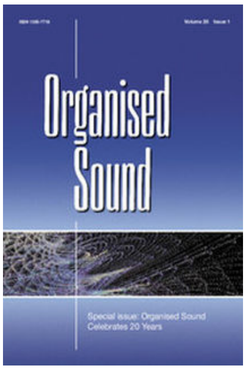

Listening *With* Machines: A shared approach

Simon Emmerson

Organised Sound / Volume 20 / Special Issue 01 / April 2015, pp 68 - 75
DOI: 10.1017/S1355771814000442, Published online: 05 March 2015

Listening *With* Machines: A shared approach

SIMON EMMERSON

Music, Technology and Innovation Research Centre, De Montfort University, Leicester
Email: s.emmerson@dmu.ac.uk

The aim of this article is to review the last twenty years of 'machine listening'[1] to sound and music, and to suggest a balanced approach to the human–machine relationship for the future. How might machine listening, and MIR[2]-based ideas of data storage, retrieval and presentation enhance both our embodied experience of the music and its more reflective study (analysis)? While the issues raised may be pertinent to almost any music, the focus will remain on electroacoustic music in its many forms, whether for interactive composition, performance or analytical endeavour. I suggest a model of listening *with* – that is, alongside – machines in such a way that our skills may be enhanced. What can we share with machines to mutual advantage?

1. *MACHINE LISTENING* – 'HOW TO TELL IT WHAT'S GOING ON?'

Over the twenty years that this issue of *Organised Sound* celebrates, the world of 'machine listening' has developed steadily. Initially driven in most part by the rise of *interactive systems* for performance and improvisation in the 1980s, there was pioneering work establishing the field by, for example, Robert Rowe (1993) and George Lewis (2000). Machine speeds meant a first generation of 'listening' in real time was generally confined to symbolic following (MIDI). But tools for analysis of signals were of course developing rapidly and were increasingly integrated into machine listening systems: Collins (2007: 178) distinguishes these as 'symbolic' and 'sub-symbolic' information systems. This is not the place to describe the burgeoning of interactive music systems in this period.[3] All to a greater or lesser extent now demand real-time machine listening to an audio (and maybe a MIDI) stream.

Is the notion of listening any different to what *any* human/machine interface for music does? Although a fuzzy boundary, performance transducers – that is, physical gesture interfaces - have generally been excluded from this discussion. Listening after all usually means to a soundstream – sometimes extended to include the symbolic domain of MIDI - but it could be that in the long run this distinction will dissolve and any sound interface be deemed a 'listening device' in some senses.[4] But listening *for what*? To extract information relevant to a defined task. This runs in parallel to changes in paradigm concerning perception itself – no longer considered the encounter of an undifferentiated stream which the mind tries to 'make sense of' but an already information-rich stream which human perception is ready equipped to interpret[5] – our listening machine needs to have some idea in what form to encode and store the apparently endless data in a way that can be searched efficiently.

In parallel to the development of listening interfaces for interactive electronics, in the same period the world of 'big data' has come upon us – and with it the challenge of handling data volumes and streams too large for any individual human to grasp and comprehend. Even pre-Internet digitisation was generating such volumes and information science developed rapidly in response, but as the libraries of information became potentially (and really) global the need for machine assistance was immediate and unarguable. We have become increasingly overwhelmed by the sheer volume of what is now available in the digitised archives connected by the Internet – we have witnessed a fast forward in evolution and risk becoming immobile in the bright headlights of this new enlightenment. This has of course driven the growing MIR (*music (and sound) information retrieval*) community, which has harnessed machine assistance to seek, sort and represent (visualise and display) information of some use and comprehension to the user.[6]

While the initial drive was indeed 'retrieval' from existing vast databases, this has increasingly developed a 'front end', the harnessing of the machine as an analytical assistant to generate the relevant data and to store it in a way it can be retrieved.[7] Until recently the

[1]'Machine listening' is a metaphor. Machines search out data from which information may be inferred. How this does or does not parallel human listening will be part of the present discussion.
[2]'Music information retrieval': the term is based on the assumption of a need to search 'big data' for information relevant to a specified task. This describes the aims of a developing research community.
[3]Winkler 2001 and Dean 2003 cover a range of options.
[4]This may go further. Why exclude video or any other perceptual mode? In the end this will mean machine *sensing* not simply listening.
[5]This ecological view of perception is well explained for musicians by Luke Windsor (Windsor 2000).
[6]Casey, Veltkamp, Goto, Leman, Rhodes and Slaney 2008 is a very useful summary of the state of play and 'future challenges' only partly addressed in the subsequent time.
[7]The analytical engines were always there of course but it is clear from recent literature how these two areas have now joined up.

