonification. The goal is generally to achieve a result which is *informative* and/or *aesthetically valuable* (Barrass and Vickers, 2011, p. 157). Some sonifications are primarily functional, such as the auditory displays used in hospital intensive care units or in airplane cockpits. The former are notoriously lacking in aesthetic value. Work on sonification does often consider aesthetics, but sometimes only in service to the functional goal, simply because people won't listen to ugly works. Roddy and Furlong (2014) acknowledge this, but call for work in which aesthetics are elevated from this position, and in which embodied aspects of aesthetics are primary.

At the other end of the spectrum, data-driven music or sound art is common, perhaps more common than data-driven visual art. Non-theoretic music, from nursery rhymes to folk music, is already abstract and patterned rather than figurative, in contrast to non-theoretic visual art: if data-driven sound art is more accessible to most audiences than data-driven visual art, then this could help explain why.

Finally, as an example from the middle of the functional–aesthetic spectrum, the BBC World Service has begun to use "audiographics" for a novel alternative way to convey data related to current topics, for example "Buzzword frequency in US presidential candidate debates"[1] and "Global infant mortality continues to fall"[2]. These short pieces are intended to convey a sense of some dataset, often but not always a time-series, together with some integrated commentary. In this they function more as the auditory analogue of an infographic, rather than of a bare plot. They have entertainment value – not quite the same thing as aesthetic value.

It seems likely that a further part of the motivation for these audiographics is that they are well suited to their medium. This medium is not only radio, but also web audio. In the modern internet, much content is consumed as video, and hence there is an audience of listeners who consume with audio speakers or headphones at the ready. Users of computers and mobile devices now use them for audio much more frequently than heretofore. This suggests that the market for sonification (whether presented via the public internet, or via private means) is growing, and will continue to grow.

---

[1] `http://www.bbc.co.uk/programmes/p03b33n7`

[2] `http://www.bbc.co.uk/programmes/p039d6mp`

1.2. When is sonification useful?

Sound lives in the *medium* of time. So, sonification seems naturally suited to datasets where change over time is explicitly represented in the data, for example in time-series data and time-stamped event-lists. Time-varying processes such as growth, decay, sudden and gradual change, cyclical behaviour, chaos and regularity can be represented and perceived through sound. Roddy and Bridges, in this volume, call for sonification of these types of processes to be linked to a long line of thinking in fields ranging from psychology to sound art, on embodiment and on environmental and gestural metaphors – for example the *Spectromorphology* framework of Wishart and Emmerson (1996).

(The distinction above, between event-list and time-series data, is important to understanding the pros and cons of both visualisation and sonification methods. Large event-list datasets are common, for example in web server logs. They cannot be treated in exactly the same way as time-series datasets, although sometimes they can be pre-processed to give time-series.)

Sonification is also potentially useful as a monitoring mechanism for “live” data. In situations where a user is required to monitor some background process even while their eyes are occupied with a foreground task, sonification of the background process may be sufficient[3]. This type of usage is simply using more of our available sensory bandwidth. Sonification is also useful for visually-impaired users in obvious ways.

Finally, sonification is useful as a novelty, or a change from visualisation, or where some entertainment or aesthetic purpose can be served through sonification which could not be served through visualisation.

1.3. Goals of this paper

The first goal of the present paper, then, is to explore the differing goals of multiple visualisation and sonification techniques on a large, real-world dataset, and to draw conclusions about their advantages and disadvantages.

We have chosen the Million-Song Dataset (Bertin-Mahieux et al., 2011) for this purpose. The MSD consists of one million songs from the 20th and early 21st centuries. It was curated by Bertin-Mahieux et al. (2011) in collaboration with Echo Nest, a music discovery service. It is available online[4]. The full dataset includes detailed audio analysis data, though it does not include the audio itself, and comes to approximately 280GB. The full dataset and various sub-setted and pre-processed versions, available from the same website, have been used as a testbed for many machine learning tasks. We will concentrate on a half-million song subset consisting of 90 timbral features, labelled with the year of release. It is thus a time-stamped event-list dataset, where each song is an event. With preprocessing it can be seen as a time-series dataset. For this reason, the MSD is an interesting dataset for our purposes. It is also thematically suitable, because it concerns music. Sonifications have an opportunity to take advantage of this for aesthetic effect.

Broadly, study of this dataset should allow insight into how timbre of popular music has changed over time. Our study of the dataset will begin with pre-processing to achieve a time-series. Both the event-list and the time-series formats will be used for visualisation and sonification.

---

[3] `https://www.oreilly.com/ideas/cameron-turner-on-the-sound-of-data`

[4] `http://labrosa.ee.columbia.edu/millionsong/`

The visualisation methods we will compare will include several methods of achieving a two-dimensional embedding of the MSD features (other than the song's year label). Ideally, an embedding should achieve a "spread" of the data, and should convey some understandable structure. We compare the methods on these criteria. We also compare visualisations of the event-list and the time-series, with and without embedding. All visualisations are produced using well-known methods.

We also provide two sonifications of the data. The first is based on the original event-list format, and granular synthesis. Each event gives rise to one "grain" of sound. In each case there is a direct mapping from a long period of time (the 90 or so years spanned by the dataset) to a short period (e.g. 1 second per year gives a piece lengthy enough to convey its message, which does not outstay its welcome).

The second sonification uses less well-known methods. The author's previous work (McDermott and O'Reilly, 2011; Crowley and McDermott, 2015) has described a system in which arbitrary time-series can be mapped through complex arithmetic transformations to MIDI output.

The two sonification systems are flexible enough to produce many very different sonifications, with very different aesthetics, from the same data. A key idea in both is that the parameters of the mapping (from the event-list, via granular synthesis, in the first case; from the time-series, via arithmetic transformations, in the second) are too complex for direct user control, but at the same time the aesthetics are too important to be left uncontrolled. A form of *indirect* user control is used instead, that is *interactive evolutionary computation* (IEC) (Takagi, 2001). Here, the user interacts with the parameters only by listening to their results, and either accepting or rejecting them, iteratively. Behind the scenes, the IEC system runs an evolutionary search process guided by the user's judgements, which gradually finds parameter settings fitted to the user's aesthetic goals.

The second goal of this paper, then, is to show how IEC can be used as a means of user control of aesthetics in sonification without requiring the user to have expertise in the design of parameters and mappings.

2. Visualisation

2.1. Event-lists and time-series

We begin by visualising the data directly in Fig. 1. For the event-list data, we produce a scatterplot, placing time on the horizontal axis and each feature on its own vertical axis. We then produce a 90-dimensional time-series by taking the mean value for each feature over all events for each year. We overlay this over the scatterplots in Fig. 1.

The scale of the data (half a million data points) already causes a problem in this visualisation – in PDF format, Fig. 1 consumes hundreds of megabytes and is unwieldy on-screen. Saving as PNG avoids this problem, at the cost of image quality when zoomed in.

The event-list scatterplots show the increasing density of the data over time. Some scatterplots seem to show a shift in distribution over time, but this is partly due to an increased sample size, hence more extreme values, in later years. The time-series (mean values) may give a better

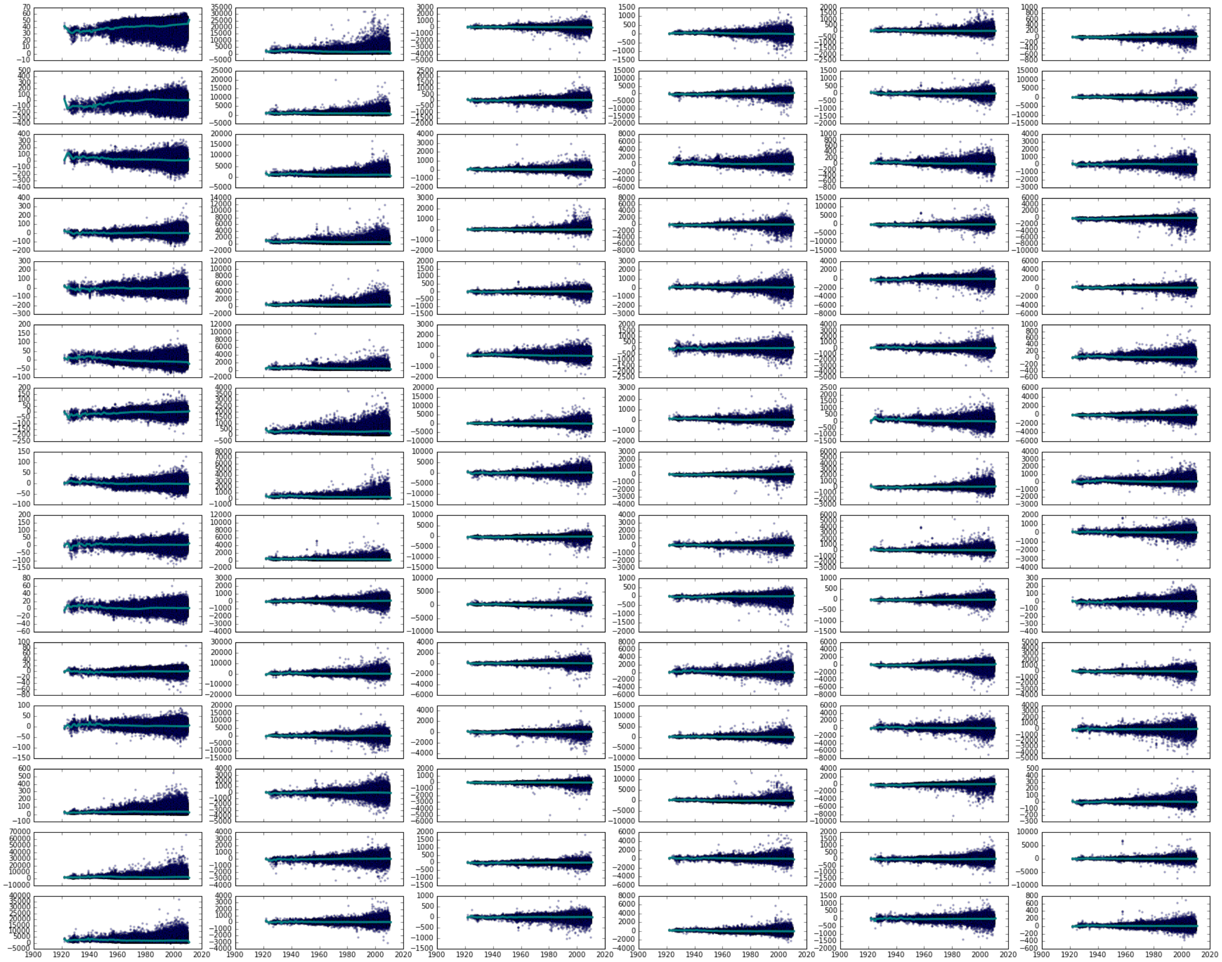


Figure 1: 90 timbral features, in event-list and time-series format.

indication of whether the distribution has shifted.

2.2. Embeddings of time-series

The fact that there are 90 features also means that visualising all at once is difficult. An alternative is to embed the 90 features into a 2-dimensional space. We will consider several methods of embedding, always using the implementation provided by `scikit-learn` (Pedregosa et al., 2011): Principal Components Analysis (PCA), Isomap Embedding, t-Distributed Stochastic Neighbour Embedding (t-SNE), and Spectral Embedding. Figure 2 shows a “trajectory” achieved using spectral embedding.

