
Peircing Fritz and Snow: An aesthetic field for sonified data

MICHAEL FILIMOWICZ

School of Interactive Arts and Technology, Simon Fraser University, 250-13450 102 Avenue, Surrey, BC V3T 0A3, Canada
E-mail: mfa13@sfu.ca

This essay elaborates a field of general aesthetic considerations relevant to the sonification of data. A set of dialectical tropes are introduced to define the possibility space for organised sonified data: data-in-itself and the listener-for-itself; cognitive support and sabotage; and the Peircean triad of rheme–dicisign–argument. Taken together, these three dialectical parameters elaborate a conceptual space in which strategies can be sought for mapping acoustic parameters to data features, data structure and sonic transformations, all with respect to listener reception. A work-in-progress is discussed in connection with this general aesthetic field, and considerations of the aesthetic space are applied to several works. Finally, the notion of data verité is explored in connection to ‘big data’ and issues related to the transformation of data into information generally.

1. THE AESTHETIC FIELD AS DIALECTICAL SPACE

In setting up an aesthetic field for data sonification, I will start with a detour through film and literature in order to establish a dialectical tension between two conceptual personae representing data-in-itself and a listener-for-itself as extreme possibilities for sonic production that takes data as source material. Our first dialectical persona will be the character Fritz in Wim Wenders’ *Lisbon Story* (1994), who in an extended monologue expresses his existential burnout with filmmaking to his sound man Winter:

Images are no longer what they used to be. They can’t be trusted anymore. We all know that, you know that. When we grew up, images were telling stories, showing things. Now they’re all into selling, stories and things. They’ve changed under our very eyes, they don’t even know how to show it any more. They’ve plain forgotten. Images are selling out the world and at a big discount! When I came to Lisbon to make this little movie, I thought I could beat the drift. We talked about it man, remember? I wanted to shoot it in black and white on this old hand cranker. Like Buster Keaton and *The Cameraman*. Grinding in the streets on my own, *A Man with a Camera*, a Dziga Vertov, pretending that the whole history of cinema hadn’t happened, and that I could just start from scratch one hundred years later. Well it didn’t work. That is, for a while it seemed to work. Then it all collapsed. I really love this city. Lisboa! And most of the time, I really saw it. In front of my eyes. But pointing a camera is like pointing a gun. And each

time I pointed it, it felt like life was drained out of things. And I cranked and I cranked. But with each turn of the ol’ handle, the city was receding, and fading further and further. Like the Cheshire Cat. Nada. It was becoming unbearable. I took a real beating. That’s when I called you for help. For a while, I lived with the illusion that Sound would save the day. That your mics with my images ... it’s, it’s hopeless. It’s all hopeless, Winter. Hopeless. But there is a way, Winter. I’m working on it. Listen. An image that is unseen can’t sell anything. It is pure, therefore, true. Beautiful and in one word, innocent. As long as no eye contaminates it, it is in perfect unison with the world. If it is not seen, the image and the object it represents belong together. Yes, it is only one sweet look at the image, the thing in it, it dies. There it is, Winter. My library of the unseen image. Every one of these tapes was shot with nobody looking through the lens. Nobody saw them while they were recorded, nobody verified them afterwards. I shot every goddamn one on my back. These images show the city as it is, not as I want it to be. Anyway, there they are, in their first sweet sleep of innocence. Ready to be viewed by some future generation, with eyes different from ours. Don’t worry, mate, we’ll both be dead.

Fritz’s solution to the inauthenticity – commodification and corruption – of media production is the stockpiling of pure data (no pun intended in reference to the well-known open source software originally developed by Miller Puckette (Pure Data n.d.), shot by a camera on his back, unpolluted by his eyes or intent, and remaining in a state of perpetual data, only possibly becoming information should some hypothetical future generation view the sealed stockpile of unwatched tapes. Fritz will be our ‘superego’ for the voice of non-corrupted data, left unblemished by human contact, the dialectical pole of data-in-itself.

Representing the listener-for-itself, the second dialectical persona to introduce is the scientist Snow in Stanislaw Lem’s classic sci-fi novel *Solaris* (1955: 81):

We take off into the cosmos, ready for anything: for solitude, for hardship, for exhaustion, death. Modesty forbids us to say so, but there are times when we think pretty well of ourselves. And yet, if we examine it more closely, our enthusiasm turns out to be all sham. We don’t want to conquer the cosmos, we simply want to extend the boundaries of Earth to the frontiers of

the cosmos. For us, such and such a planet is as arid as the Sahara, another as frozen as the North Pole, yet another as lush as the Amazon basin. We are humanitarian and chivalrous; we don't want to enslave other races, we simply want to bequeath them our values and take over their heritage in exchange. We think ourselves as the Knights of the Holy Contact. This is another lie. We are only seeking Man. We have no need of other worlds. We need mirrors. We don't know what to do with other worlds. A single world, our own, suffices us; but we can't accept it for what it is. We are searching for an ideal image of our own world: we go in quest of a planet, of a civilisation superior to our own but developed on the basis of a prototype of our primeval past. At the same time, there is something inside us which we don't like to face up to, from which we try to protect ourselves, but which nevertheless remains, since we don't leave Earth in a state of primal innocence. We arrive here as we are in reality, and when the page is turned and that reality is revealed to us – that part of our reality which we would prefer to pass over in silence – then we don't like it anymore.