Organised Sound **20**(1): 68–75 © Cambridge University Press, 2015. doi:10.1017/S1355771814000442

term 'metadata' has been used to describe *verbal* information 'tags' *about* the (for us) stored sound. In addition, analytical engines have increasingly generated vast amounts of 'pure data' especially when analysing sound signals – for example, timbral similarity measures, auditory stream segregation, or onset and duration information. But, increasingly, 'music' level information may also be generated or mined – for example, melodic transcription, tonality and polyphonic pitch information, and pattern recognition of all these (Casey et al. 2008).

Clearly, if the user defines the criteria for 'similar' sound qualities then many sounds may be searched out by machine. A good example of human–machine collaboration is given in a system created by Arne Eigenfeldt and Philippe Pasquier (2010) to harness fast machine search of an audio sample bank to increase choice possibilities for the live performing musician. Here, the similarity ideas from other MIR software developments are harnessed to serve the needs of a performer, presenting options which would not be possible in the absence of the machine, and using (relatively) simple visual display to allow the choice to be made in real time.

A glance at recent ISMIR annual conference proceedings (www.ismir.net) suggests a wide range of applications from looking at structures within western tonal music (classical and popular), to extracting information from the recorded music of aural cultures. This community (as we might expect) crosses over strongly with the NIME (New Interfaces for Musical Expression) group (www.nime.org). Mital and Grierson (2013) have developed an MIR system to mine a database of the work of Daphne Oram and present results in a 3D interactive visualisation. This combines all the historical streams of research discussed above, generating data, retrieving, organising and visualising using newly developed screen presentation techniques.

One approach to data mining has been termed *soundspotting* – tracking down sounds through defined criteria, usually similarity to a source target (Spevak and Polfreman 2001). Many versions of this approach tend to address problems of searching a global jukebox for music of a chosen nature. Michael Casey has developed a creative version of this in his *SoundSpotter* application intended to create a real-time soundstream dealing with shorter sound segments – in effect a composition system where 'similarity' (he stresses this should be perceptual, not just any measurable parameter) becomes part of the creative play (Casey 2009).

2. PROSTHESIS OR INDEPENDENT AGENT?

In discussing how machines and humans interact I have suggested two distinct paradigms of human–machine relationship (Emmerson 2009: 168–71). The first extends our (musical) reach, the expressive possibilities of our physical and mental 'event space'. In this view the 'deep' history of our technology in the world has been about *prosthesis*, both physical and mental – we want to stretch out and contact others way beyond the boundaries of our immediate senses. From memory to writing, then to communications technologies, machines have allowed us to offload and store first plain data, then sound and video, and now processing (and perhaps eventually thinking itself).

The second has a different emphasis – the machine as a clone 'other', eventually acting and interacting with us indistinguishably from any other human agent. At first this is more easily conceived if we think of the agents as remote from each other (as in Alan Turing's original model) – say, operating over a network. Behaviour is limited to interaction through text exchange and interaction.[8] But the technology of robotics maintains physical action in the world; those which aim directly to resemble humans are sometimes referred to as *androids*. The most famous recent example in musical terms is probably the *WF-4* series from Waseda University (Japan), which seeks literally to model the entire physical mechanisms of the flute player (Solis, Chida, Taniguchi, Hashimoto, Suefuji and Takanishi 2006). In the same issue of *Computer Music Journal*, however, we find *Haile* (Georgia Institute of Technology), which deliberately possesses *superhuman* performance options within an anthropomorphic design (Weinberg and Driscoll 2006). Collins reports on robots programmed to show 'affective (emotional) responses' (2007: 171, and footnote 1) and a new generation of 'companion' robots has recently been launched.

Of course these two approaches have not developed apart and perform a sometimes unbalanced dance with real human life forms. I intend to discuss this distinction with respect to *listening* both creatively and analytically. I believe we may be led into making assumptions about the role of machines which can be misleading. In brief, the question becomes: are we developing new approaches to listening and understanding? And how might machines best assist in the task?