2.3. Embeddings of event-lists

We can also embed the original event-list data into a lower-dimensional space. The large scale of the data poses a problem for all embedding methods. Therefore, we show results for just 1% of

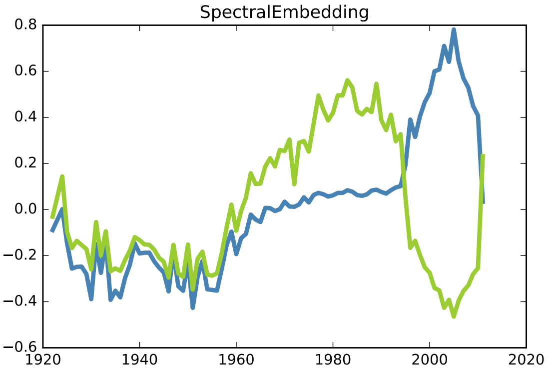
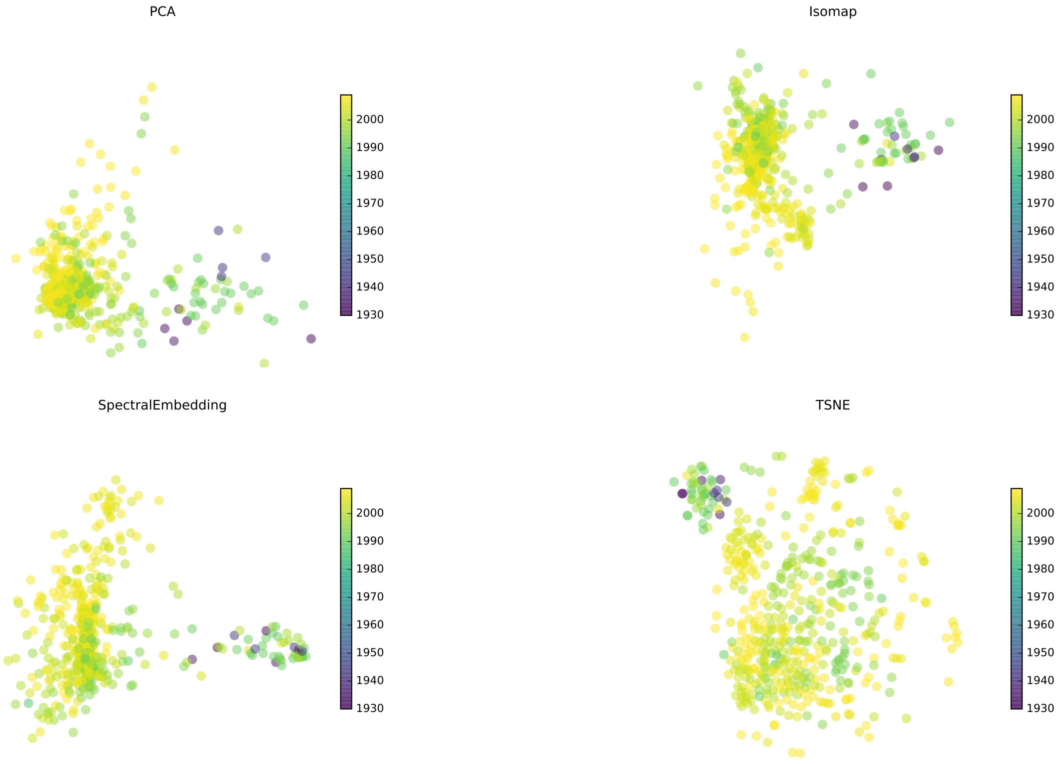


Figure 2: The 90-dimensional time-series, embedded into a two-dimensional space using spectral embedding. The two dimensions are plotted against the time axis.

the data in Fig. 3. Here, the axes represent the two abstract dimensions found by the embedding method, while time is represented by colour.


PCA

Isomap

SpectralEmbedding

TSNE

Figure 3: Embeddings of event-list data into a 2-dimensional space.

These embeddings show that in all cases, the earliest music (darkest points) tend to cluster together, indicating similar timbral features. The Spectral and t-SNE embeddings achieve a broader "spread" of the data, compared to PCA and Isomap.

3. Sonification

We next consider two methods of sonifying the data. We map the event-list data via granular synthesis to a textured, non-tonal waveform; we map the time-series data via arithmetic transforms to simple, tonal MIDI output.

3.1. Granular synthesis

In the first method of sonification, for each song in the event-list data, we add a "grain" of sound to the output, at a time corresponding to the song's year. The grain is extracted from a grain source, that is a live recording of *How Blue Can You Get* by BB King[5]. This was chosen as a classic performance by an artist of exceptional longevity – which seems appropriate for this dataset. It also contains a variety of textures (guitar, band, vocals, crowd noise, and an MC) suitable for granular synthesis.

3.2. Music as a function of time

In the second method of sonification, we divide time into short time-steps (quarter-notes, or somewhat shorter). At each time-step, we feed values of the time-series variables into the input nodes of a *directed acyclic graph*. Each of the non-input nodes executes some arithmetic function: for example, a '*' node will have exactly two incoming edges, and it will compute and output the product of the incoming values. Fig. 4 shows an example graph of typical size, with input nodes at the top.

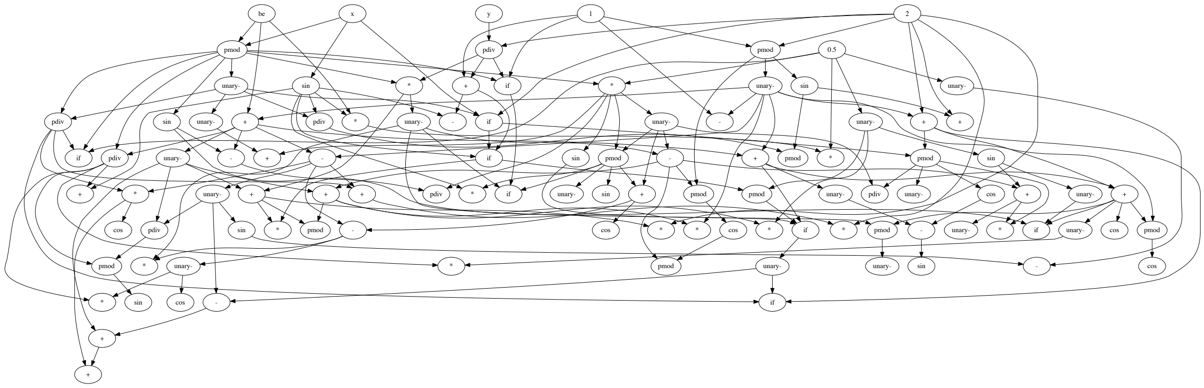
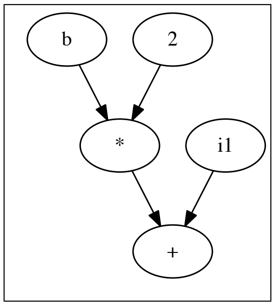


Figure 4: An example directed acyclic graph, mapping arbitrary time-series to music.

Nodes which are sufficiently far from the input (a path length of 3 or more) will additionally implement some MIDI output processing. They maintain an accumulator which represents "activity", adding an incoming value to it at each time-step. Whenever activity is above a threshold, a MIDI note-on is triggered: its volume is determined by activity, and pitch is determined by another

---

[5] URL `https://www.youtube.com/watch?v=LWLAAzOBoBI`



Time-step	0	1	2	3	4	5
3/4 beat (at b)	0	1	2	0	1	2
Data signal (at $i1$)	50	50	50	58	58	58
Output (at $+$)	50	52	54	58	60	62
Resulting pitch	D	E	F#	A#	C	D’

Table 1: Example: a fragment of a directed acyclic graph (left) and the resulting computation at several time-steps (right).

incoming value, mapped via a sigmoid mapping and a diatonic mapping. The accumulator is then drained.

In addition to the time-series data, data-streams representing musical structure are also input to the graph. For example, in Table 1, at each time-step the current position in the bar is input as variable b. As a result, the musical output has both a “bar” structure as well as being controlled, indirectly, by the time-series data. The method is partly inspired by that of Hoover et al. (2008).

3.3. Controlling the sonifications

In both sonification methods, the mapping from data to sound is controlled by many parameters. In the granular synthesis method, the extraction and playback of the grains are under the control of several numerical parameters such as start location, length, and playback rate. In the arithmetic transform method, the node types and connections between nodes must be chosen. Ideally, the user would be able to control these parameters in order to produce a sonification which achieves both the desired functionality (i.e. being informative) and desired aesthetics. However, because there are so many parameters, and they interact in complex ways, direct control is rather difficult. Instead, a method of indirect control is provided. Interactive evolutionary computation (IEC) (Takagi, 2001) is an interactive search method, inspired by Darwinian natural selection, explained in Fig. 5. More details are available in previous work (Crowley and McDermott, 2015).

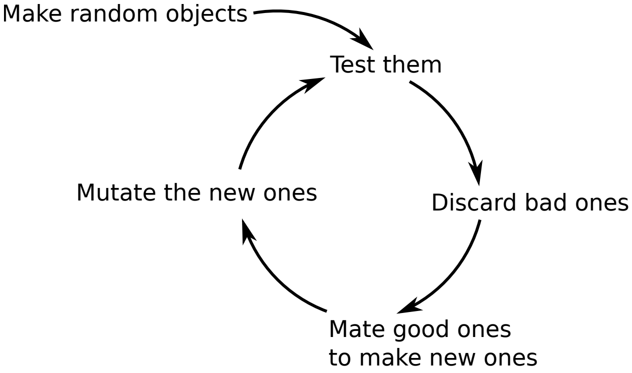


Figure 5: Evolutionary Computation. A *population* of parameter settings is initialised randomly. Each gives rise to a sonification. The user listens and discards the worst, keeping the best. The system recombines and mutates the best to produce a new population, and iterates. Over time, the population evolves to suit the user’s preferences.

3.4. Results

The result of the granular synthesis is a dense, textured, and noisy piece, in which elements of the source material are sometimes audible. A sense of increasing density is clearly audible. Different settings change the aesthetics noticeably, but maintain a sense of trajectory.

The result of the arithmetic mapping is variable, from very sparse, gestural music, to repetitive and rhythmic pieces. There is a large variety in subjective quality. The IEC method allows the user to direct search towards the desired aesthetics.

In both methods, we maintain a mapping of 1 year to 1 second, so the two results are synchronized. Because of their very different sonic qualities, it is easy to listen to both at the same time, and they balance each other's shortcomings (the texture can be "non-musical", while the tonal music can be overly simple) quite well. Sample output is available[6].

4. Conclusions

Visualisation has obvious advantages, especially for analytical settings. It can show the entire dataset at once. Most of us are more familiar with analysing visual representations of data. Powerful modern feature reduction methods such as t-SNE and spectral embedding can deal with data of moderately high dimension, bringing it down to two dimensions for screen or page, though they can struggle with datasets of many rows.

However, sonifications have their place also. The sonifications produced in this work use a direct mapping between time in the data, and time in the sound. As a result they place the user "inside" the data in a way that visualisation does not. Combining two very different sonifications takes advantage of our sophisticated source-separation abilities (cf. the cocktail party effect) to convey more information in a fixed time-frame.

One of the major challenges in sonification is how to produce aesthetically pleasing results which retain an informative structure. This work responds to this challenge by setting out two sonification methods – one well-known, the other less so – each with many parameters, which can be placed under the control of interactive evolutionary computation. These methods can be used by others – non-experts – to impose aesthetics on sonifications of arbitrary event-list and time-series data. Code is available[7].

The cultural context (the grain source and the dataset) gives the (abstract, mechanical) music a sense of "referring" to something. This is an advantage in artistic terms, and changes our perception of the sonification, even though the aesthetics themselves are unchanged.

Future work will concentrate on expanding the granular synthesis model with more user-controlled parameters, and biasing the creation of the directed acyclic graph towards more high-quality music in the early stages of search.

---

[6] URL `http://www.skynet.ie/~jmmcd/xg.html`.

[7] See `http://jmmcd.net/music.html`

References

M. L. Anderson. Embodied cognition: A field guide. *Artificial Intelligence*, 149(1):91 – 130, 2003. ISSN 0004-3702. doi: http://dx.doi.org/10.1016/S0004-3702(03)00054-7. URL `http://www.sciencedirect.com/science/article/pii/S0004370203000547`.

S. Barrass and P. Vickers. Sonification design and aesthetics. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*. Logos Verlag, Berlin, 2011.

T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

K. Crowley and J. McDermott. Mapping brain signals to music via executable graphs. In J. Timoney and T. Lysaght, editors, *Proceedings of SMC2015, the 12th Sound and Music Computing Conference*, pages 503–508, Maynooth, Ireland, July 2015. Music Technology Research Group, Maynooth University. ISBN 9-7809-92746629. poster.

T. Hermann and A. Hunt. Guest editors' introduction: An introduction to interactive sonification. *IEEE multimedia*, pages 20–24, 2005.

T. Hermann, A. Hunt, and J. G. Neuhoff. *The sonification handbook*. Logos Verlag, Berlin, 2011.

A. K. Hoover, M. P. Rosario, and K. O. Stanley. Scaffolding for interactively evolving novel drum tracks for existing songs. In *Proceedings of EvoWorkshops*, volume 4974 of *LNCS*, page 412. Springer, 2008.

G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff. Sonification report: Status of the field and research agenda. `http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1443&context=psychfacpub`, 2010.

J. McDermott and U.-M. O'Reilly. An executable graph representation for evolutionary generative music. In *GECCO '11*, Dublin, 2011.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

S. Roddy and D. Furlong. Embodied aesthetics in auditory display. *Organised Sound*, 19:70–77, 4 2014. ISSN 1469-8153. doi: 10.1017/S1355771813000423. URL `http://journals.cambridge.org/article_S1355771813000423`.

H. Takagi. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proc. of the IEEE*, 89(9):1275–1296, 2001.

T. Wishart and S. Emmerson. *On sonic art: a new and revised edition*, volume 12 of *Contemporary Music Studies*. Psychology Press, 1996.

Music Performance Science: Analytics and Music Recordings

Eamon O Doherty, MA, MSc,
DIT Conservatory of Music and Dance

Abstract This paper describes one type of analytics model for recorded music performances. It includes the derivation of an "average" performance or Performance Norm (PN), and calculation of distance matrices (DM) for a collection of recordings. It then describes some experiments using both point rates of acceleration, and of change of dynamic level, as key factors in determining expressivity and ultimately performance style. A software program was developed in R to automate calculations. This model has possible "fingerprinting" applications in music fraud detection, music information retrieval (MIR) and, perhaps most importantly, in pedagogical applications for music education.