Snow presents a picture of subjectivity as a closed self-reinforcing feedback loop, structurally incapable (non-open) to Otherness (the alien, other worlds) and in perpetual search of only repeating itself while overlooking or repressing the unpleasant, which gives us our opposite dialectical pole, the listener-for-itself (indeed, of itself), the extreme possibility at the farthest remove from data captured and stored to be unperceived by anyone. If we draw an analogy with astrophotography, Fritz would accuse us of corrupting the data of the radio telescopes by wanting to transform it into images of distant galaxies (he is our superego of representational authenticity), while Snow would make sure the hues, contrasts and saturations presented a scenic cosmic landscape that we could take a nice weekend drive through, once faster-than-light personal vehicles come onto the market (Snow is our cosmic brochure producer). Fritz and Snow should be understood as *conceptual personae*, in the sense advanced by Deleuze and Guattari as affective figures prior to but essential for concept formation: 'It is possible that the conceptual persona only rarely or allusively appears for himself. Nevertheless, he is there, and however nameless and subterranean, he must always be reconstituted by the readers' (1995: 63).

If Fritz were in charge of our data sonification project, he would merely bill us for the accumulation of hard drives or cloud storage, while Snow would assure us of a pleasant musical experience not unlike whatever other experiences we would consider to be pleasant and musical. An aesthetic field, however, is either a plane or a space, and by posing two points (Fritz: data-in-itself and Snow: listener-for-itself) we have so far only created a line. Other vertices and vectors are needed to fill out the area of the field under consideration and construction.

The next dialectical tension that is suggested by data sonification is the classic opposition between strict determinacy (procedural serialism) and indeterminacy (chance operations), with perhaps stochastic algorithms mediating these two poles of possibility. Here we ask, to what extent should the data *determine* the sonified form? Fritz would say (begrudgingly, since he would claim that listening itself is already data corruption) that the form of the data should equate completely to the form of the sound, while Snow would suggest ready-made stylistic parameters to which we can map the data set (perhaps salsa or gagaku). If the ghosts of Schoenberg and Cage were to appear, bidding us toward either strict or non-determinacy, Bob Snyder (another conceptual persona who has just shown up in this aesthetic field) would remind us that serialist and chance music can sound exactly the same to a listener depending on how the composition addresses basic cognitive parameters, such as what can reasonably fit within the duration of short-term memory. In *Music and Memory* (2001), Snyder describes the set of compositional strategies we should keep in mind if we want to take into account the hard-wired limitations of memory that condition our overall sense of aesthetic form:

Memory is an important consideration when talking about music. Indeed, we can divide music into two broad categories based on the use of memory:

1. Music that attempts to exploit long-term memory by building up hierarchical and associative mental representations of large time structures; and
2. Music that attempts to sabotage recognition and expectation by frustrating recollection and anticipation, thereby intensifying the local order of the present. (Snyder 2001: 234).

Given that the listener's cognitive capacities over-determine the composer's aesthetic decision to map data strictly or loosely to sonic parameters, we can rename the second dialectical aspect of our aesthetic field as cognitive *support or sabotage*. If interested in the former, the Snowist sonifier of the data set would need to respect the limits of intelligibility, 'consisting of events that clearly exemplify categories of primary parameters such as pitch and rhythm, [supporting] the possibility of identifiable patterns' (Snyder 2001: 234). The Fritzian cognitive saboteur, on the other hand, could make use of either high/low-information or memory-length strategies:

Strategies that involve high information and low redundancy, such as using random pitch and rhythm patterns, produce music that cannot be fitted into a schema (standardised musical form). (Snyder 2001: 235)

At the opposite extreme, low-information music keeps change to a minimum, with only the bare minimum of contrast necessary to sustain interest. This is usually accomplished through a great deal of repetition on various levels. (Snyder 2001: 236)

Although overlapping with low-information strategies, memory length strategies involve, not particular types of patterns and their information content, but duration and silence, using these in ways that make musical information difficult to process. Some memory length strategies attempt to break up the continuity of the present by creating situations in which short-term memory cannot perform its usual function. (Snyder 2001: 237)

Categories such as serialist, random or chance procedural refer to the intentions and processes of the composer. However the listener-for-itself is only interested in what s/he is hearing, not why or how the sounds came to be produced (the backstory in the programme notes). Data-in-itself, on the other hand, is such a purist that it would prefer only the harshest of machinic translations.