3. ANTICIPATE AND IMAGINE

Listening is not so much about what is happening now as what is likely to happen next. The human system is built to anticipate the future, to be ready to act (London 2004). The embodied decoding of the current flow of sound includes the special case – when events are periodic - of synchronisation (referred to as *entrainment* or *attunement*). At its most embodied this

[8]At the time of writing (June 2014) the popular press is full of reports of the computer program *Eugene Goostman* (Reading University) 'passing the Turing test'.

results in so-called mirror neuron activity in the listener that appears to mimic or follow the action we suppose provoked the sound – we don't just think along, we may hum along, move limbs to the beat, play air guitar. Performing embodies 'what next?' strategies. Collins suggests that more generally we create an *expectancy field* ('a probability distribution for future events calculated from recent ones' (2007: 181)), which affords us a range of anticipation possibilities and is thus fundamental to performing with others. There is a non-real-time version of this kind of skill which is much used in electroacoustic music (studio or live). Anticipation of an action may give way to imagination of a sound – a kind of internal representation which many of us claim is almost as vivid as the real thing – and certainly as vivid as memory of the real thing. This we might in some contexts describe as 'being creative'.

4. THE ANALYSIS OF ELECTROACOUSTIC MUSIC

The analysis of electroacoustic music has been relatively slow to develop and has remained separated from these developments until quite recently. There are many reasons, including the lack of an agreed visualisation process, difficulty with deciding on 'meaningful units' (salient features), as well as its position in a group of experimental arts practices without an agreed syntax – and with very little relationship to previous practices with an established analytical history. So it seems an apparently natural development to harness machines to assist in the examination of such a technology-based practice.

A recent example of MIR-based electroacoustic music analysis is the ongoing EASY/SQEMA project led by Tae Hong Park. This is at present a two-part process. Park takes a classic 'pyramid' approach to parsing the sound flow, teaching the machine (through the EASY toolbox) to detect different levels of boundary, from sections to sound events. This information is passed to SQEMA through 'multiple listenings', through several layers of analysis, culminating in 'aesthetic analysis' (Park, Hyman, Leonard and Wu 2010; Park, Hyman, Leonard and Hermans 2011). Related to MIR concerns, Dean and Bailes (2010) apply machine analytical techniques to a wide range of electroacoustic musics (both studio created and improvised works). Here the task is one of extensive measurement and data processing (in this case, of rise–fall amplitude characteristics). Yet the discussion of the results remains one of embodied human interpretation: 'We suggest that the electroacoustic creator is probably carrying over the ideas of the instrumentalist's force-energy input to the music judged by listeners as effort-loudness-affect' (2010: 154). It is this kind of alliance of the machine with human judgement I wish to explore further.

A very creative approach to this alliance challenges the separate roles of listener and composer, encouraging the user interactively to engage with the sound work. Michael Clarke's interactive analysis of Denis Smalley's *Windchimes* was a major landmark (Clarke 2010).[9] The accompanying software allows the user to 'recreate' sections of the work, to mimic compositional decisions, changing parameters to hear how that influences the outcome. While in this case much of the analysis is fixed by the author, subsequent developments indicate an increasing flexibility, with current work at time of writing suggesting that even more radical intervention by the user is possible (Clarke, Dufeu and Manning 2013). I shall examine other such human–machine alliances after discussing in more detail the changing role of listening.

5. HUMAN LISTENING IN THE AGE OF THE MACHINE LISTENER

More specifically I want to focus on the relationship of such a key skill – listening is fundamental to human evolution – to the emerging machine-centred paradigm and *its* evolution. Not so much the question 'can machines listen?' but more how human listening may be influenced by an alliance with the machine. I am suggesting the need for a carefully worked out relationship – one in which each party understands the other. In an early draft of this essay I wrote 'Exactly what do we want to delegate to the skill of the machine?' – but this illustrates a misunderstanding. Put in a more positive way we might better ask: 'How can the machine enhance human skill and understanding?'

5.1. The modes of the listening machine

There are many discussions on listening in the literature from physiology, psychology, psychoacoustics, cognitive science and of course the sonic arts (Handel 1993; McAdams and Bigand 1993; Bregman 1994; Smalley 1996; Clarke 2005; Smalley 2007). I intend the following discussion not seriously to imagine that machines can indeed listen in the ways that humans do, but to use such an assumption to interrogate our use of language, and more importantly to see if the advent of machine listening changes this use. Let us start by seeing what our machine might do for Pierre Schaeffer's four listening modes (Chion 1983: 25–6).[10] The first two seem to map well onto what is currently their main function.