Performance science

The nineteenth-century witnessed the primacy of western art music (WAM) *works* which are represented in the physical world as *scores*. Analysis of scores became a fundamental bedrock of musicology and of music studies generally, independently of any instantiations of works in performance. The pinnacle of such analyses is perhaps the types of quasi-mathemetical analyses by scholars such as Babbitt, Lewin and, more recently, Mazzola [Mazzola, 2011]. Even Babbitt's program notes could be impenetrable: "models of similar, interval-preserving, registrally uninterpreted pitch-class and metrically durationally uninterpreted time-point aggregate arrays".

Music Performance Science (MPS) is concerned not so much with procedures that treat scores as objects nor with aesthetic principles. It seeks after quantitative analytics of performances -- in order to fully understand performance as an art phenomenon -- the psychology of communication, movement, and even pathways to the types of injury that performers can acquire from repetitive actions [Cook, 2013]. Treating recordings as proxies for contemporaneous performance styles may help in detecting evolution in performance style over time [Fabian, 2015]. This paper addresses an idea that Eitan and Granot term *intensity contours*, referring to the ebb and flow of intensity over the life-cyle of a musical performance [Eitan and Granot, 2007]. Intensity is assumed to be a characteristic result of expressivity: which has been remarked by Cook as something that is not notated in any score. Expressivity is what a performer adds to a score to bring a performance to life and largely determines how successful is the communication with listeners [Cochrane and Rooset, 2012]. How would one go about defining the make-up of an intensity contour, of measuring it, and using a basket of such contours to compare performances?

Computation and performance variables

There are significant computational issues implied by music performance data which may involve multivariate datasets of enormous scale – the Big Data problem arriving in the music domain. A critical issue in handling large datasets is the general problem of detecting meaningless correlations or clusters, termed the *curse of dimensionality*. One might choose to apply machine learning, or other, algorithms to deal adequately with volume and complexity, assuming one can identify the important factors to measure. Another approach is to start with techniques that can reduce both the dimensionality and numerosity of the datasets. The latter approach will also help algorithms whose runtime can not be guaranteed to be a linear multiple of the volume of input data.

A starting point is to assume that any musical performance may be represented as a multivariate time-series (TS). It is not important that the constellation of all possible variables be included since it is not possible to assure such completeness. It must be assumed that some variables will be more important than others in influencing results, and that the incremental benefit of adding further variables at the margin would be small. Each performance variable may be thought of as a column of a matrix (a *column vector*), where each row represents measurements for each of those variables, taken at specific time points. Quite how to choose performance variables is beyond the scope of this paper, but many performance analyses have focussed on tempi and dynamic levels. The complete set of vectors for a performance forms a Performance Matrix (PM). The independent variables must be transformed to a common time-base. This may be achieved by moving the reference point for measurements to a beat-within-bar and by keeping track of the elapsed time for arriving at each measurement point.

Saxify is a generalised analytics model and software program, applied herein to two first-derivative variables – rate of acceleration, and rate of change of dynamic level. There is no reason to limit the model to those two variables. In general the PM can include any additional performance variable, the only constraint being that such variables may be measured at identifiable time intervals. The variables do not necessarily need to be numeric, nor even continuous. One may, for example, conceive of a scale (perhaps 1-10) to represent different degrees of articulation ranging, on piano, from perfect legato to very accented staccato. Or one might represent different bowings such as col legno, legato, pizzicato, by a numeric or symbolic scale.

SAX

Symbolic Approximation (SAX) is one technique for comparing TS, originally proposed by a computer science team at University of California, Riverside [Lin, et al., 2003]. The technique relies firstly on a quantization step – Piecewise Aggregate Approximation (PAA) – to reduce dimensionality. PAA divides any original timeseries into a smaller set of equal-sized frames. The second step, to permit comparison of TS on different scales, is to normalize the reduced dataset by subtracting the arithmetic mean and dividing by the standard deviation. This will allow us to compare series comprising different variables and is essentially a shift of the original data to a common base. The third step is to assign a symbolic alphabet (typically 4-8 symbols) to the quantized, Gaussian aggregated values in such a way as to guarantee equi-probable selection. This achieves two important results. It means that standard statistical tests can be applied. It also means that string-comparison techniques (of which Levenshtein Distance is a well-known one) can be used to provide a measure of similarity with reduced computational complexity. Each original TS is now reduced to an arbitarily small string and the strings for multiple variables may be aggregated by applying weightings. Since a set of musical performance variables comprises multiple series flowing in time, it is useful to combine them into one symbol string using weighted values from each individual series. This has achieved a significant result for musical applications. It allows us to work with much smaller data volumes, and to use proven techniques to assess relationships, such as clustering or classifiction, between performances. One symbol string might typically be :

aabbccdd

This results in a symbolic representation of the original, numeric series that may be subjected to appropriate techniques to determine similarity. The SAX technique uses the Mindist measure – which is one such way of establishing (dis)similarity between two TS.

$$MINDIST(Ss, Ts) \equiv \sqrt{\frac{n}{N}\sqrt{\sum_{i=1}^{n}(dist(s_{si}, t_{si}))^2}}$$

Where n is the length of two original time series S and T, N is the length of the strings (the number of the frames), S_s and T_s are the symbolic representations of the two time series, and where the function dist() is implemented by using the Nornal Distribution lookup table. Mindist guarantees to lower-bound other measures such as Euclidean Distance and Dynamic Time-Warping Distance.

Experimental approach

The current research project applied the model to a work by Arnold Schoenberg – his Phantasy for Violin with Piano accompaniment, Op.47. The null hypotheses at the start of the study was that performance style (represented by expressivity) did not materially change over the period of these recordings, which are treated as proxies for contemporary performance sstyles. The complete discography from 1951 (46 performances) was acquired, and converted to a common digital format. It was decided to use two higher-order variables (rate of acceleration and rate of change of dynamic level) rather then the more conventional tempo and loudness variables. The selected variables are more complex to work with but are valuable in respect of identifying more subtle point-effects. The piano part was ignored in all measurements. One interesting fact about dynamic markings on conventional WAM scores is that, unlike tempo variations which may be noted with unlimited detail, the 'hairpins' appear to imply a linear gradient. That may not be the composer's intent, nor the performer's execution, but there is no conventional indication otherwise.

Measurements were taken at beat intervals, and had to be done manually since automated beat timing software does not deal well with polyphonic scoring, nor with music where the metre changes frequently. The acceleration rate at any instant is the of change in tempo which is indicated by the slope of a tangent to a tempo curve. We can approximate the slope, S, at an intermediate point on a tempo curve (x_2,y_2) from two arbitrarily adjacent points (x_1,y_1) and (x_3,y_3) as follows:

$$S \equiv \frac{y_3 - y_1}{x_3 - x_1}$$

Performance Norm

A Performance Norm (PN) is a conceptual *average* performance for a set of performances of a single musical work. On the assumption that we can derive a PN, it can then be used to measure how far any individual performances lies from the average. The concept is relatively straightfoward. A PN is a MTS structured in the same way as any other performance, and for the same musical variables calculated at the same instants. It is created by extracting the average value of all performance values at each instant. Simple arithmetic averages are not appropriate since we are dealing with ratios (velocity) and logarithms (decibels). In the case of ratios we need to use the harmonic mean H of n values ($x_i \dots x_n$):

$$\frac{1}{H} \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}$$

where H is the Harmonic Mean, n is the number of values, and x_i is the *i*th value.

For decibels, the average value of n decibel values is:

$$L_p \equiv 10log_{10}[(\frac{1}{n}) \sum_{i=1}^{n} log^{-1}(\frac{L_{pi}}{10})]$$

where L_p is the average, n is the number of values, and L_{pi} is the ith value.

Distance Matrix

A DM shows how far each performance lies from others, including the PN. The DM is triangular (symmetric across the diagonal), and the top left to bottom right diagonal is all-zeroes. An extract from a DM is shown in Fig 1, with row titles referencing each individual recording. Values are in Mindists (which are related to standard deviations). Since Mindsit does not satisfy all the criteria for a metric, this matrix is in reality a Similarity Matrix. A DM may be used for many purposes, including performing a clustering analysis.

PN	0	3.51	2.97	5.63
1951a	3.51	0	4.97	6.78
1951b	2.97	4.97	0	6.72
1953a	5.63	6.78	6.72	0
1953b	4.22	5.42	4.83	5.37
1954	3.49	3.88	5.72	6.83
1962	3.89	4.89	4.93	6.9
1964	3.68	4.37	4.37	5.75
1965a	3.35	4.84	4.97	7.91
1965b	4.29	5.08	5.25	7.03
1966a	4.02	5.35	5.84	7.69

Fig 1. Extract from a DM

Saxify also holds a performance map, which is a section-breakdown of a piece of music specified in bars. The purpose of the performance map is to allow the program also to calculate DMs for each section of the work, across all performances, in order to achieve more fine-grained analysis. Each section is represented as an R6 object. The data stored for each of these objects are:

- Section Number (eg. 0)
- Section Name (eg. 1st Episode)
- Start Bar

- End Bar,
- Start Beat
- End Beat
- Sub-Section Name (eg. *poco meno mosso*)
- Bpm (beats per minute as indicated on score)
- Time Signature (e.g. 6/8)
- Total Beats

Results

Beat timings and dynamic levels were obtained from 44 recordings with the aid of Sonic Analyser. The resultant data were processed using software written in R. This software, with parameters to vary key factors, reads all the performance data from an Excel spreadsheet - each individual recording's data was stored on one tab of the spreadsheet (which could equally be a suitable database). Saxify converted the numeric data to SAX symbol format, and calculated a Distance Matrix (DM). PAA used 150 frames (compressing the data signals to ~1/3rd of their original size). The symbol alphabet comprised 8 symbols. The distance matrix results were plotted by year in the scatterplot shown in Fig 2. Since the plot is sensitive to framesize (shifting up/down) and to symbol alphabet (slope) we should not make too much of this plot on its own. However it is notable that the overall trend is downwards for the two separate variables as well as the combined values. The general level of dispersion tightens from left to right.

When the DMs for separate sections of the work are plotted, as in the exanple in Fig 3, the same trend persists across most of the work (barring clear outliers).

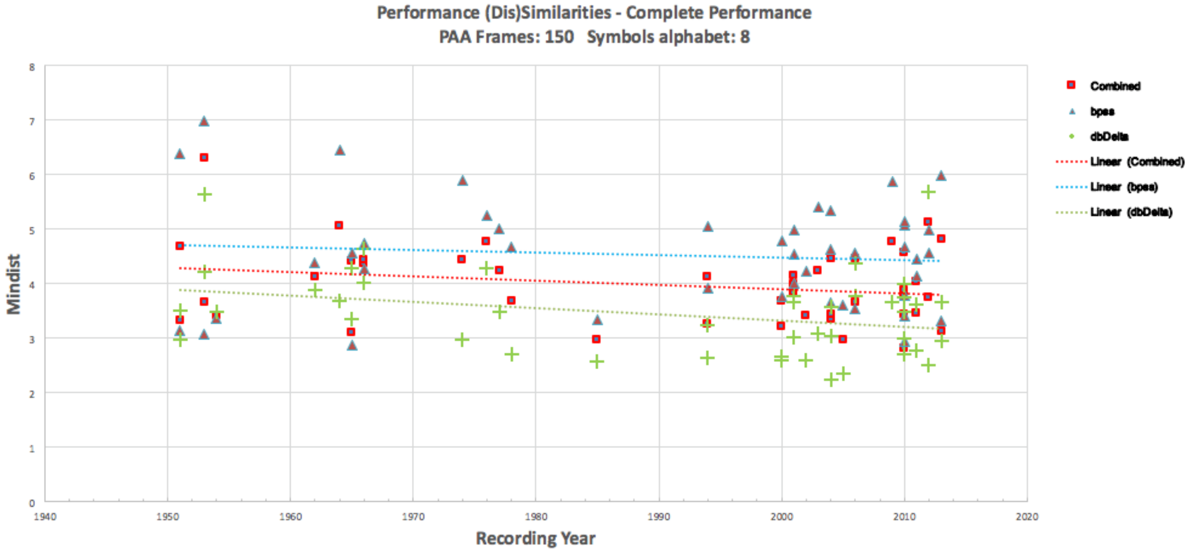


Fig. 2. Performance dissimilarities – complete work

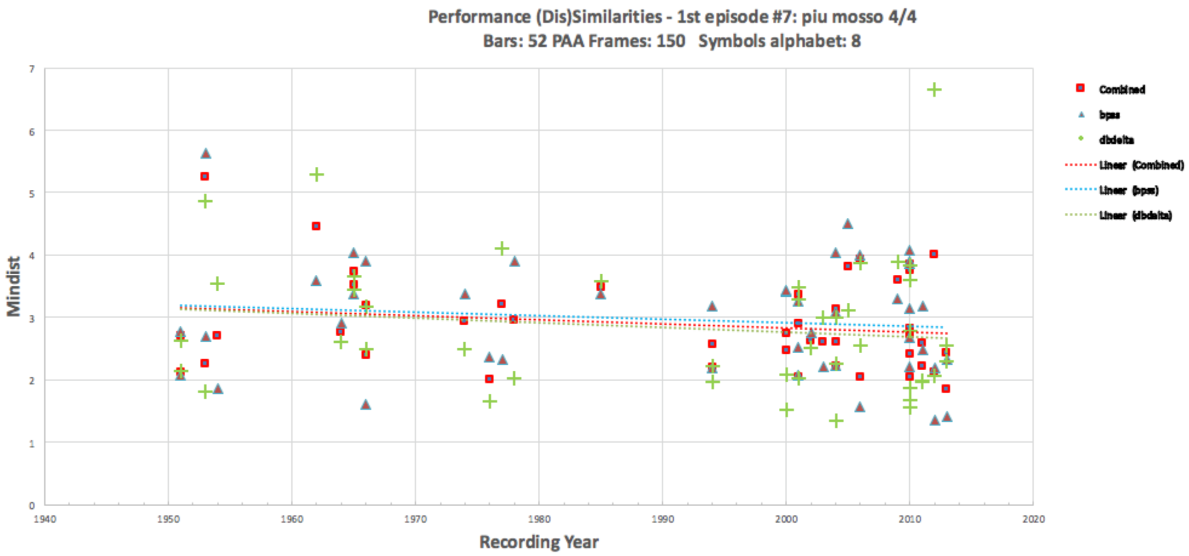


Fig 3. Performance dissimilarities – piu mosso

Concluding remarks

The two chosen factors of expressive performance indicate that performing style trended away from innovation in the earlier period and towards a long-term norm. The earlier innovations are possibly due to unfamiliarity with the music and how it should sound. The performers in the 1950/60 period, such as Menuhin, received their formative education very early in the twentieth-century and brought older performing values to this new music. Perhaps risk-taking, indicated by greater dispersion in the DM, was more prevalent early on since it would have taken time to build consensus on how to deliver effective performances of this music. It may also point to a feedback loop in that commercially successful recorded music perhaps has an influence on style and taste which dictates to what and how audiences wish to listen.