With this second dialectical tension (cognitive support or sabotage) I have for the moment ‘resolved’ the tension between differing modes of compositional intending or producing by deferring or subsuming both to the perceived ‘end result’ in the listener’s experience. However, it is worth a brief revisiting of process to round out the aesthetic field of data, which will give us another dialectical tension and thus transform our aesthetic field from a plane into a space. John Cage’s *Atlas Ellipticalis* suggests that there are three constitutive sources of audible parameters in data: parameters ascribed to features of the data itself, parameters ascribed to the structural organisation of the data (e.g. a matrix, array or table structure), and parameters attributable to the specific means of sonification. Thus, in the case of sonified star data (Cage’s cosmic atlas), one may ascribe parameters in relation to size and brightness of stars (e.g. duration and amplitude); the atlas itself spatialises the data in two dimensions, so that sequences and simultaneities can be produced; and finally, the overlay of the musical staff introduces the order of pitch space. Our puritan superego (Fritz) might see three opportunities here for impurity in the translation, while our accommodating superego (Snow) will see three opportunities for making us ever more comfortable with the organisation and transformation of data into sound. These three parametric orders (data features, data structure, sound structure) restore a dimension of practice, materiality and intent which above we had elided into the cognitive domain through consideration of memory and form. Compositional strategies of support or sabotage would then determine the larger formal properties of the work, with Fritz and Snow operating somewhere in the wings as dictatorial superegos or conceptual personae. However, rather than refer making and intending back to past musical aesthetic practices, we will instead take a detour through semiotics for an alternative explication of formal strategies.

2. PEIRCEAN TRIADS

We can complete the dialectical space described above through the introduction of the Peircean triad of firstness, secondness and thirdness. Peirce’s well-known distinctions between icon (firstness), index (secondness) and symbol (thirdness) refer to the sign’s *relation to its signified object*. However Peirce’s semiotics also introduces triadic categories of a sign’s *relation to itself* (qualisign, sinsign, legisign) and a sign’s *relation to the interpretant* (rheme, dicisign, argument). It is this last triad of Peirce’s categories that are germane to a semiotic perspective on data sonification, to which he also gave the alternate terms ‘seme’, ‘pheme’ and ‘delome’ (Peirce 1906: 506–7). Because data is already an abstraction from some source, the icon–index–symbol triad does not come into play in data sonification. The data itself will typically be symbolic (in Peircean terms), discretised and subsumed under a rule of interpretation (array, matrix, etc.), while some forms of data may be iconic or indexical (analogically resembling or caused by its referent, as in the case of seismographic recordings or ECG graphs). However, data sonification *works with the data, not the object abstracted into data*, so this rules out the icon–index–symbol triad as a major area of concern. Likewise, the quality of the data itself (the sign’s phenomenological character vis-à-vis itself: qualisign, sinsign, legisign) is typically not perceived directly by listeners, but only insofar as it is sonified. However, some sonifications include a parallel ‘visual channel’ in which a visualisation of the data is also shown (Youtube n.d.). In the combined interaction of a sonification with a visualisation, the sign’s self-relation can come more prominently into play. However, since this Peircean dimension also occurs when, for instance, listening to any sound while also looking at its visualised time-domain waveform representation on a screen, I will not be considering this semiotic relation here. Since the ultimate concern of data sonification is the representation that is presented to a listener, it is the triad pertaining *to the interpretant* that is germane for semiotic considerations. A sign in the mode of rheme, dicisign or argument ‘direct[s] an interpreter to its qualities, to its existence, or to its generality’ (de Waal 2001: 75).

The rheme (alt. seme) represents the qualities of objects or processes represented by the data. This category of firstness comprises analogously ‘the characters undergoing the story’ of the composition, the things or objects that are being ‘represented’. The dicisign (alt. pheme) would represent the actual occurrence or existence of the sonified data. As a relation of secondness (only between two elements), it is either in contrast or juxtaposition to the listener (as object against a subject) or in contrast to another sound (an object against an object). In a time-based

phenomenon such as sound, existence is indissociable from sequence, so a gestalt such as a clang/silence contrast (Tenney 1992) could be a *dicisign*. Finally, the argument (*alt. delome*) is a synthetic conceptual unity of a larger relation, integrating the listener, the sound and the referent (data or original object represented by both the sound and data). In this way larger movements of change and relation can be grasped by the listener.

What has been described above as the cognitive strategies – support or sabotage – would pertain primarily to the overall assemblage of multiple or composite arguments, as the overall coherence of the elaborated sound form. Each particular argument would be analogous to lower-order groupings such as clusters or phrases. To the *rheme* we can assign the material qualities of sound (*timbre, pitch, amplitude*) while sequence and simultaneity would be initiated at the level of *dicisign* (a relation between two elements). Peirce continuously elaborated and refined his categories of *firstness*, *secondness* and *thirdness* throughout his life, and these categories (he also called them the *cenopythagorean* categories) are regular themes and preoccupations of his thought, as they derived from his early attempt to reduce Kant's twelve categories of understanding to a simpler triad. In order to provide an aesthetic character to these semiotic divisions, the following excerpts from Peirce's own writings (Peirce 1955) will serve as some indication as to the compositional prospects that can be related to the semiotic triad, and which moves us to considerations beyond determinacy. An aestheticised understanding of *rheme, dicisign* and argument frees us from positing any habitual mapping or simple cause–effect relation between the source data and translated sound object, which is what considerations around determination might lead us toward as a potential compositional trap (too easily reiterating past stylistic thematics or aesthetic concerns).