[9]See also Clarke (2006) for earlier work based on Jonathan Harvey's *Mortuos Plango Vivos Voco*.
[10]The first two are discussed by Schaeffer in reverse order (*écouter*, *ouïr*) because of his construction of a 2 x 2 binary table dividing the modes into abstract/concrete and subjective/objective (Chion 1983: 25–6).

5.1.1. Ouïr

This is the ability to register the presence of sound – to 'know it is there' even when paying no particular attention to it. It has sometimes been assumed to be a dumb function of mechanical 'fact': the A-to-D converter and associated memory tools would be its simplified equivalent. But we should be careful – the human system is effectively always on the lookout for threat or opportunity. Other listening modes are never fully switched off, such as the next one, *écouter.*

5.1.2. Écouter

This is everyday listening for survival, the search for basic information in the sound stream: What? Where? Who? How? It appears to map well onto the new technologies of data analysis and mining. We need to define how we 'teach' the machine to recognise specific sounds – there are many models, pattern matching, neural network learning and so forth. We must note that in this mode the search includes the discovery of source and cause, not just sound quality.[11]

More recent ecological models of perception do not separate these two to such an extent (Windsor 2000) and we shall attempt to refine them with ideas from Barry Truax and Katharine Norman below. The second group of two listening modes move towards the world of *aesthetic* listening through *reduced listening*.[12]

5.1.3. Entendre

This is the attention to specific sound qualities with an increasing focus on traits and characteristics that are *interesting* and potentially *useful* (salient features but with an aesthetic potential). Schaeffer defined further criteria to create an extensive taxonomy of sound types and behaviours – typology and morphology (Schaeffer 1966).

Our machine will need help here – extraction of data with specified features is clearly an elementary function of feature matching – there needs to be definition of salient data to be searched for or an analysis process to create such data. The user will define a subset to be searched for and reported back in some sort of comprehensible form. 'Find me a set of sounds with features A, B, … X' will be quasi-objective. In this sense, for the machine, *reduced listening* is simply a different set of salient features – one which does not try to identify sources and causes, focusing on acoustic and psychoacoustic properties alone. Thus a Schaefferian typology data set would be (in principle) simple to construct. But, that said, Schaeffer's programme is at the service of something more creative: features of the sound which might be described as 'interesting', 'having potential' (even 'beautiful')[13] will be more problematic to define. (We shall discuss this further.)

5.1.4. Comprendre

This is the comprehension of a sequence of sounds according to criteria relevant to their type and organisation – for Schaeffer this is an aesthetic (musical) understanding quite separate from real world causes, one discovered by experience and through the ear – what *sounds right*, what *makes sense*. This is Schaeffer's 'search for a language' – (a *solfège*)[14] (Chion 1983: 90–4).

5.2. Refinements: Truax and Norman

Before exploring Schaeffer's move towards a *comprehension* of sound I wish to add in two alternative views of listening that do not always value the *bracketing out* of the real-world associations of sound – in fact, rather the contrary.

Barry Truax (1984) distinguishes three levels of attention within listening that span Schaeffer's first two listening modes:

- *Background listening* – the awareness of sound is present to our consciousness (and may persist for a while in memory) but has no specific impact or significance.
- *Listening-in-readiness* – specific significant sounds, if and when they occur, will provoke conscious attention (even if we are asleep).
- *Listening-in-search* – we consciously and continuously listen out for specific sounds, characteristics and qualities.

The examples Truax gives (1984: 16–22) all come from listening to the sound-world around us with respect to the information it might carry, hence this final designation (listening-in-search) is clearly a subset of Schaeffer's *écouter*. But the idea of *search* may be extended: when sources and causes are bracketed out (*écoute réduite*) we may move forward to the more aesthetic *entendre*, searching out the sonically interesting features. While Schaeffer believed this to be an inevitable outcome of the recorded medium (affording repeated listening), a wider view might include listening to the world's sounds as they occur. This is the basis of *soundwalks* (Drever 2009) and Denis Smalley's detailed and meditative reflections on space (Smalley 2007), as well as Cage's famous remark 'My favorite piece of music is the one we hear all the time if we are quiet.'

[11]There is a substantial literature on this from the MIR community referenced throughout.

[12]Bracketing out the source and cause of the sound (Chion 1983: 33–4).

[13]All these terms are impossible to define clearly – 'potential for development' presupposes an approach if not a method.