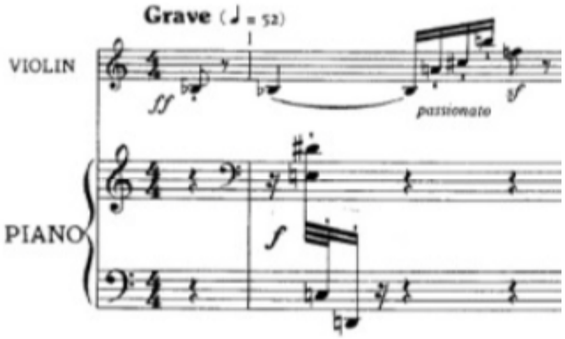



Fig 4. Opening bar showing two motifs

Finally, having listened to all of the recordings repeatedly and exhaustively the two key motifs (Fig. 4) of the Phantasy suggest something: the short-long, short-short-short rhythmic patterns in the violin part, and repeated many times throughout the work, are the Morse Code dot-dash dot-dot-dot representations of ‘*A S’* (for Arnold Schoenberg).

References

Cook, N. (2013). *Beyond the Score: Music as Performance*, Oxford: Oxford University Press

Cochrane, T, and Rosset, O. (2012). ‘Expressive Dimensions in Music’, in *Proceedings of CMMR 2012* (presented at the 9th International Symposium on Computer Music Modeling and Retrieval, Queen Mary University, London).

Eitan, Z. & Granot, R. (2007). 'Intensity changes and perceived similarity: Inter-parametric analogies'. *Musicae Scientiae*, Discussion Forum 4A.

Fabian, D. (2015). *A Musicology of Performance* (http://www.openbookpublishers.com: OpenBook Publishers, 2015) http://dx.doi.org/10.11647/.

Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). 'A Symbolic Representation of Time Series, with Implications for Streaming Algorithms', *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego: ACM.

Kozinn, A. (2011). ‘Milton Babbitt, a composer who gloried in complexity, Dies at 94’, New York Times, (Jan 29, 2011).

Mazzola, G. (2011). *Musical Performance: A Comprehensive Approach: Theory, Analytical Tools, and Case Studies* (Heidelberg: Springer).

Cloud Chamber Music: an installation utilising computer vision to tie music making to sub atomic particles

Thomas Power
Music & Media Technology, Trinity College Dublin

Synopsis

This project aims to tie composition directly to the fundamental particles in our universe. To unite the worlds of particle physics and music. The cloud chamber will be used as a source of music generation using a computer vision system. The visible trails that ionized particles leave in the cloud chamber will be used to trigger notes in a compositional framework giving a musical voice to one of the most fascinating physical phenomena available to the naked eye.

Description

The cloud chamber has already contributed so much to the world of particle physics. From Wilson's first experiments to its role in the discovery of the positron, muon and kaon. It is possibly one of the most important discoveries humans have made and without it the LHC in CERN would perhaps never have been conceived. This project aims to explore the cloud chamber's musical potential. The creation of music already relies so heavily on physical phenomena why not tie our music making directly to subatomic particles? I believe that the cloud chamber has much to offer us musically. It offers us another way of experiencing the connection with the subatomic particles that it so beautifully visualises.

The aim of the project is to use a computer vision framework to achieve this. Computer vision will be used to extract musically relevant data from the visible trails that ionised particles leave in the cloud chamber. The result will be a aleatoric music system that derives its performance information from the radioactive decay of particles. Expanding John Cage's ideas of indeterminacy to a fundamental physical process.

A performance piece involving cloud chamber and violin has already been achieved by Alexis Kirke from Plymouth University. His project illustrates the potential of the cloud chamber as a music device. A paper by Bob Sturm from CCRMA in Stanford on "Composing for an Ensemble of Atoms" comments very fittingly, "*It is only a matter of time before a real system of particles could be used—the particle accelerator will become an expressive musical instrument, and the particle physicist will become the composer/scientist.*"

The desired outcome of this project is to convey an experience to the audience. A visceral experience of the unseen forces in our universe. The public it seems are far removed from connecting with the world of particle physics. The aim of this project then is to provide that connection through the medium of music.

Prototype Links

`https://www.youtube.com/watch?v=U6HY2uyYQ0w`
`https://www.youtube.com/watch?v=EMqx9qjQkfA`

Sounding Human with Data: The Role of Embodied Conceptual Metaphors and Aesthetics in Representing and Exploring Data Sets

Stephen Roddy
CONNECT, Centre for Future
Networks and Communications,
Trinity College Dublin, Ireland
`roddyst@tcd.ie`

Brian Bridges
School of Creative Arts and Technologies,
Ulster University, Magee campus
`bd.bridges@ulster.ac.uk`

Introduction

Auditory display is the use of sound to present information to a listener. *Sonification* is a particular type of auditory display technique in which data is mapped to non-speech sound to communicate information about its source to a listener. Sonification generally aims to leverage the temporal and frequency resolution of the human ear and is a useful technique for representing data that cannot be represented by visual means alone. Taking this perspective as our point of departure, we believe that sonification may benefit from being informed by aesthetic explorations and academic developments within the wider fields of music technology, electronic music and sonic arts. In this paper, we will seek to explore areas of common ground between sonification and electronic music/sonic arts using unifying frameworks derived from musical aesthetics and embodied cognitive science (Kendall, 2014; Lakoff & Johnson, 1999).

Sonification techniques have been applied across a wide range of contexts including the presentation of information to the visually impaired (Yoshida et al., 2011), process monitoring for business and industry (Vickers, 2011), medical applications (Ballora et al., 2004), human computer interfaces (Brewster, 1994), to supplement or replace visual displays (Fitch & Kramer, 1994), exploratory data analysis (Hermann & Ritter, 1999) and, most importantly for the current milieu, to reveal the invisible data flows of smart cities and the internet of things (Rimland *et al.*, 2013; Lockton et al., 2014). The use of sonification as a broad and inclusive aesthetic practice and cultural medium for sharing, using and enjoying information is discussed by Barrass (2012). As networked smart societies grow in size and becomes increasingly complex the ubiquitous invisible data flows upon which these societies run are becoming hard to monitor and understand by visual means alone. Sonification might provide a means by which these invisible data flows can be monitored and understood.

In order to achieve this type of usage, sonification solutions need to be applicable to and intelligible to an audience of general listeners. This requires a universal shared context by which sonifications can be interpreted. Embodied cognition researchers argue that the shared physical features of the human body, and the capacities and actions which our bodies afford us, define and specify mid-level structures of human cognitive processing, providing shared contexts by which people can interpret meaning in and assign meaning to their worlds (Lakoff and Johnson 1980; 1999; Varela et al., 1991). At present, embodied perspectives on cognition are infrequently explored in auditory display research, which tends to focus on either higher level processing in terms of language and semiotics (Vickers, 2012) or lower level processing in terms of psychoacoustics and Auditory Scene Analysis (Carlile, 2011).

Sonification as Data-Framing: Structural, Narrative, Aesthetic.

Broadly speaking, sonification is useful when representing/investigating data sets which embody significant temporal variation. Sonification gives access to a temporal evolution (and can speed up or slow down temporal processes, depending on the nature of the data sampling). This temporal variation implies narrative and causality; the mapping of data to temporally–evolving sound may reveal significant events through audible significant deviations within the sound. In this way, sonification has been used within the context of data analytics using data from business (Worrall, 2009) or socioeconomic data and may be helpful in searches for patterns within data which may not be as apparent using visual representational strategies. The temporal and frequency sensitivity of the auditory system may tend to make sudden (transient) differences in the formal structure of incoming signals quite obvious, as long as basic sensitivity/just noticeable difference (jnd) ranges are taken into account (or various temporal/frequency scale mappings are investigated).

Fitch and Kramer (1994) provide an example of a rich mapping strategy for an auditory display that is designed to help users monitor and respond to medical complications across eight physiological variables of a digital patient. Two physiological variables are mapped to control sounds to which they resemble. Heart rate was mapped to a rhythmic thudding sound and breathing rate was mapped to a breath like sound. Atrio-ventricular dissociation and fibrillation were mapped to modulate the heart rate sounds in the same way these factors modulate a heartbeat in the real world. These mappings leveraged the users previous knowledge embodied knowledge of human anatomy. Four other mappings, body temperature to a filter applied to the heart beat sound, blood pressure to the pitch of the heart sound, brightness of the heat sound to CO_2 level and pupillary reflex to a high pitched tone, were abstract and so required learning on the part of the listener. Empirical evaluation showed the auditory display system to be more effective than a visual display for helping users monitor and respond to changes in the condition of the digital patient.

Sonification practices are prevalent throughout science and the arts. This mixed appeal has led to its use as a public outreach tool for popularising scientific research (Supper, 2014). The search for the Higgs Boson at CERN’s Large Hadron Collider installation in Geneva is embracing sonification as a means of public outreach[1]. Researchers at NASA have are also making extensive use of sonification as a public outreach tool[2]. They have also made new discoveries using sonification, which would not have been possible through visual media alone[3]. Researchers at the Climate Change Research Center used sonification techniques to create The Climate Change Symphony in order to communicate climate change data to the public[4].

Sonification is emerging as a critical technique for understanding and communicating large complex data flows in the context of increasingly networked and data-driven societies (Rimland et al, 2015; Hermann and Hunt, 2005). This has resulted in increased interest in sonification as a tool for observing network activity patterns and monitoring network security, as evidenced by a number of notable contemporary projects (see Worrall, 2015; Wolf and Fiebrink, 2013; Fairfax et al, 2014). A number of artists have used sonification techniques to reveal the hidden data

---

[1] http://tedxtalks.ted.com/video/Listening-to-data-from-the-Larg

[2] https://www.youtube.com/watch?v=kcqiLvHiACQ

[3] http://www.economist.com/news/science-and-technology/21694992-scientific-data-might-be-filled-important-things-waiting-be-discovered

[4] http://www.scidev.net/global/data/multimedia/climate-symphony-data-sound.html

flows of smart cities and the IoT. Kasper Fangal Skov's Sonic Particles 2.0 project sonifies data provided by smart sensors placed in major cities for Data Canvas' Sense Your City competition[5]. Stanza, a UK based sound artist, makes extensive use of environmental sensor data in his artistic sonification practices[6]. The Social City Detector project and the Citigram project (Park et al, 2013) used sonification to integrate the digital and physical layers of the city by making social data visible through sound[7]. The Phantom Terrains project used a repurposed hearing aid to reveal the electromagnetic signals of the wireless networks, which pervade the contemporary built environment.[8] Composers Natasha Barret and Andrea Polli make extensive use of the sonification of environmental data in their compositional practices, often with the activist intent of raising awareness about important environmental issues (see Barret and Mair, 2014; Polli 2012).

Beyond the arts, technology researchers are also examining the use of sonification practices to help reveal, analyse and understand the rich data flows of the IoT. Eva Sjuve's (2015) Metopia is a research project that explores sonification as a medium for representing data in the context of the IoT and Big Data. The Sound of Things project aims to add sound to the IoT. It also reveals the invisible IoT networks of smart cities through novel applications including agogic maps[9], geo located tweet sonificaiton[10]. A number of hardware applications that generate live sonified sound streams when physically attached to IoT devices have also emerged in recent years (Barrass and Barrass, 2013; Lockton et al., 2014).