2.1. Firstness

Firstness is the mode of being which consists in its subject's being positively such as it is regardless of aught else. That can only be a possibility. (Peirce 1955: 76)

The first [category] comprises the qualities of phenomena, such as red, bitter, tedious, hard, heartrending, noble. (77)

The idea of First is predominant in the ideas of freshness, life, freedom. The free is that which has not another meaning behind it, determining its actions. (78)

2.2. Secondness

The second category of elements of phenomena comprises the actual facts. The qualities, in so far as they are general, are somewhat vague and potential. But an

occurrence is perfectly individual. It happens here and now. (77)

The idea of second is predominant in the ideas of causation and of statical force. For cause and effect are two; and statical forces always occur in pairs. (79)

The second category ... is the element of struggle. (89)

2.3. Thirdness

Thoughts are neither qualities nor facts. (78)

Secondness is the predominant character of what has been done. The immediate present, could we seize it, would have no character but Firstness But we constantly predict what is to be. (91)

It differs from immediate consciousness, as a melody does from one prolonged note. Neither can the consciousness of the two sides of an instant, of a sudden occurrence, in its individual reality, possibly embrace the consciousness of a process. This is the consciousness that binds our life together. It is the consciousness of a synthesis. (97)

Elsewhere I have used the term *imago mundi* (Filimowicz and Stockholm 2010: 10) to mean a 'world image' or image of worldliness, to describe the production of acoustic images that are devoid of or attenuated in subjective investment. Data sonification has the potential to offer us percepts that are uninvested by subjectivity, through the production of 'world images' that can decentre or displace our general egocentrism with regards to our world. For an example of what a 'true' *imago mundi* would be (a nod to the Fritz persona), we can consider radio frequencies from deep space that are brought into our perceptual registers *without* being translated into astrophotography. However, given the social fate of electroacoustic and contemporary music generally, it remains to be heard whether disinterested sonifications of data would be of much appeal to the general public. Some attention to Snow's remark – our existential need for mirroring and familiarity – is no doubt called for. Peirce's triad helps us to articulate the concern of the audience: the relation of signs to interpretant maps well to the general concerns of relating sounds to listeners, which is of a different order from the concern of mapping data into sound.

3. THE SOUND OF GRADING

Thus far I have discussed dialectical tensions that pertain to the compositional prospects of data sonification at the level of a general conceptual aesthetic field. Here I will integrate a work-in-progress into the discussion in order to better ground the theoretical concerns in praxis. The particular data set under consideration for sonification consists of the

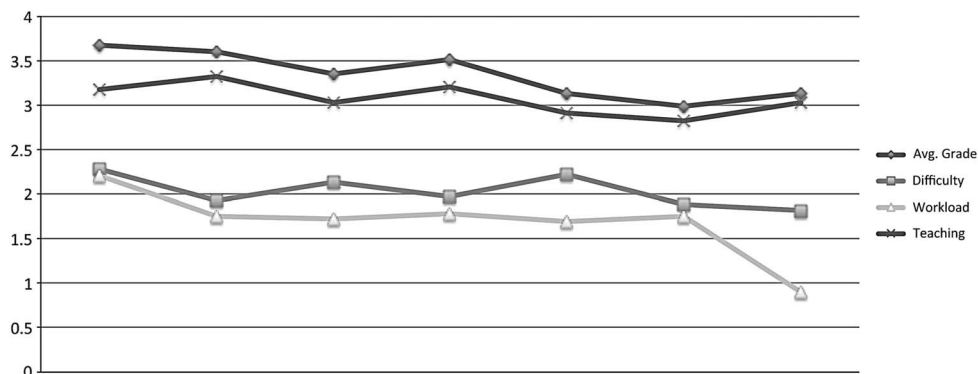


Figure 1. Digital photography selected course survey data.

quantitative component of five years' worth of student course evaluations. The reader may suspect that this is motivated perhaps by some desire for revenge against the website *Ratemyprofessors.com*; however, I can offer assurances that my evaluations are in fact not below my university's average. As I have added research into pedagogy as part of my overall disciplinary practice, partially in response to the ever-increasing demands of measuring learning outcomes in a university environment coming ever more under the influence of the audit culture, I have gradually discovered that course evaluations initially present themselves as information when in fact, under further analysis, they appear to be more like data in need of formal interpretation methods that would turn them into more meaningful information.

Let us take, for example, a chart – as sonified data, suggestive of a building crescendo of two voices or timbres – showing the parallel development in the rise in both GPA (grade point averages) and post-secondary tuition fees at the University of Michigan.¹ While this is thought provoking, it has more often been speculated in the research literature and popular media that the primary driver of grade inflation in postsecondary education has been the administrative use of student course evaluations in order to scale and determine instructor merit and pay increments. Since I have more control over the grades I give in my classes than I do the tuition fees my students are charged, I have tested the so-called 'leniency thesis' (correspondence between teacher ratings and student grades) in my own course evaluations. In the data set here being utilised for purposes of sonification, I related the mean or average grades given in all of my courses over a five-year period against the scores given on all survey questions in the student evaluations, with of course a particular interest in the 'summative question' (i.e. the

instructor's overall perceived teaching ability). For example, in my digital photography class, mapping some of the survey questions in relation to the average grade over a five-year period yielded the graph shown in Figure 1.

From a sonification perspective, the course shown in Figure 1 suggests a monkish or medieval 'unison melody' between the students' ratings of my teaching and my grading of the students' projects, since they progress almost perfectly in parallel, with a nicely contrapuntal middle voice (the survey question on perceived course difficulty) complemented by what is mainly a constant drone in the bassy registers of the course workload survey question that has suddenly dipped south at the last minute (due to a major course redesign in my most recent iteration).