[14]Thus (for example) jazz, gamelan, classical symphony may be comprehended in terms of their particular (and different) modes of organisation ('codes' – based on conventions and practices). Similarly that for musique concrète must be discovered ('searched for' in Schaeffer's terms).

Katharine Norman (1996) stresses the multi-layered nature of listening and suggests we employ different strategies dynamically.[15] Her initial distinction of *referential* and *reflective* listening are then both enveloped within *contextual* listening, which is grounded in 'an amassing of individually experienced knowledge, that extends beneath all our new experiences to influence and constrain our perceptual direction' (1996: 8). Memory and personal experience are thus crucial parts of what she terms *composed listening*. This poses a real challenge for the machine listener. We could reduce the argument to an absurdity by simply ruling this out – we could say that it is simply impossible for a machine to have the kind of personal memory bank and associated 'meanings' of an individual feeling being. But there is one line of escape. Norman does not argue for a solipsistic individualism where all listeners are somehow locked into worlds of personal memory and resonance. She argues – almost in archetypal terms – that engaging with personal memory opens up greater possibilities of generating meanings for all listeners.

We may share more than we know. Of course the resonances of memory may depend on social and family circumstances, urban and rural environments, and many other individual biographical events. But in the end the shared heritage of sound may be no more inclusive or exclusive than the practice of any genre of music. Our machine might then be taught to search out likely candidates for not just an archetype (which suggests cross-cultural reference) but a more personal set of sound options.[16]

6. MACHINE-LEARNED ACOUSMATIC PASTICHE?

For our machine we are in a realm new for acousmatic music but well trodden for tonal music since the earliest days of computer applications (Cope 1991). Style analysis and pastiche composition have been its mirror twins – if listeners cannot tell the difference between a Bach chorale pastiche and the real thing then the machine creator has passed a kind of musical Turing test.

Now we face the greatest dilemma yet in interpreting this simple machine mapping. Schaeffer's rules for musique concrète are strictly empirical – discovered through practice and listening, and not invented *a priori*. Our analytical machine may observe: 'it appears that human "Parmegiani" tends to choose[17] sound quality X transforming into sound quality Y under condition Z etc.' – maybe elaborating the observation using a multi-dimensional Markov transition probability matrix which has been built up from analysis of many instances of transformation within Parmegiani's work. This is in the first instance descriptive of the 'real' Parmegiani. But we are thus creating potential *generative rules* from this empirical evidence – just as in the tonal example – which may rapidly become *a posteriori* a *Parmegiani generator*. But that is the nature of pastiche - and debates on the nature of 'machine comprehension' (and originality) are clearly bound up with this dilemma. While the possibilities for machine-created taxonomies are advancing, Schaeffer's 'comprehension' may be more elusive: it will remain a matter of debate to what extent we will be able to teach a machine what is a 'good piece of music' or even a 'good sequence of sounds'.

Using this kind of analytical model we are locked into a behaviourist paradigm. If our pastiche generation produces something apparently indistinguishable from the original we can never make firm statements that these procedures are anything near the likely choice mechanisms used by the original composers (Bach or Parmegiani). Of course creative computational strategies may in time be developed which make such a claim. Geraint Wiggins, Pearce and Müllensiefen argue that some computational models do indeed aspire to model real cognitive processes eventually 'to help us understand how human composers work' (2009: 383).

7. WHAT IS IT 'TO BE PRESENT'?

Our machine listener risks introducing a potentially dangerous ambiguity here: what is it for something *to be present* in the act of listening? The ambiguity lies in the complex, evolving and interactive relationship between two simple notions of *being present* – what is *measured as present* and what is *perceived as present*. If we are going to discuss this relationship there emerges a further complication. How do we grasp the essence of either in a form that can be understood by us (humans)? Data representation, whether from measurement (machine) or from a kind of evocative transcription (human perception) is itself a form of filter, inevitably incomplete and aimed at performing a specific task. Both have historically focused on *visualisation* as a tool, at first by hand[18] but increasingly using machine assistance.[19] This has become a crucial level of human interaction with the machine as assistant: to what extent can the user *see* represented as well as *make sense of* the features, formations, processes and organisation of the music? What exactly

[15]'Our ordinary listening is itself a complex, multi-layered activity of which hearing is but a part. In going about our everyday listening lives we take – I suggest – several different, but interdependent, stances, which amount to a dynamic construct.' (Norman 1996: 2)
[16]Witness the crude attempts of current search and sales engines to profile our likes or needs – imagine that forward twenty years.
[17]Is that the same as 'like'?