Barrass (2012) discusses how the current aesthetic turn in the field is driving the adoption of sonification as a mass cultural medium by making sonification more appealing to the listener. Rimland et al (Rimland et al, 2015) discusses how as societies become increasingly networked and the IoT grows in size and complexity sonification will be needed as a means of making sense of the complex data flows that can no longer be effectively understood by visual means alone. Listeners will turn to sonification for enjoyment, aesthetic appreciation and to learn about the data sources represented therein. In the coming years sonification will become a popular method by which the invisible data flows of smart cities and smart environments are revealed and by which the digital environment is manifest in the physical environment.

What some of these applications illustrate is that sonification may bring with it a consideration of aesthetics; how data may be rendered in sound such that its structure is not only accessible, but 'attractive'/'engaging', perhaps even 'beautiful'…certainly, sufficiently engaging to hold interest over longer time-spans spent interrogating data sets in this way. More broadly, the sense of narrative causality and dynamism within sonification makes it an emotive technique; data may become more 'impactful', less 'neutral', when perceptualised. The temporal evolution and frequency variation of sonification may be seen as corresponding to a basic model of emotion (arousal/valence), which has previously been identified as one underpinning music's emotional efficacy (Huron, 2006). Technology artist Luke Dubois[11] has made the point that music is an 'emotional technology', a data–art…the similarity relationship between sound–data–art

---

[5] http://www.kasperskov-audiodesign.dk/projects_SonicParticles2.html

[6] http://www.stanza.co.uk/emergentcity/?tag=sonification

[7] http://thecreatorsproject.vice.com/the-makers-series/the-makers-series-social-city-detector

[8] http://phantomterrains.com/

[9] http://www.soundofthings.org/Proj/AgogicMaps.html

[10] http://www.soundofthings.org/Proj/SonifyTweet.html

[11] https://www.ted.com/talks/r_luke_dubois_insightful_human_portraits_made_from_data?language=en

(sonification) and music may also run in the opposite direction! As such, both (a) basic structural framing (accessible to auditory parsing processes), and (b) narrative/affective qualities, may be relevant considerations when exploring sonification strategies and may be unified within an aesthetic domain.

Care must be taken when considering aesthetics in a sonification context and to this end it is useful to draw a sharp distinction between aesthetic (for structural ends) and cosmetic sonification practices. Cosmetic sonification practices simply aim to produce an attractive and easily listenable sonic result while aesthetic sonification practices aims to frame and shape the qualities of the listeners' sonic experience as a means of communicating information about a data source. While it is important that a sonification should sound pleasing and be easy to listen to, especially if a listener is expected to attend to it repeatedly or for an extended period of time, aesthetic considerations reach far beyond this concern to the framing and shaping of the listeners very experience of the sonification, and so their understanding of the original data. It has been argued that the aesthetic dimensions of sound are those best suited to the communication of data in a sonification context (Roddy & Furlong 2014; Roddy, 2015).

Even if some sort of narrative of dynamic change is obvious via the audible changes within a sonification 'signal', does that render the sonification inherently meaningful? How might abstract data be converted to sound such that the structure of the data is accessible to interpretation? There is no uniform consensus in terms of strategies; sonification designers must currently answer questions of *mapping* (transposing from data structures to sound parameters) during the design phase (Flowers, 2005). Successful sonification is not simply a straightforward case of mapping data to isomorphic sound parameters (an approach which is basically formalist; one which found favour in early explorations of auditory display techniques). Listeners may not always be able to easily interpret sounds in the absence of consideration of certain perceptual–conceptual predispositions. For example, certain dimensions within a sonification, whilst isomorphic in terms of a dataset, may appear 'counter–intuitive'; for example, an increase in a dimension could be represented by a similar change in magnitude of the frequency of a tone, but if the resulting parametric change were a *decrease* in frequency (albeit with similar magnitude), the polarity of the sonification might appear to be reversed. This phenomena is explored in depth by Walker (2000) and Roddy (2015).

The Mapping Problem and the Dimensions of Sound

The mapping problem represents a significant open problem within the field of auditory display (Worrall 2009). It was first introduced by Flowers (2005) who criticised the central claim of auditory display research: that submitting the contents of complex data sets to sonification will necessarily lead to the emergence of meaningful relationships in the data. In reality, this is rarely the case. It has been argued that the dominant conceptualisation of sound in the West is tailored to describing the physical acoustic waveform and its perceptual correlates but that it cannot adequately account for how sound communicates information to a listener (Truax, 1984). An analogous argument has been made about Western art music, which reduces the rich multi-dimensional spectra of musical and sonic possibilities to just three primary dimensions (pitch, duration and timbre) which can be easily represented on a written score. Worrall (2010) argues that this reductive approach to music is informed by the computationalist theory of mind and auditory display researchers often impose these same limits upon their own work by using the music and sound synthesis software developed for these paradigms, failing to account for the role of the embodied performer and the perceptual and cognitive configuration of the embodied listener. The mapping problem may be seen as the result of a tendency amongst auditory display researchers to adopt the acoustic/psychoacoustic and Western art music paradigm when specifying sound, thus imposing a set limits on how sound can be conceptualised, parameterised and used to communicate data to a listener. This *psychoacoustic paradigm* can result in auditory display solutions that are not designed to exploit the full range of communicative dimensions provided by sound and music and which do not account for the perceptual and cognitive faculties of the listener.

This issue is exemplified in one of the complications resulting form the mapping problem: *dimensional entanglement*. This is the intermingling of auditory dimensions traditionally assumed to be separable within the computationalist framework. For example, in *PMSon* pitch, loudness, duration and timbre are often mapped to unique data (see Grond and Berger, 2011). However, these dimensions are not perceived independently of one another but are perceived as individual aspects of larger sonic wholes. They are integrated and changes in one dimension can cause changes in another making it confusing for the listener to interpret a *PMSon* sonification during listening (Peres and Lane, 2005; Worrall, 2010; Peres, 2012). Ideas of discrete sonic dimensions such as timbre, pitch and amplitude have little to do with the listeners' everyday experience of sound. From the perspective of embodied cognition and the ecologically–grounded theories of perception which have influenced it, these are not meaningful dimensions along which to sonify data. Mapping data to such parameters is not *sonification* but *acoustification*, the straightforward mapping of data to acoustic features. This results in mapping strategies that seeks to communicate data to a listener by means of acoustic symbols that are seen to be made meaningful through the application of some internal set of syntactical rules in the mind (cognitive grammars). Ryle (1949) and Searle (1980) and Harnad (1991) have variously shown that meaning cannot be generated for a listener in this way because, as Dreyfuss (1965) and Polyani (1966) point out, objects of meaning require a background context against which their meaning can be assigned, and the act of processing information provides no such context.

However, contemporary sonification research and practices seek to solve the mapping problem whilst preserving the structure of the data during encoding so that it can be accurately represented to a listener. Much contemporary sonification takes place in a context which recognises *ecological* perspectives on the problem; that the very act of encoding data into sound may also allow for aspects of its structure to be revealed via a listener's engagement (Walker & Kramer, 2004). Ecological, in this sense, situates the problems of perception within its inter-dependent relationship with the sensory environment. In this context, the *perceptualization* (Hermann, Hunt, Neuhoff, 2011) of data is important because different sensory modalities may be useful for revealing different aspects of data structures. The heuristic processes of parsing environmental sound and music (Bregman, 1990) may therefore be particularly helpful in a search for meaningful patterns within data sets.

From this contemporary perspective, sonification may be seen as a structural investigation *at both the encoding and decoding stages*. Whilst much attention has previously been focused on strategies for *encoding* (also referred to here as *mapping*), we believe that considering these alongside frameworks for *decoding*, based on contemporary theories of embodied auditory perception and cognition may be of great significance for improving the efficacy of sonification. While ecological approaches to sonification have been explored by a number of researchers in the field, the current paper is concerned with approaches informed by embodied cognitive science and musical aesthetics.

An additional problem with formalist auditory display/sonification approach is its treatment of complexity. The discretized treatment of these materials also conforms to a computational–formalist information processing paradigm (the Shannon–Weaver model of communication (Shannon & Weaver, 1949) which considers (discrete) channels/dimensions with particular bandwidths, influenced by external sources of noise (via signal interference, or, from some broader perspectives, a lack of contextualization allowing reconstruction of an ambiguous communication). A formalist mapping approach based on the assumption of discrete dimensions may be viewed as encompassing such a set of discrete channels, with the idea of additional noise/ambiguity which that entails. An early commentary on the channel capacity/mapping problem is to be found in George Miller's[12] commentary on human information–processing and recognition abilities via working memory, 'The Magical Number Seven, Plus or Minus Two' (Miller, 1956). Miller's work was informed by an analysis of Pollack's (1952) early auditory display studies which showed a communicative potential of c. 2.5 bits (6 elements/values channel capacity) for *unidimensional*, i.e. single-modality, stimuli. Miller analyses the experimental results of various contemporaries who were investigating auditory perception as an information processing task to conclude that, for *unidimensional judgments,* working memory, the kind of memory responsible for the short-term holding and processing of information, has a capacity of 7 ± 2 chunks, or discrete objects, of information. He argues that a listener can identify and remember roughly 7 ± 2 distinct pitches, amplitudes, rates of interruption, on-time fraction, durations, and spatial locations when presented as domains for representing data. Miller also noted that information capacity could be increased through additional parsing known as *chunking*; items to be remembered could be associated with one another, freeing capacity in formal, working memory.

This case is reserved for high–level cognition of precise rank-ordering relationships, etc. Although it may be tempting to take this as a hard–and–fast rule (within its own somewhat imprecise boundaries), it should be noted that Miller states clearly that this only holds true for unidimensional stimulus cases and cannot describe the parsing of multidimensional stimuli e.g. facial recognition, or, indeed, more complex musical cases. For example, even in more 'unnatural' formalist multidimensional cases, such as integrated frequency and amplitude displays (Pollack, 1953), Miller comments that capacity has increased beyond the 7+-/-2, albeit not in an obvious pattern noting the complexity of music[13]. Our own perspective is that such a comparably unpredictable increase in capacity is due to ecological concerns.

Although the psychoacoustic paradigm expanded to encompass ecological psychoacoustics in auditory display research (Walker and Kramer, 2004), the methods employed in this framework draw heavily from the *operationalist* framework developed by Stanley S. Stevens, the founder of psychoacoustics. Operationalism is a form of positivism which holds that a concept that cannot be reduced to a measurement is meaningless. Stevens developed his cross-modal matching and magnitude estimation techniques in order to reduce psychophysical information, e.g. heard sounds, to simple measurements. However as Dreyfus (1965), Searle (1980), Johnson (1987) and Polyani (1966) point out a large spectrum of human experience and human knowledge cannot be

---

[12] Godøy *et al.* (2010) argues that coarticulation, the fusing together of smaller units into larger wholes, is critical to Miller's theory of information chunking. While the miller limit might define the mechanical limits of perception, at the level of musical/sonic experience the information in the chunks is co-articulated so that the listener can experience rich musical imagery and gestures.

[13] Pollack (1954) saw 6 different acoustic variables of 5 different values each, yielding 7.2 bits (c.150 values), a value which may be seen closer to real-world cases of cognitive–perceptual recognition abilities and closer to what might be viewed as reasonable 'channel capacities' for music (note how close this figure is to the 7-bit range of the MIDI (Musical Instrument Digital Interface) standard!

reduced to simple measurements, an issue which Miller's (1956) account of memory and information processing grapples with even as it seeks ways to quantify its capacity.

It is in these contexts that we argue that one crucial avenue to consider in the development of sonification is an ecologically–considered model of auditory dimensions which is meaningful to a listener as it aligns with the perception and interpretation strategies which we use within our everyday environmental experience.

For more richly perceptualised sonification, as opposed to the narrower discretised signals of auditory display, we are not concerned with chunking, per se, but with pattern and correlation recognition, an entirely different problem within a perceptual, rather than formalist, context. Leveraging our (integrative) cognitive–perceptual parsing systems may help us identify meaningful patterns within multiparametric data (if degree of change is relatively constrained). It is in this context that we contend that an ecological perspective on the mapping problem is crucial in certain contexts: those where the data is mapped for clear communication; for example, for didactic purposes. In this context, our interpretative framing of multiparametric data may be aided by a consideration of models (schemas) derived from common perceptual experience cases. These may, in part, be explained as the 'environmental regularities' which underpin heuristic principles within our perceptual processes. But these regularities are more basic structural framing principles rather than necessarily supporting *interpretative framing*. To consider how interpretative framing beyond the basics happens, we may need to consider how our experience of the environment impacts upon our conceptual systems.

Sense-making in Sound from Perception to Cognition

To restate our perspective on sonification, it is our contention that sonification is not just rendering, but also the act of framing data as it is 'filtered' through our perceptual transduction and expectancy schemas; our 'sense-making' apparatus. This *perceptualization* (Hermann, Hunt, Neuhoff, 2011) is thus more than just a mapping from one domain to another, but also entails an act of structural framing which derives from our perceptual systems. As such, approaching sonification simply from the perspective of formalism as opposed to being *perceptually and cognitively informed* may impede its ability to meaningfully represent data (Worrall has compared with serial music 'problem', see also (Lerdahl, 1988, McAdams, 1987).