Over my five years of teaching at Simon Fraser University, it appears that I have been deflating my grades with respect to the overall trend of grade inflation typically noted in both the mass and scholarly media (Figure 2). Compositionally, this can be represented metaphorically as both progressive motion (professional development, stricter grading), and a potential downward movement (in student appreciation).

The reader will be disappointed to know that, since this is a journal article, it is not possible, in fact, to hear the sound of my grading (though it will be released digitally in the near future – please refer to updates on my personal website for this Peircean 'actual occurrence').

The aesthetic field of data sonification described in Figure 3 is indeed of practical benefit to this project. My superego Fritz, of course, is shaking his head, adamant that the data remain encased and unperceived by human sensory apparatus. Snow, on the other hand, is reminiscing of pleasant unison chants in stepwise melos as the student grades and my teaching ratings proceed linked together and apace. However, if a more brooding Romantic counterpoint is decided upon, I could focus the composition only

¹This chart can be seen at <http://usactionnews.com/2012/07/dumb-dumber-inflated-gpa-and-tuition>.

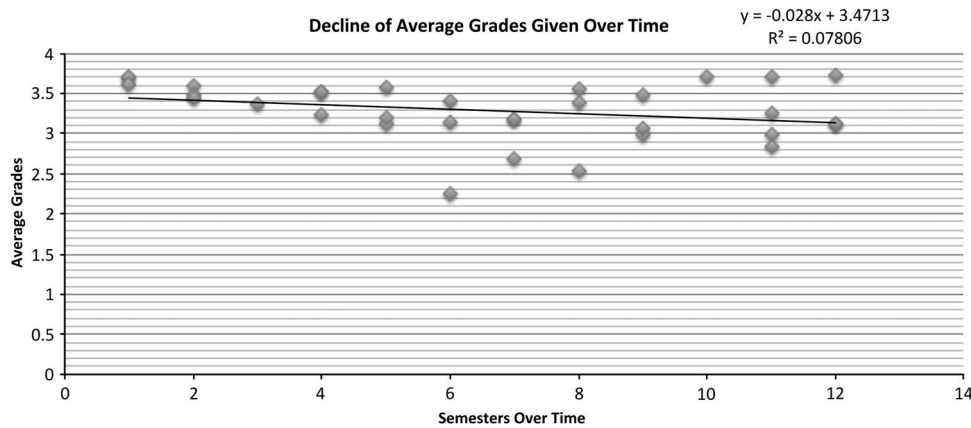


Figure 2. Grade deflation (slope of lowering grades over a five-year period).

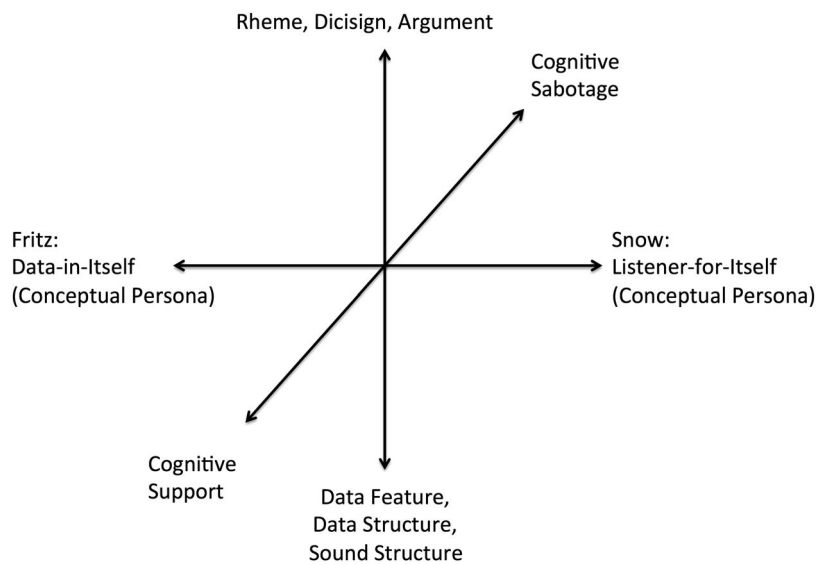


Figure 3. An aesthetic field for data sonification.

on the inverse relationship between poor teaching ratings and low grades given, which would yield sequences of contrapuntal inversion. No doubt happy moments in the surveys should be rendered in consonant harmonies while the deflationary moments tilt toward dissonance. Five years is a long time, however, so I have decided not to pursue the cognitively sabotaging memory-length strategy of a five-year-long composition, and most probably the final form will exhibit a structure close to low information and high redundancy. The pointillistic character of the data (mean scores on a four-point scale) suggests the tone space of temporal-octave division rather than spectral evolution. In deference to Fritz’s spirit of pure data I will score the work in Max (a slight lean toward Snow here). However, I have reconsidered the previously stated use of harmony (consonance and dissonance). That is too old fash-

ioned, or at least too ‘old school’. I have settled on whole tones. Lacking a tonal centre or chord-dominance of any kind – both of which require significant logic gating – whole-tone scales are the easiest to score algorithmically and thus map perfectly to the overall sound of grading.