[18]Wishart 2012 includes a beautiful collection of Trevor Wishart's hand-drawn diffusion scores.
[19]The *Acousmographe* (GRM, Paris) and *EAnalysis* (Pierre Couprie, De Montfort University) are current examples.

is being visualised, suggested and eventually reified, and how does this influence what is claimed to be *heard*?

7.1. Measurement, representation and evocation

The mapping of physical feature to perceptual entity has long been the matter of much discussion. It is broadly accepted that FFT-based spectral information does not well represent sound quality. The developing use of the *spectral centroid* has helped us to grasp 'brightness' as a measure but it is by no means perfect. Human perception retains a head start over machine strategies for auditory stream segregation which has only slowly arrived and is only now coming closer to the real time of human perception.[20] The mapping has at least two components: identification of what physical features contribute to the perceived event, and then how to present these to the user.[21] It is clear that for a substantial part of the literature what is present is deemed to be what is measurable.

The degree of detail within evocative transcription depends upon the function of the document produced. From the earliest days of electroacoustic music evocative transcriptions have primarily functioned as mnemonic support.[22] For those that practise active sound diffusion this was usually a timeline with graphics to indicate salient features that need accentuation in the performance; for the listener, a reminder of something already memorised, or to indicate a simple outline with important features to 'listen out for'. Refinements for performance with instruments are likely to be more time accurate.[23] Those for scholarly analytical use also vary from relatively detailed transcription to the micro-detail that might in principle allow the work to be recreated on newer technology.

In summary, this group was usually intended to supplement and enhance a human-centred listening approach. But as we shall see this isolation cannot remain so simply stated. These may ideally combine with – even cross over into - the 'measured' kind of transcription discussed above. There is interaction between the two as well as the influence of memory and repetition. Analytical listening is clearly not the same as listening at a performance – especially when recording affords repeated listening.

[20]And then only really for the simplified aim of 'melody and bass line' pitch tracking used to identify tonal popular song materials (Casey et al. 2008). See also Wang and Brown (2006) for a broader view.

[21]Visualisation is not confined to graphics except in the very broadest sense embracing text, symbols and shapes (Gray 2013).

[22]As were the earliest notations in the western music tradition.

[23]For a classic example of the 'diffusion score tradition' see Mion, Nattiez and Thomas 1982, which includes Bernard Parmegiani's diffusion score of *De Natura Sonorum*; the published scores of Stockhausen's electronic works cover a range of these possibilities (*Studie II*, *Kontakte* (*realisation* and *performance* scores), *Telemusik*, *Hymnen*).

7.2. The dynamic relationship of measured and evocative

Something detectable by the machine is clearly not the same as something consciously perceived, although the relation is not simple. Yet in any analytical method the presence (for example) of particular frequencies in an FFT representation is taken to be an indication that they are 'really there' even if not all are perceived. Conversely such a representation may omit perceived tones generated in human perception but not physically present – missing fundamentals, sum and difference tone effects and so on. Jonathan Harvey famously pointed this out in his description of the analysis of the cathedral bell during composition of *Mortuos Plango, Vivos Voco*. He could clearly hear a middle F pitch which was absent from the measured spectrum and which he described as a 'supernatural attribute' of the sound (Harvey 1981: 22–3). Perhaps we will be able with more recently developed techniques to teach the machine to recognise and display these 'fictitious' tones.

There is also the tantalising paradox that there may be something the machine detects which works on human perception only at a subliminal level; we cannot identify it or separate it out but it has a real effect – and we might detect its absence or, better put, detect a change when it is absent.[24] Also, there may be cases where at first perception we miss features of the sound flux which the machine 'points out' to us and thus on subsequent listening we detect. The machine acts as a pointer, to 'listen out for', certain features. This may often be constructive and helpful. But more ambiguously this may lead to problematic possibilities of 'over-suggestion' – you are told something is there so it 'is' – you 'hear' it. I can never deny you the genuine belief that this is the case – I cannot ever know exactly what you heard.[25]

Thus 'what is present' is both ambiguous and unstable, depending on the interaction of signal, environment, perception, memory and machine. Possibly a core group of features will be stable and held in common but many, especially those which contribute to the 'taste' of a particular style, may be subject to vast variation. That will not of course stop a reciprocal reaction. The evocative perceptual transcription may in turn influence the machine-measured representation – indeed, that is often an avowed aim of newer software developments (Mital and Grierson 2013, *Acousmographe* and *EAnalysis* software).