Perceptualization involves not only leveraging familiar environmental contexts but also leveraging from embodied contexts in ways that are compatible with the faculties and processes of embodied cognition. Patterns within complex data may be revealed by rich mappings which are engaging enough to support careful listening and communicative enough to furnish the listener with the required information. But, beyond even these concerns, there is the question of how conceptual framings arise out of the basic perceptual–structural framings through which we approach sonification. Theories of embodied cognition place the emergent patterns of experience that arise in the interaction between structural regularities of the environment and the structural regularities of the human body at the center of the 'problem' of conceptualisation.

Electroacoustic Music Theory, Embodied Cognition and Sonification

Embodied cognitive science examines the relationship between the body and mind with a specific focus on how the physical and perceptual dimensions of the human body shape and define the cognitive processes and conceptual systems of the human mind. It emerged in the late 20th century as researchers began to study emotion, culture and aesthetic experience. It has shaped the development of a number of important disciplines related to sonification research and practice, e.g. computer science, artificial intelligence and human computer interaction (Brooks, 2003; Dourish, 2004; Imaz and Benyon, 2007), computer music (Leman, 2008; Klemmer *et al.*, 2006), cognitive sciences (Varela *et al.*, 1991), visual perception (Noë, 2009), aesthetics (Johnson, 2013), music (Godøy, 2006;, 2005; Brower, 2000; Larson, 2010; Cox, 2001), linguistics and philosophy (Lakoff and Johnson, 1999).

Embodied cognition researchers have presented a number of theoretical models which describe how the embodied mind perceives meaning in and applies meaning to it's world. Image schemas were first introduced by Lakoff and Johnson (1987) and can be thought of as 'gestures of thought' in that they are basic gestural patterns derived from sensorimotor experience which we draw upon to structure our thinking and conceptual systems. The process by which these basic patterns of experience are imported into cognition is referred to as conceptual metaphor (Lakoff and Johnson 1980). The process by which multiple mental spaces are integrated to create new mental content s referred to as conceptual blending (Fauconnier and Turner 2002). A number of researchers have described musical listening and composition in terms of embodied schemata, conceptual metaphors and conceptual blends: Kendall (2014), Cox (2000), Brower (2000); Adlington (2003) Godøy (2003; 2006) Wilkie et. al. (2010). These thinkers make the argument that the embodied components of cognition represented in these theoretical models play a key role in the listener's experience of music. In electroacoustic music theory, Spectromorphology is a descriptive framework for electroacoustic music consisting of detailed categorisation schemes deriving from basic gestural shapes called primal gestures that are extended to add a meaningful low-level organisational structure to musical domains (Smalley, 1997).

Sound(ing) Schemas and Embodied Functions in Electronic/Electroacoustic Music

Previous work by one of the authors (Graham and Bridges 2014a; 2014b) has investigated points of compatibility between Smalley's spectromorphology and the embodied image schema theory of Lakoff and Johnson. They argue (*ibid.*) that Smalley's gestural surrogacy and the dimensions of his gestures are compatible with image schema theory and its extension by Johnson (2007) in terms of *qualitative dimensions of movement* (essentially, gestural details of more regular or chaotic behaviour which alter some of the contours within image schemas).

Particular points of comparison between Smalley (1997) and Lakoff and Johnson's work, especially its extension by Johnson (2007):

- **General:** environmental models/embodied–ecological models of causality, ideas of musical 'forces' based on environmental analogues
- **General:** idea of 'embodied functional associations' of particular movements
- **Specific:** image schema structures (cycles, verticality, source–path–goal, container) identifiable within spectromorphologies
- **Specific:** dimensions of Smalley's sound gestures are similar to Johnson's qualitative dimensions of movement

The similarity between the dimensions of Smalley's sound gestures (termed *energy–motion* profiles) and Johnson's *qualitative dimensions of movement* can be seen below (Bridges and Graham, 2014a).

Johnson (2007)	Embodied Association	Smalley (1997)
Tension	Rate–effort=>overcoming inertia	*Motion rootedness*
Projection	Sudden rate-change / transient movement	*Motion launching*
Linearity	Coherence of path	*Contour energy/inflection*

Table 1.1 Embodiment and Spectromorphology (Bridges and Graham, 2014a)

Not only are these dimensions of movement similar in terms of the division of embodied associations, but they also relate closely to the basic schematic forms of verticality and source–path goal. Smalley's logic of environmental causality sees certain sound gestures as embodying more rootedness or dynamism (projection/motion launching). We believe that these ideas of certain timbres as providing temporal structural dynamics may help us to develop more convincing 'sonic narratives' using data if theories such as spectromorphology and exploratory practices within electronic music can help to move us in the direction of embodied theories of musical timbre.

This emphasis on physicality within novel musical structures is also to be found beyond the world of electronic and electroacoustic music. Whilst common practice tonal music could be said to be based on structures explicable via the metaphor *Music–as–Moving–Force* (Johnson, 2007), Adlington (2003) explores image schema theory and contemporary music from the perspective that salient metaphors may relate more to ideas of changes of material and changes of physical state. The key developmental aspect which this highlights for our present purposes is that sonic 'image' schemas may be best viewed as temporally dynamic and morphologically/structurally plastic.

There is much still to explore in terms of how the specific domains of sound and music can be addressed via image schema theory. The common auditory–perceptual affordances of stream segregation and integration (Bregman, 1990) and the material metaphors of *glitch/rupture*, stretching and *bouncing/inertial effects* which are observable in a variety of contemporary electronic musical processes have the potential to be useful in sonification mappings (indeed, where these configurations occur unintentionally within existing sonifications, they may already act as clues to significant elements within the data). Sound's perceptual–ecological *interpretative frames* (contextual framing) occurs within commonplace perceptual experience due to the alignment of perceptual–heuristic processes with 'environmental regularities' (Bregman, 1993). Combining these inbuilt dynamics with more attention to potential embodied timbral/textural mappings could lead to a much more sophisticated integrating approach which avoids the obscuring of meaningful sonic dimensions behind inappropriate formal models. Exploratory creative processes which investigate embodied mapping strategies may help to suggest further avenues for the development of accessible sonic mappings.

A Consideration of *The Human Cost*: A Data-driven Composition Using Embodied Sonifiction Techniques

The Human Cost is a piece of data driven music composed by one of the authors (see Roddy, 2015), in which principles from embodied cognitive science are applied to organise mapping strategies from data to sound in a sonification context. Some of the embodied dimensions of this piece are considered in this section. The piece is intended to communicate a sense of the human cost of Ireland's economic recession.

The Human Cost was motivated by Smalley's statement that "in electroacoustic music the voice always announces a human presence" (Smalley, 1996). It was thought that as result the human voice might prove effective for representing data which measured the lives and behaviors of people. As such it was decided to sonify socioeconomic data sets from the period of Ireland's economic crash and recession. Deprivation, unemployment, emigration and GNP rates in Ireland from 2007 to 2012 were chosen as data sets for sonification. Rich multi-layered mapping strategies were employed to sonify this data. GNP is mapped to control a parameterised sounding object intending to reflect the sound of a heartbeat as GNP falls the heartbeat slows and as it rises the heartbeat increases. The choice of the heartbeat sound to represent GNP data was informed by White (2003) who argues that the economy is often conceptualised as a living organism. Deprivation, unemployment and emigration are mapped to control the prosodic features of three synthesised vocal simulations. The simulation to which the emigration rate was mapped acts as a "lead line" and the pitch and prosodic content of the vocal gesture are modulated to imitate the kinds of structure found in the old Irish laments, a type of song sang at a wake, a kind of funeral celebration which was often held to honour either a deceased relative, or a relative who was emigrating with no prospect of return.

Laments represent a cultural connection with the historical (and contemporary) experience of the emigration of the Irish Diaspora, cultural forms in which the singer's personal experiences of the passing or emigration of a loved one are expressed and communicated through vocal gesture. This transduction of human experience to physical, sound-producing gestures represented a useful physical-emotive mapping of relevance to the data sonified. The data is mapped so that the lead voice takes the foreground while the other two voices present a form of backing and the heartbeat performs a grounding percussive role in the piece. Deprivation rate and unemployment are mapped to these backing voices. Data is mapped to control the vowel shape in each vocal simulation so that as the economy worsens the open vowel sounds shift to closed vowel sounds to communicate a sense of tension. It is also mapped to spatial parameters so that both vocal simulations move through space and as the data increases and decreases the speed at which they move through space also increases and decreases creating a sense of frenzy in the piece as the economy crashes!

Conclusion

There is more to sound and music than pitches, durations, timbres and amplitudes. Sound is a powerful medium for the representation of data precisely because of its communicative dimensions; some of which are unaccounted for in standard sonic models based on discrete dimensions and parameterisation.

We have argued that a new conceptual model and specification of sound which recognises the embodied and aesthetic dimensions of sound is crucial to the development of effective data to sound mapping strategies. If sonification involves the mapping of data to sound for the purposes of communicating information about a data-source, this necessarily re-frames the information in the data in terms of the embodied and aesthetic and dimensions (and dimensional integrating effects) of the chosen sound materials. Such a process of reframing has the potential to integrate insights from a diverse range of sonic practices and theories, from embodied cognitive science and ecological psychology to electronic/electroacoustic music composition and production.

Framing of this nature can be used to ensure that sonification mapping strategies are a good fit for the listener's cognitive meaning-making faculties, thus supporting the efficient communication of the data. They can also be used to explore emotional and affective dimensions to a sonification, thus presenting a richer representation of the data than would otherwise be possible. The development of sonification within this context is best seen as an integration of the arts and sciences as their interests intersect within the spheres of sound, perception and meaning–making.

Acknowledgements

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

References

Adlington, R. (2003). Moving beyond motion: Metaphors for changing sound. *Journal of the Royal Musical Association*, *128*(2), 297-318.

Ballora, M., Pennycook, B., Ivanov, P. C., Glass, L., & Goldberger, A. L. (2004). Heart rate sonification: A new approach to medical diagnosis. *Leonardo*, *37*(1), 41-46.

Barrass, S. (2012). The aesthetic turn in sonification towards a social and cultural medium. *AI & society*, *27*(2), 177-181.

Barrass, S., & Barrass, T. (2013). Embedding sonifications in things.

Barrett, N., & Mair, K. (2014, May). Sonification for geoscience: Listening to faults from the inside. In *EGU General Assembly Conference Abstracts* (Vol. 16, p. 4489).

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. *Thinking in sound: The cognitive psychology of human audition*, 10-36.

Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces* (Doctoral dissertation, University of York).

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1), 139-159.

Brower, C. (2000). A cognitive theory of musical meaning. *Journal of music theory*, *44*(2), 323-379.

Clark, A. (2001). *Natural-born Cyborgs.* New York: OUP USA .

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae*, *5*(2), 195-212.

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. MIT press.

Dreyfus, H. L. (1965). *Alchemy and artificial intelligence.* Santa Monica, CA: The Rand Corporation, Research Report.

Fairfax, T., Laing, C., & Vickers, P. (2014). Network Situational Awareness: Sonification and Visualization. *Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance*, 334.

Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In *Santa Fe Institute Studies in the Science of Complexity Proceedings* (Vol. 18, pp. 307-307). Addison-Wesley Publishing Co.

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. *Faculty Publications, Department of Psychology*, 430.

Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, *36*(4), 317-319.

Godøy, R. I. (2006). Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, *11*(02), 149-157.

Godøy, R. I., Jensenius, A. R., & Nymoen, K. (2010). Chunking in music by coarticulation. *Acta Acustica united with Acustica*, *96*(4), 690-700.

Graham, R., & Bridges, B. (2014a). Gesture and Embodied Metaphor in Spatial Music Performance Systems Design. In *NIME* (pp. 581-584).

Graham, R., & Bridges, B. (2014b). Strategies for spatial music performance: the practicalities and aesthetics of responsive systems design. *Divergence Press*, 3, December 2014. University of Huddersfield Press/Centre for Research in New Music (CeReNeM). [Online journal article]. Available at: DOI: 10.5920/divp.2015.33. [last accessed 22 July 2016].

Grond, F. & Berger, J. (2011). Parameter mapping sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., (Eds) *The Sonification Handbook,* (pp. 363–397). Berlin, Germany: Logos Publishing House.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-346.

Hermann, T., & Hunt, A. (2005). Guest editors' introduction: An introduction to interactive sonification. *IEEE* *multimedia*, (2), 20-24.

Hermann, T., Hunt, A. & Neuhoff, J. G. (2011). Introduction. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook,* (pp.1–6) Berlin, Germany; Logos Publishing House.

Hermann, T., & Ritter, H. (1999). Listen to your data: Model-based sonification for data analysis. *Advances in intelligent computing and multimedia systems*, *8*, 189-194.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Imaz, M., & Benyon, D. (2007). *Designing with blends: Conceptual foundations of human-computer interaction and software engineering methods*. Mit Press.