4. DATA VERITÉ

An aesthetic space constructed of dialectical tensions suggests that perhaps the most ‘successful’ sonifications are to be found toward the centre of the construct, since dialectics employed creatively involves negotiating and maintaining tensions rather than tilting toward the implied forms of one-sidedness. To illustrate the productivity of this approach, it may be useful to provide three examples of works, two of which can be positioned toward unsatisfying extremes, and one of

which seems exemplary of the kind of sonifying practice ‘promoted’ by the aesthetic space.

For an example of Snow producing too pleasant and familiar a musical experience out of the raw alien stuff of the cosmos, we might consider the attempt by researchers at CERN to translate the Higgs Boson into music. Various instrumentations of the so-called ‘God particle’ can be found online at the GÉANT website. A comforting, schmaltzy, almost lullaby variation is to be found on Soundcloud posted by user ‘andycowling’ featuring a slow soothing acoustic guitar plucking away a rendering of the music score posted by CERN.²

Here the sonification is presented in terms of familiar musicality, easily grasped by everyday music listening habits, and encoded in instrumental formats of little semiotic challenge. This of course is not to speak ill of the good folks who work hard to ‘visualize the invisible’, to use Don Ihde’s phrase (2009: 45). With large amounts of public money come public relations responsibilities, and if pleasing melodies of the Higgs Boson galvanise public support for science, these sonifications should have our social blessing (just not our aesthetic enthusiasm). This is not to argue that all sonifications via the music staff are by default unsuccessful in some way, just that this presents opportunities for skewing toward the all too musically familiar, or, as Snow suggests, too much mirroring of our everyday existence. Perhaps an alternate tuning (e.g. micro or just intonation) would benefit the Higgs Boson melody. A work like *Atlas Ellipticalis* is clearly amongst the aesthetic historical precedents of the CERN melodisation, which seems to have been programmed in a somewhat too straightforward manner direct to the music staff, without the kind of awareness of relations to other similar works that is typical of artistic production.

Shifting scales from quantum depths to the signals of the human body, much sonification deals with biometric data, with EEGs being favourite sources for sonification (ECGs seem to have been relatively eschewed, perhaps due to the overuse of ‘scary heartbeats’ as a trope in filmic suspense, or the over-familiarity of the heart monitor as a sound effect in hospital scenes, or even the doctoral stethoscope and muddy mulchy ‘guh-glub-glub’ beat in the routine health checkup). Much biometric sonification indeed has little to distinguish it from other sensor data, such as seismographs or solar flares, since a basic feature of data in general is that, when it is all taken together at once, it is basically noise. Patterns only emerge when certain strands of data are selected out of what William James called the ‘blooming

buzzing confusion’ (1890: 462) which, while initially a description of an infant’s sensory experience, can just as well be applied to an unsorted, uncorrelated, ‘full-on’ data set (every piece of data that is there sounding at once).

For an EEG sonification of some pedigree (presented at no less a venue than the Sydney Opera House), we might consider ‘Untidy Mind – EEG data sonification (2004)’ posted on Soundcloud by user ‘Carnivorous Algae’.³ Like many sonifications, it presents a machinic ‘start/stop’ temporality (not unlike much minimalist music), rather than having any compositional sense of ‘begin/end’, which we can of course ascribe to an inherent limitation of data tables in general. Most complex data sets taken as an undifferentiated aggregate of data points indeed tend to present noise – for example, in reproducing the visual noise of a scatter plot. The *Untidy Mind* excerpt indeed gives us ‘blooming buzzing’ brain signals in a relatively static drone of microsonic variations. The work of statistical analysis is typically to pick out of the total collection of available signals the pertinent few strands of strong correlations in order to support some overall argument or narrative. There are of course other uses of data sets: for instance, a graphing of data points might show a set of outliers that defy the overall tendencies and provoke new research questions and so forth.

Note that *Untidy Mind* is no different from data collected in a techno-scientific context and presented in a relatively ‘unadulterated’ sound format (very far indeed from the Higgs Boson ‘piano, bass, percussion, marimba and xylophone’ found on the GÉANT website). The material presented at the Sydney Opera House could just as well be presented in a laboratory, scientific conference or YouTube illustration of objective phenomena. In other words, there is here no distinction being made between aesthetic and experimental context. This would come close to satisfying the Fritsian superego, subjecting concertgoers to lab equipment, essentially. We should applaud the analytic, experimental and knowledge-discovery potential of this type of sonic representation, since these practices show that research is expanding well beyond ‘visual centric’ modes of scientific inquiry, as Ihde and others have noted. For our purposes, however, the aesthetic space described above would situate these technical practices as a one-sidedness toward data-in-itself, a form of machinic translation which, though presentable (as anything is) in aesthetic contexts (opera house, gallery), is not quite in dialectical tension with the listener-for-itself. For that possibility I will conclude with the third example.

²See www.forbes.com/sites/alexknapp/2012/07/11/scientists-set-the-higgs-boson-to-music.

³<https://soundcloud.com/carnivorous-algae/untidy-mind-eeeg-data>.