Research trying accurately to describe the elements of a sound-world using everyday (even emotive)

[24]I am grateful to Pierre Couprie, the developer of *EAnalysis* software, for discussions on the complexities of this relationship (personal communication and comments in his presentations of the software).

[25]A neural nightmare might be with us sooner than I fear: 'The map of neural firings in your brain suggests you did *not* hear that sound quality whatever you say.'

language has been under way for several decades. Starting with verbal descriptors such as 'bright' for spectra, much progress has been made in such a mapping. But searches for 'frightening' sounds might need greater social sophistication in their mapping, especially when a demand for machine-readable terms is made (Casey et al. 2008 summarises such work). As we shift steadily from *sound* (signal) description to *music* description Schaeffer's third and fourth modes show us clearly that the idealisation of a 'complete' description which in principle encapsulates all possible salient features has severe limitations. A salient physical sound feature may not be a salient musical feature.[26]

8. CONCLUSION

Perhaps there will always be a problematic relationship of machine listening to real embodied listening. To give a preliminary indication of the kind of partnership I envisage I will describe a first case study. In the three-year AHRC-funded project *New Multimedia Tools for Electroacoustic Music Analysis*[27] the emphasis remained with the human listening experience completely at the focal point. The role of the software was thus systematically to extend the imaginative graphic externalisation *of the human listening experience*, embracing existing typological descriptive systems and allowing the addition of the user's own. This was intended (in part) to allow greater ability to uncover patterns and processes derived from listening. But as the project evolved the notion of *machine suggestion* arose. Although not fully developed in the initial project this is part of Couprie's continuing developments of the software.[28] The ideal might be described as an intelligent assistant – not primarily analysing the music but (for example) assisting the user to spot patterns in the listening transcriptions.[29]

In conclusion, I suggest an approach that encourages the active collaboration and participation of the user with the machine partner, one that enhances and does not offload the fundamental skills of listening. Care must be taken to keep transparent and open to question the complex and dynamic relationship of *measured as present* to *perceived as present*. It is clear that robotic assistants (real or virtual) will play a key role in the future of music but their job descriptions, which will no doubt change over time, must be clear.

[26]This is another version of the difference between *etic* (measurable) and *emic* (meaningful) features long discussed by our ethnomusicological colleagues (see for example Nattiez 1990).

[27]At De Montfort University, directed by Simon Emmerson and Leigh Landy, with additional assistance from research student Michael Gatt and software development from Pierre Couprie.

[28]For the software download, manual and further information see http://logiciels.pierrecouprie.fr/?page_id = 402.

[29]The degree to which the machine will in future risk *suggesting* such patterns is a matter of current research but not taken for granted as in other MIR software approaches.

REFERENCES

Bregman, A. S. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Casey, M. 2009. Soundspotting: A New Kind of Process?. In R. T. Dean (ed.), *The Oxford Handbook of Computer Music*. Oxford: Oxford University Press.

Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M. 2008. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE* **96**(5).

Chion, M. 1983. *Guide des objets sonores: Pierre Schaeffer et la recherche musicale*. Paris: Buchet/Chastel, English translation by J. Dack and C North available from www.ears.dmu.ac.uk.

Clarke, E. F. 2005. *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press.

Clarke, M. 2006. Jonathan Harvey's *Mortuos Plango, Vivos Voco*. In M. Simoni (ed.), *Analytical Methods of Electroacoustic Music*. New York: Routledge.

Clarke, M. 2010. Wind Chimes: An Interactive Aural Analysis. In *Denis Smalley: Polychrome Portraits*. Paris: GRM.

Clarke, M., Dufeu, F. and Manning, P. 2013. Introducing TaCEM and the TIAALS Software. In *Proceedings of the 2013 International Computer Music Conference (ICMC), Perth, Australia*. San Francisco: ICMA.

Collins, N. 2007. Musical Robots and Listening Machines. In N. Collins and J. d'Escriván (eds.), *The Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press.

Cope, D. 1991. *Computers and Musical Style*. Madison, WI: A-R Editions.

Dean, R. T. 2003. *Hyperimprovisation: Computer-Interactive Sound Improvisation*. Middleton, WI: A-R Editions.

Dean, R. T. and Bailes, F. 2010. A Rise–Fall Temporal Asymmetry of Intensity in Composed and Improvised Electroacoustic Music. *Organised Sound* **15**(2): 147–58.