Isaacson, R. E., & Aust, R. J. (2007, June). The Use of Sonification in Synthetic Voice Messages. In *EdMedia: World Conference on Educational Media and Technology* (Vol. 2007, No. 1, pp. 957-965).

Johnson, M. (1987). The body in the mind: The bodily basis of meaning, imagination, and reason. University of Chicago Press.

Johnson, M. (2013). II. Identity, Bodily Meaning, and Art. *Art and Identity: Essays on the Aesthetic Creation of Mind*, *32*, 15.

Kendall, G. S. (2014). The Feeling Blend: Feeling and emotion in electroacoustic art. *Organised Sound*, *19*(2), 192.

Klemmer, S. R., Hartmann, B., & Takayama, L. (2006, June). How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems* (pp. 140-149). ACM.

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Indiana University Press.

Lakoff, G.. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind. University of Chicago Press.*

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic books.

Leman, M. (2008). *Embodied music cognition and mediation technology*. Mit Press.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception: An Interdisciplinary Journal*, *5*(3), 315-349.

Lockton, D., Bowden, F., Brass, C., & Gheerawo, R. (2014). Powerchord: Towards ambient appliance-level electricity use feedback through real-time sonification. In *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services* (pp. 48-51). Springer International Publishing.

McAdams, S. (1987). Music: A science of the mind?. *Contemporary Music Review*, *2*(1), 1-61.

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, vol. 63, pp. 81-97.

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. Macmillan.

Park, Tae Hong, Jonathan Turner, Christopher Jacoby, Alex Marse, Michael Musick, Ajay Kapur, and Jinyin He. *Locative sonification: Playing the world through citygram*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2013.

Peres, S. C. (2012). A comparison of sound dimensions for auditory graphs: Pitch is not so perfect. *Journal of the Audio Engineering Society, 60*(7/8), 561-567.

Peres, S.C., & Lane, D.M. (2005). Auditory graphs: The effects of redundant dimensions and divided attention. *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)* (pp.169-174), Limerick, Ireland.

Polli, A. (2012). Soundscape, sonification, and sound activism. *AI & society*, *27*(2), 257-268.

Pollack, I. The assimilation of sequentially encoded information. *Amer. J. Psychol.,* 1953, **66,** 421-435.

Pollack, I. The information of elementary auditory displays. *J. Acoust. Soc. Amer.,* 1952, **24,** 745-749.

Pollack, I. The information of elementary auditory displays. II. *J. Acoust. Soc. Amer.,* 1953, **25,** 765-769.

Polanyi, M. (1966). The Tacit Dimension. Anchor. *Garden City, NY.*

Rimland, J., Ballora, M., & Shumaker, W. (2013, May). Beyond visualization of big data: a multi-stage data exploration approach using visualization, sonification, and storification. In *SPIE Defense, Security, and Sensing* (pp. 87580K-87580K). International Society for Optics and Photonics.

Roddy, S. (2015). Embodied Sonification. Doctoral Dissertation. Trinity College Dublin. Ireland.

Roddy, S., & Furlong, D. (2014). Embodied aesthetics in auditory display. *Organised Sound*, *19*(01), 70-77.

Roddy, S., & Furlong, D. (2015, July). Sonification listening: An empirical embodied approach. In *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015)*.

Ryle, G. (1949). *The concept of mind.* London, UK: Hutchinson.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(03), 417-424.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of information.* Urbana, Illinois: University of Illinois Press.

Sjuve, E. (2015, July). Metopia: experiencing complex environmental data through sound. In *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015)*.

Smalley, D. (1996). The listening imagination: listening in the electroacoustic era. *Contemporary Music Review*, *13*(2), 77-107.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised sound*, 2(02), 107-126.

Stevens, S. S. (1975). *Psychophysics.* Transaction Publishers.

Supper, A. (2014). Sublime frequencies: The construction of sublime listening experiences in the sonification of scientific data. *Social Studies of Science*, *44*(1), 34-58.

Varela, F. J., & Thompson, E. &E. Rosch. (1991). *The Embodied Mind: cognitive science and human experience*. Cambridge, Mass.: MIT Press.

Walker, B. N. (2000). *Magnitude estimation of conceptual data dimensions for use in sonification*. Doctoral Dissertation. Rice University, Houston, TX.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. *Ecological psychoacoustics*, 150-175.

White, M. (2003). Metaphor and economics: the case of growth. *English for specific purposes*, *22*(2), 131-151.

Wolf, K., & Fiebrink, R. (2013). SonNet: A Code Interface for Sonifying Computer Network Data. In *NIME* (pp. 503-506).

Worrall, D. (2009). Sonification and Information: Concepts, instruments and techniques. PhD thesis, University of Canberra.

Worrall, D. (2010). Parameter mapping sonic articulation and the perceiving body.

Wilkie, K., Holland, S., & Mulholland, P. (2010). What can the language of musicians tell us about music interaction design?. *Computer Music Journal*, *34*(4), 34-48.

Yoshida, T., Kitani, K. M., Koike, H., Belongie, S., & Schlei, K. (2011, March). EdgeSonic: image feature sonification for the visually impaired. In *Proceedings of the 2nd Augmented Human International Conference* (p. 11). ACM.

Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press.

Learning music production practice through evolutionary algorithms

Alex Wilson
University of Salford
Greater Manchester, UK

Bruno Fazenda
University of Salford
Greater Manchester, UK

The field of intelligent music production has been an active research topic for over a decade. The aim is to develop systems which are capable of performing common tasks in music production, such as level-balancing, equalisation, panning, dynamic range compression and application of artificial reverberation.

Many systems developed are modelled as expert systems, where the music production task is solved by optimisation, and domain knowledge, obtained by examining industry "best-practice" methods, is used to determine the optimisation target [1]. Drawbacks to this method include the fallibility of domain knowledge and the assumption that there is a global optimum – a mix which all users would agree is best. Results suggest that many systems can perform basic technical tasks but struggle to compete with human-made mixes, due to a lack of creativity.

We propose to use interactive evolutionary computation to solve this problem. These methods are well suited to aesthetic design problems, which are highly non-linear and non-deterministic. In the case of music mixing, the problem is highly subjective: research has shown that mix engineers typically prefer their own mix to those of their peers [2]. Consequently, intelligent music production tools would benefit from interactivity, to determine "personal" global optima in the solution space, instead of one "universal" global optimum.

The space to be explored is a novel "mix-space" [3]. This space represents all the mixes that it is possible to create with a finite set of tools. Currently, basic level adjustment has been implemented, while mix-space representations of panning and equalisation are currently under development.

The fitness function for optimisation is subjective, allowing mixes to be generated in accordance with any perceptual description, such as "warmth", "punchiness" or "clarity". Clustering techniques are used to increase the population size beyond that which a user could realistically rate, by extrapolating the fitness function to nearby individuals. When optimising the overall "quality" of the mix, we introduce findings from recent, large- scale studies of music mixes [4], which can be used to calculate the fitness of the population, alongside the subjective rating.

Early results indicate that the system can produce a variety of mixes, suited to varying personal taste. We believe this approach can be used to further the study of intelligent music production, to deliver personalised object-oriented audio and increase the understanding how music is mixed.

The presentation slides can be found at the following location http://usir.salford.ac.uk/39220/.

References

[1] Pestana, P., & Reiss, J. (2014, January). Intelligent audio production strategies informed by best practices. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society.
[2] De Man, B., Boerum, M., Leonard, B., King, R., Massenburg, G., & Reiss, J. D. (2015, May). Perceptual evaluation of music mixing practices. In *Audio Engineering Society Convention 138*. Audio Engineering Society.
[3] Wilson, A., & Fazenda, B. M. (2015, July). Navigating the mix-space: theoretical and practical level-balancing

technique in multitrack music mixtures. In *Proceedings of the 12th Sound and Music Computing Conference*. Sound and Music Computing.
[4] Wilson, A., & Fazenda, B. (2015, October). 101 mixes: a statistical analysis of mix-variation in a dataset of multi-track music mixes. In *Audio Engineering Society Convention 139*. Audio Engineering Society.

The Design of Tangible Digital Musical Instruments

Gareth W. Young
Dept. of Computer Science / Music
University College Cork
G.Young@cs.ucc.ie

Katie Crowley
Trinity College Institute of Neuroscience
Trinity College Dublin
K.Crowley@tcd.ie

Abstract Here we present guidelines that highlight the impact of haptic feedback upon the experiences of computer musicians using Digital Musical Instruments (DMIs). In this context, haptic feedback offers a tangible, bi-directional exchange between a musician and a DMI. We propose that by adhering to and exploring these guidelines the application of haptic feedback can enhance and augment the physical and affective experiences of a musician in interactions with these devices. It has been previously indicated that in the design of haptic DMIs, the experiences and expectations of a musician must be considered for the creation of tangible DMIs and that haptic feedback can be used to address the physical-digital divide that currently exists between users of such instruments.

1. Introduction

General advances in technology have always influenced the field of music technology, facilitating the modern musician's requirement for new devices and encouraging creative expression. Music technology has a deep-rooted history of performance and a close-knit relationship with human interactions with musical devices. Through the use of natural sound resonating objects, such as reeds, bells, pipes, and others, humans have made possible the creation of musical instruments. Traditionally, it was the limitations of the sound-generating device that determined the design of an instrument. However, this fact has never deterred the making of music from most any man-made or naturally resonant object.

With the discovery and widespread application of electricity in the early part of the 20th century, the number of new mediums for sound generation increased. The majority of early electronic musical instruments were keyboard based, drawing upon the universal success of acoustic instruments such as the piano and harpsichord. Notable exceptions that deviated from this design principle are instruments such as the Theremin [1] and other instruments that made use of gesture sensitive inputs to control timbre, pitch, and/or volume. Another example would be the Trautonium [2], which operated via a touch sensitive strip across its length. Of the keyboard-based instruments, advancements in functionality were achieved via increasing the devices sensitivity or manipulation through the development of additional knobs and buttons, for example, the Ondes Martenot [3] and the Electronic Sackbutt [4].

Digital Musical Instruments (DMIs) have very few limits and the potential design possibilities are vast. Beyond musical performance, in devices that operate on a one-to-one interaction, new DMIs are also encompassing other jurisdictions of musical composition. Artists may become proficient in the use of a singular instrument or they may choose to become the master of a multi-instrumental controller. Musicians may concentrate all of their efforts into increasing their skill in playing a stand-alone instrument or they may choose to master the control of multiple sound sources through digital manipulation. A performer may also have an indirect influence over an installation or live recital, becoming a unique and often difficult to control aspect of a performance. Beyond the musician, performance itself has also changed. The musical medium is also no longer a static performance, as a single musician or ensemble on stage, it moves beyond this. It can be inclusive of the movements of a dancer, a dance troupe, and even the audience itself. The inclusion of multiple free movements into music production paves the foundations for a more expansive interaction.

2. Background

In computer music, virtually any gesture can be captured and translated into a control signal. In the application of DMIs, these gestures are often used as a control source for complex sound synthesis modules. With the separation of interface from sound source, new musical devices are afforded near endless freedom of form. However, they are becoming unrecognisable, as the gestures captured by a device do not require resemblance of anything ever applied before. The multiple combinations of different styles of interface design have protracted the performance techniques that musicians are afforded in performance. This is indicative in the increased popularity of DMIs in contemporary music, as they have been embraced and accepted as a new means for artistic expression.

Musical interface models based upon the playing principles of musical instruments have previously been seen [5]. In Figure 1, we present our own concept of a closed-loop model of a tangible musical instrument. It is proposed here that if a DMI wishes to be considered as tangible, a number of steps must be followed to ensure this. In this context, both artist and instrument can be observed as two separate entities that are independent of each other. The link between user and instrument mediates between the minor components contained within. The relationship between these two modules is realised through gestures made and gestures captured. The musician or artist is independently providing the intention (often attained through training and previous experience) and the necessary gestures specific to the interface. The instrument captures these physical interactions and processes them into a form of control data. The sound generator makes use of the data collected from gestures captured by applying control parameters to a physical sound generating design. The physical separation of these modules is impossible to achieve in acoustic instruments (represented in Figure 1, as the gesture interface is rarely removed from the sound source. DMIs allow us to separate the user from the instrument, permitting us to rethink the relationships formed between the two. For example, a gesture can be made and the sound generated varies in some way; however, the gesture does not necessarily have to relate to a control change in the sound generator, as it may also convey performance information that is not audible. What has become apparent from observing current DMI trends is that whilst performers have been given absolute freedom of gesture capture, they have at the same time eliminated haptic feedback, a key feedback channel of information through which they can measure the response of the instrument and the accuracy of their movements. In the realm of gesture capture, synthesis algorithms and control rate data have been separated from the sound producing mechanisms along with the performer. The capture of human performance with such devices forces the user to rely heavily on the proprioceptive, visual, and aural data cues, or more simply put: “...the computer music performer has lost touch” [6].