Yolande Harris's *Satellite Sounders* offer the listener neither a calming melodyscape nor a barrage of continuous undifferentiated 'flat line' (in the sense advanced by R. Murray Schafer) machinic noise:

In great contrast, the internal combustion engine brought into the soundscape the 'flat line,' which was followed by many other generators of low-information, high-redundancy sound: ventilation, heating, electrical systems and, of course, aircraft. With the increase of these heavy, droning noises, the soundscape thickens into an infrasonic drone. (Schafer 2005)

The fact that much sonified data matches so well the baseline noise-features of almost any industrial equipment should give us pause in being too quick to pursue sonifications that yield only starting/stopping noise drones as a result. Harris's *Satellite Sounders* are handheld portable devices that translate satellite GPS data into sonifications. Latitude and longitude are translated into what Harris describes in the online video documentation as a 'frogging sound, it's like a little frog, it's always much the same, sometimes longer sometimes shorter Then the rest of the sound, which is most of it, comes from the position of the satellites in the sky.'⁴ These two main categories of data (longitude/latitude, satellite position) make for a strikingly affective minimal composition with intriguingly meditative qualities. The resultant sonification occurs within a formal economy consisting of a few sparse oscillator elements undergoing unpredictable though not random variation, which is also productive of many silences (non-soundings). Typically one to three oscillator voices may be sounding at any given time, within pitch registers that translate well within the limitations of headphones worn outdoors (the *Sounders* are used as part of soundwalks Harris has led in various northern European cities). In the terms of the aesthetic space described above, the *Sounders* are not too-easily graspable within the limits of short-term memory (e.g. there is no stressed pulse, downbeat or repeating metre), nor do they produce effects of cognitive alienation and overload (the slowness of the *Sounders*' pitch phrases, the sparse soundings and long silences are conducive of steady reflection and active listening). The electronic tonalities are neither the 'factory sounds' of the CERN melodisation nor the 'flat line' noise of EEG signals, but rather exhibit a retro electronic feel with simple modulations suggestive at times of minimal animal signals (the 'frogging'). There is an objective data source on auditory display, the invisible satellites miles overhead above the sunny sky, but the human pedestrian going on a soundwalk is also attended to, particularly

since the sporadic tones and silences of the sonification allow for the easy bleed-through of street-level ambient sound, which acts as a second layer to the composition, juxtaposing cosmic and human scales. The Peircean distinctions of sound quality, sound occurrence and sound relation are clearly differentiated and identifiable in this mix. Finally, the notion of *imago mundi* (a non-egocentric 'world image') is well in effect in this work, since the *ekstasis* of being taken 'outside oneself' is produced through a headphonic intimacy which brings the distant satellites close to the listener through their intermittent 'frogging' against the street noise. Furthermore, the portability of this sounding GPS receiver thematises the listener-in-motion, whose street wandering movements contrast the orbits of satellites coming and going in and out of 'view' of the GPS antennae.⁵ In short, the work is data-centric and listener friendly simultaneously, keeping the dialectical tensions described above in play and in tension.

Recalling that for Peirce signs can relate to objects, to themselves and to interpretants, we might make a last gesture toward 'the newest thing' in data sonification, namely the relation to so-called 'big data'. I would argue that with big data these concerns become even more acute than they have been in the past. In general terms we can say that the form of data ranges from the table (e.g. a spreadsheet, rows and columns), to the database (a collection of tables, typically relational or 'fixed scheme' in organisation), to the 'volume, velocity, and variety' (O'Reilly 2012: 4) of ubiquitous unrelationally unorganised data capture, storage and streaming. If the aesthetic 'temptations' of sonification are to veer toward the poles of either noise or comfortably familiar music models, big data promises yet more noise, if not much more difficulty in achieving standard musicality. In other words, what was above called data feature and structure in the aesthetic space has moved beyond the paradigm of the table or relational database into a world that is much more 'untidy' than the dataverse of *Untidy Mind*. 'The process of moving from source data to processed application data involves the loss of information. When you tidy up, you end up throwing stuff away. This underlines a principle of big data: *when you can, keep everything*' (italics in original text, O'Reilly 2012: 7). 'Big data is messy' (O'Reilly 2012: 9).

To invoke a cinematic style, we might recommend that the sonifier of big data utilise the aesthetic space in a manner that we could call 'Data Verité.' The correlations between classic cinema verité practice and current data reality are perhaps not immediately apparent, but if we note that for verité filmmakers the

⁴See http://yolandeharris.net/?k_work=sun-run-sun-satellite-sounders.

⁵See www.nimk.nl/air/sunrunsun/index.php?m=05&y=08&entry=entry080525-193328.

mantras were ‘shoot first, ask questions later’ and ‘verité was wanting what you got rather than getting what you want’ (Wintonick 2000), there is a pertinent analogy to be drawn with the new trends of ‘capture everything, analyse later’. Or, as data analyst Alistair Croll states the current condition, ‘With the new data-is-abundant model, we collect first and ask questions later. The schema comes after the collection’ (O’Reilly 2012: 56), which is an almost exact restatement of the verité ethos of figuring out the story in the editing room, well after the shoot (in the case of verité the reels of footage are the data). I would like to propose that perhaps a kind of ‘data verité’ would be an appropriate strategy for the sonification of big data. Verité respected the ‘raw data’ of the material in its documentary methods of capturing people in action, yet these filmmakers were also not adverse to staging actions at times – for instance, in asking Stravinsky and Nabokov to sit down, drink vodka and speak Russian to each other in front of the camera in *A Stravinsky Portrait* (1966). Verité mixed the enthusiasms of having no plan in advance with collecting vast amounts of footage without concern as to how it might all fit together into a coherent story, and when conditions were right staged a scene or two when it seemed acceptable to do so (bending the rules of documentary for cinematic pleasure). The finished film would come to find its ‘internal logic’ after the fact of shooting, in the editing suite, in the search for meaningful patterns worthy of presenting out of the messy collection of time-based captured reality. In other words, Fritz was given leeway in the shoot, Snow in the edit, and the result could be understood as being in aesthetic tension between these pulls toward potential one-sidedness in the realisation.

It is interesting to note that while the practices of visualisation utilise a range of terms such as ‘data visualisation’ and ‘information visualisation’, sonifications almost always refer to data, not information (the former term yields 21,700 Google hits versus 4460 for the latter, using quotes around the words). The term ‘information sonification’ is almost never used; in the academic literature, ‘data sonification’ returns 109 hits on my university library’s website, whereas ‘information sonification’ returns 11 (again using quotes around the words). The parallel ratios when it comes to visualisation

Term:	Google	SFU Library
"data sonification"	21,799	109
"information sonification"	4460	11
"data visualization"	3,800,000	12,655
"information visualization"	1,020,000	5263

Figure 4. Data versus information sonification and visualisation in Google and academic searches.

are 3,800,000/1,020,000 and 12,655/5263 respectively, summarised in Figure 4.

While it is beyond the scope of this paper to fully analyse why sonifications ‘skew’ toward the tropes of data rather than information, we can perhaps reasonably surmise that this occurs because of the traditional association of sonic practices with conditions of music and sound art which eschew verbalisation, overt meaning-making, communicating and so forth, preferring to remain in a more ‘abstract’ (to use a word within easy reach) modality toward the listener, whereas with visualisation there is much social impetus toward connecting the visual with the verbal in general. To paint the distinction in broad strokes, we can say that ‘data’ relates to the project of pattern recognition, whereas ‘information’ deals with the higher-level inferences (arguments and stories) that we fashion out of the found patterns. As Seth Kim-Cohen has noted, ‘In music, and in what later came to be known as sound art, there is an evident resistance to questioning established morphology, materials, and media. There is a sense amongst practitioners and theorists alike that sound knows that it is: sound is sound.’ (2009: xx). ‘Pursuing the acousmatic epoché, we are then responsible for bracketing out all information that might shade our auditory experience with signification, with historical contingency, with social import. From this reduction, we can identify that which, within sound, simply *is*’ (Kim-Cohen 2009: 13, italics in original). Thus a further value of employing semiotic parameters with respect to sonifications is that they open the door to further considerations as to what ‘information sonification’ may be relative to ‘data sonification’, addressing listeners who may wish to extract inferential information from pattern perception.

There may be a tad too much ‘humanism’ (for some sensibilities) in the approach described above; however, as far as we know, it is not the machines who are listening to us.

REFERENCES

- De Waal, Cornelis. 2001. *On Peirce*. Indianapolis: Indiana University-Purdue University Press.
- Deleuze, Gilles and Guattari, Félix. 1995. *What is Philosophy?* New York: Columbia University Press.
- Filimowicz, Michael and Stockholm, Jack. 2010. Towards a Phenomenology of the Acoustic Image. *Organised Sound* 15(1): 5–12.
- Ihde, Don. 2009. *Postphenomenology and Technoscience*. New York: State University of New York Press.
- James, William. 1890. *Principles of Psychology*. Cambridge, MA: Harvard University Press, 1981.
- Kim-Cohen, Seth. 2009. *In the Blink of an Ear: Toward a Non-Cochlear Sonic Art*. New York and London: Continuum Press.

- Lem, Stanislaw. 1955. *Solaris*. New York: Berkley Medallion Books, 1971.
- O'Reilly Radar Team. 2012. *Big Data Now: 2012 Edition*. O'Reilly Media Inc. ebook at <http://oreilly.com/data/radarreports/big-data-now-2012.csp>
- Peirce, Charles Sanders. 1906. Prolegomena To an Apology For Pragmatism. *The Monist* 16(4) (October): 492–546.
- Peirce, Charles Sanders. 1955. *The Philosophy of Peirce: Selected Writings*, edited by Justus Buchler. New York: Dover.
- Pure Data. n.d. Pure Data Development Wiki, <http://puredata.info/dev>.
- Schafer, R. Murray. 2005. I Have Never Seen a Sound. Twelfth International Congress of Sound and Vibration, Lisbon, Portugal, July, 2005. Online at <http://www.arch.ksu.edu/seamon/Schafer06.htm>.
- Snyder, B. 2001. *Music and Memory: An Introduction*. Cambridge, MA: The MIT Press.
- Tenney, James. 1992. *Meta-Hodos/Meta Meta-Hodos: A Phenomenology of 20th Century Musical Materials and an Approach to the Study of Form*. Lebanon, NH: Frog Peak Music.
- Wenders, Wim. 1994. *Lisbon Story*. Production Company: Road Movies Filmproduktion/Berlin.
- Wintonick, Peter. 2000. *Cinéma Vérité: Defining the Moment*. National Film Board of Canada documentary.
- YouTube. n.d. 'Sonification of Tohoku Earthquake', <http://www.youtube.com/watch?v=3PJxUPvz9Oo>.