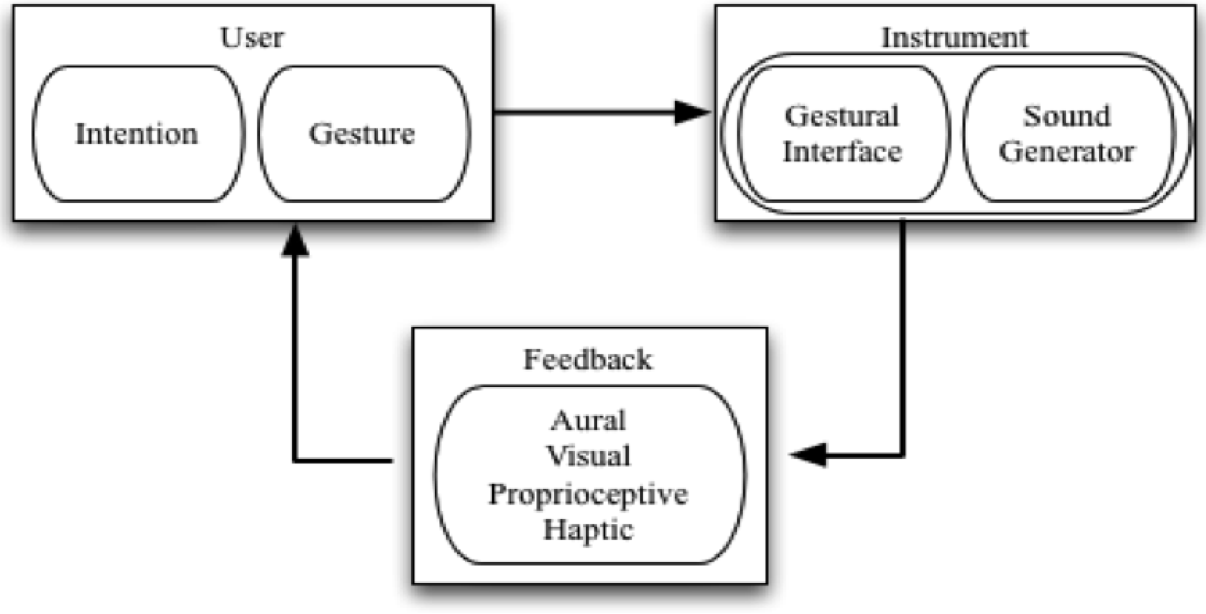


Figure 1: Closed Feedback Loop of a Tangible Musical Instrument.

3. Design Guidelines

Through the amalgamation of digital music technology and electronic musical instruments, DMIs have emerged. A DMI is a musical instrument that is capable of producing sound via digital means. They are specifically constructed with a separable control interface and sound generator; however, these are not always separate. The mapping of a gestural interface to a sound generator translates the input gestures into sound control signals that are applied to a sound generator. The separation of these two elements enables musicians to approach the creation of music differently from how they would with an acoustic instrument, as the physical constraints of sound generation and input gesture are no longer inseparable. This approach allows for the sonification of gestures or the creation of a sound-generating algorithm that is controlled via an unknown or undefined input gesture.

DMI designs, such as the Rhythm’n’Shoes [7], The Sound Flinger [8], the Haptic Carillon [9], The Vibrobyte [10], StickMusic [11], and The Plank [12] have demonstrated the successful application of haptic feedback in musical devices. However, the majority of commercial interfaces in the field of digital synthesis have focussed on simulating the effects of acoustic keyboard instruments (such as the piano, harpsichord, or organ) and bowed instruments (such as the violin, viola or cello) [13]. Furthermore, previous research has highlighted that many of these DMIs fail to balance complexity with usability and that they lose transparency due to the separation of sound generator and gestural interface [14]. In the guidelines outlined herein, haptic information that can be used to address these issues will be focused on and will attempt to resolve problematic issues of interaction. We therefore propose consideration of the following guidelines when creating a tangible DMI:

- Transparent in use.
- Reactive and communicative to as many of the user's senses as possible.
- A tangible interpretation of the device's reaction must be possible.
- Clarity of affordances delivered via the sensor technologies applied.
- Consistency in the information displayed to the user.
- The application of clear and stable mapping methodologies.
- Clear and consistent constraints for the interpretation of gestures made.

3.1. Transparent

It must be possible to clearly determine the function of the instrument, for both the musician and the observing audience, as it is easier to recognise an action than to recall one. The musician should be able to quickly and easily identify the options offered by the instrument, and how to access them. In contrast, options that are "out of sight" can be difficult to find and their functionality unintuitive.

3.2. Reactive

In relation to a devices transparency, information must be presented to as many of the user's senses as is possible in a timely and logical manner to emphasise the effect of the input interaction upon the system in use (see Figure 1). The provision of feedback gives the user confirmation that an action has been performed successfully, or otherwise. Physical or auditory activational feedback can be used to indicate to the musician whether or not their intended action has successfully been performed.

3.3. Tangible

All information related to the system's reaction should be presented to the musician clearly and they should also be able to interpret meaning easily, this will serve to enhance discoverability and improve the musician's overall understanding of the device.

3.4. Clarity of Affordance

A DMI that is designed with familiar features should be done so with clarity in how these features react and should therefore respond in a recognisable and familiar way. For DMI designers, affordance means that as soon as the musician sees the instrument, they should have some knowledge as to how to use it. Users need to be able to tell how to access the full spectrum of options available to them. When the affordances of a physical object are perceptually obvious it is easy to know how to interact with it. The standard example used for explaining affordance is the door handle. A round doorknob invites you to turn the knob. A flat plate invites you to push on the plate to push the door outwards. A graspable handle invites you to pull on the handle to pull the door open towards you.

3.5. Consistent

The location, appearance, significance, and behaviour of an interface must be consistent for it to be effectively learned. In achieving this, when errors are made the interface will allow musicians to recover and continue without any additional mental or physical strain. Good design principles suggest that the same action has to cause the same reaction, every time. This is particularly important in the context of a DMI where the musician can experience flow [15] during the performance and requires the certainty and assurance that their intended action will be achieved, without having to focus on interface-related concerns. Consistency is also important in terms of learning how to play and master the DMI. People learn by discovering patterns and this is particularly relevant in a musical context. Consistency is key to helping users recognise and apply patterns. Conversely, inconsistency can causes confusion and uncertainty because something does not work in the way the user expects. Forcing users to memorise exceptions to patterns and rules increases the cognitive burden and leads to distraction, particularly during a performance. Consistency is important for instilling confidence in both the system and instrument.

3.6. Clear and Stable Mapping

The mapping of gestures in a spatial context and the systems temporal responses should be clear and stable. Controls should be positioned in logical and intuitive ways, for example, it is obvious that a slider control designed to manipulate volume maps the direction of ”up” to increase volume and ”down” to decrease. Nonconventional mappings need to be learned and can conflict with consistencies guidelines, however, they are permissible when an appropriate or valid reason exists.

3.7. Constraints

The introduction of physically identifiable, logical, and clear limitations upon an interaction will prevent errors and assist in interaction interpretation by both the musician and the system in use. Interfaces must be designed with restrictions so that the system can never enter into an invalid state - the same principle applies to DMI design. Constraints can be physical, such as a size limitation on an object or restriction threshold on the angle of movement of a lever, for example.

4. Discussion

It is hoped that through the application of these design guidelines that advances in the field of Computer Music will be made. Specifically, it is foreseen that the study of interactions between performers and digital instruments in a variety of contexts will continue to be of interest in this field far beyond the scope presented here. Further research on digital musical instruments and interfaces for musical expression should continue to explore the role of haptics, previous user experience, and the frameworks that are constructed to quantify the relationship between musical performers and new musical instruments. The complexities of these relationships are further compounded by the skills of musicians and are far more meaningful than a physically stimulating interaction and should therefore be explored further.

The designer of a DMI is often the performer and a DMI may take many forms; from concept to performance tool. In a similar vein, in the design processes of computer interfaces, evaluation tools are applied iteratively, in cycles that address the design issues raised within the previous sequence. An example of this can be seen in Norman's Seven Stages of Action as a design aid in interaction design [16]. Whilst appraising a DMI, the musician must constantly questions certain aspects of usability when applied to specific tasks. For example:

- Can I achieve my goals accurately and effectively?
- Am I working productively and efficiently?
- Is the device functioning as I expect it to?
- At what rate am I acquiring new skills?

Emergent DMI systems require further measures for an accurate appraisal of the user's experience when applying the device in a musical context. In a traditional HCI analysis, a device is evaluated in a specific context and the evaluation methods are expert-based heuristic evaluations or user-based experimental evaluations. Only by determining context is it possible to interpret correctly the data gathered. Therefore, it is suggested that to fully understand the effects of tangibility upon DMI performance, specific evaluation techniques must be formulated, such as the application of functionality, usability, and user experience evaluation methods [17].

The ideas presented in this paper have only begun to explore the possibilities of tangibility in future DMI designs. The ideas presented endeavoured to present thoughts towards the influence of feedback on a user's perception DMI tangibility. Beyond this, future research goals will include the development of laboratory tools that will assist in the creation of a DMI design environments that will allow designers to experiment with different communication paradigms and gestural interface models. Within this space, composers, performers, and DMI designers will be able to explore the affordances of new sensor technologies in the creation of new instruments for musical expression.

5. Conclusion

By following the guidelines presented here, haptically enabled DMI designs will be fully communicative to all senses and present computer musicians with an array of carefully designed tools for their own artistic endeavours. Furthermore, we believe that the experiences of the audience will also be improved upon as clarity between the musician's actions and the system's response will be achieved.

References

[1] Theremin World, “What’s a Theremin?” (December, 2005). [Online]. Available: `http://www.thereminworld.com/Article/14232/whats-a-theremin`

[2] New York Times, “ELECTRICITY, ETHER AND INSTRUMENTS; Some Considerations, Reflections and Inferences Regarding the Modern Cult of Vitamineless Art and the Synthetic Esthetic” (6 September 1931). [Online]. Available: `http://timesmachine.nytimes.com/timesmachine/1931/09/06/96381853.html`

[3] T. Bloch, “Ondes Martenot” (12 March 2015). [Online]. Available: `http://www.thomasbloch.net/en_ondes-martenot.html`

[4] Canada Science and Technology Museums Corporation, ”HUGH LE CAINE *ELECTRONIC SACKBUT* SYNTHESIZER,” (27 February 2015). [Online]. Available: `http://cstmuseum.techno-science.ca/en/collection-research/artifact-hugh-le-caine-electronic-\sackbut-synthesizer.php`

[5] J. Rovan & V. Hayward, “Typology of Tactile Sounds and their Synthesis in Gesture-Driven Computer Music Performance,” in *Trends in Gestural Control of Music*, M. Wanderly & M. Battier, Eds., Editions IRCAM, 2000.

[6] N. Castagné, C. Cadoz, J. L. Florens & A. Luciani, “Haptics in computer music: a paradigm shift” EuroHaptics, 5-7 June 2004.

[7] S. Papetti, M. Civolani & F. Fontana, “Rhythm’n’shoes: a wearable foot tapping interface with audio-tactile feedback” in *New Interfaces for Musical Expression*, Oslo, 2011.

[8] C. Carlson, E. Marschner & H. McCurry, “The Sound Flinger: a Haptic Spatializer” in *New Interfaces for Musical Expression*, Oslo, 2011.

[9] M. Havryliv, F. Naghdy, G. Schiemer & H. T., “Haptic Carillon – Analysis & Design of the Carillon Mechanism” in *New Interfaces for Musical Expression*, Pittsburgh, 2009.

[10] K. Mcdonald, D. Kouttron, C. Bahn, J. Braasch & P. Oliveros, “The Vibrobyte: a Haptic Interface for Co-Located Performance,” in *New Interfaces for Musical Expression*, Pittsburgh, 2009.

[11] H. C. Steiner, “StickMusic: Using haptic feedback with a phase vocoder” in *New Interfaces for Musical Expression*, Hamamatsu, 2004.

[12] B. Verplank, M. Gurevich & M. Mathews, “The Plank: Designing a simple haptic controller,” in *New Interfaces for Musical Expression*, Singapore, 2002.

[13] E. R. Miranda & W. M. Wanderley., “Haptic Music Controllers: Tactile and Force Feedback,” in *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, vol. 21, Midleton, WI: A-R Editions, 2006, pp. 71-83.

[14] T. Murray-Browne, D. Mainstone, N. Bryan-Kinns & M. D. Plumbley, “The medium is the message: Composing instruments and performance mappings,” in *New Interfaces for Musical Expression*, Oslo, 2011.

[15] M. Csikszentmihalyi, & I. S. Csikszentmihalyi, (Eds.), “Optimal experience: Psychological studies of flow in consciousness,” Cambridge University Press, 1992.

[16] D. A. Norman, “The design of everyday things: Revised and expanded edition,” New York: Basic books, 2013.

[17] G. W. Young & D. Murphy, “HCI Models for Digital Musical Instruments: Methodologies for Rigorous Testing of Digital Musical Instruments,” in the *Int. Symp. on Computer Music Multidisciplinary Research*, Plymouth, UK, 2015.