Drever, J. L. 2009. Soundwalking: Aural Excursions into the Everyday. In J. Saunders (ed.), *The Ashgate Research Companion to Experimental Music*. Aldershot: Ashgate.

Eigenfeldt, A. and Pasquier, P. 2010. Real-Time Timbral Organisation: Selecting Samples Based upon Similarity. *Organised Sound* **15**(2): 159–66.

Emmerson, S. 2009. Combining the Acoustic and the Digital: Music for Instruments and Computers or Pre-Recorded Sound. In R. T. Dean (ed.), *The Oxford Handbook of Computer Music*. Oxford: Oxford University Press.

Gray, D. 2013. The Visualization and Representation of Electroacoustic Music. PhD thesis. Leicester: De Montfort University.

Handel, S. 1993. *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Harvey, J. 1981. *Mortuos Plango, Vivos Voco*: A Realization at IRCAM. *Computer Music Journal* **5**(4): 22–4.

Lewis, G. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal* **10**: 30–9.

London, J. 2004. *Hearing in Time*. Oxford: Oxford University Press.

McAdams, S. and Bigand, E. (eds.) 1993. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

Mion, P., Nattiez, J.-J. and Thomas, J.-C. 1982. *L'envers d'une oeuvre: 'De Natura Sonorum' de Bernard Parmegiani*. Paris: Buchet/Chastel.

Mital, P. K. and Grierson, M. 2013. Mining Unlabeled Electronic Music Databases through 3D Interactive Visualization of Latent Component Relationships. *NIME Proceedings*.

Nattiez, J.-J. 1990. *Music and Discourse: Towards a Semiology of Music*. Princeton, NJ: Princeton University Press.

Norman, K. 1996. Real-World Music as Composed Listening. *Contemporary Music Review* **15**(1–2): 1–27.

Park, T. H., Hyman, D., Leonard, P. and Wu, W. 2010. SQEMA: Systematic and Quantitative Electro-Acoustic Music Analysis. *Proceedings of the 2010 International Computer Music Conference (ICMC), Stony Brook*. San Francisco: ICMA.

Park, T. H., Hyman, D., Leonard, P. and Hermans, P. 2011. Towards a Comprehensive Framework for Electro-Acoustic Music Analysis. *Proceedings of the 2011 International Computer Music Conference (ICMC), Huddersfield*. San Francisco: ICMA.

Rowe, R. 1993. *Interactive Music Systems: Machine Listening and Composing*. Cambridge, MA: MIT Press.

Schaeffer, P. 1966. *Traité des objets musicaux*. Paris: Éditions du Seuil.

Smalley, D. 1996. The Listening Imagination: Listening in the Electroacoustic Era. *Contemporary Music Review* **13** (2): 77–107.

Smalley, D. 2007. Space-Form and the Acousmatic Image. *Organised Sound* **12**(1): 35–58.

Solis, J., Chida, K., Taniguchi, K., Hashimoto, S. M., Suefuji, K. and Takanishi, A. 2006. The Waseda Flutist Robot WF-4RII in Comparison with a Professional Flutist. *Computer Music Journal* **30**(4): 12–27.

Spevak, C. and Polfreman, R. 2001. Sound Spotting: A Frame-Based Approach. In *Proceedings of the Second International Symposium of Music Information Retrieval (Bloomington Indiana)*: 35–6. (www.ismir.net).

Truax, B. 1984. *Acoustic Communication*. Norwood, NJ: Ablex Publishing Corporation.

Wang, D. and Brown, G. J. 2006. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Piscataway, NJ: Wiley-IEEE Press.

Weinberg, G. and Driscoll, S. 2006. Toward Robotic Musicianship. *Computer Music Journal* **30**(4): 28–45.

Wiggins, G. A., Pearce, M. T. and Müllensiefen, D. 2009. Computational Modeling of Music Cognition and Musical Creativity. In R. T. Dean (ed.), *The Oxford Handbook of Computer Music*. Oxford: Oxford University Press.

Windsor, L. 2000. Through and Around the Acousmatic: The Interpretation of Electroacoustic Sounds. In S. Emmerson (ed.), *Music, Electronic Media and Culture*. Aldershot: Ashgate.

Winkler, T. 2001. *Composing Interactive Music: Techniques and Ideas Using Max*. Cambridge, MA: MIT Press.

Wishart, T. 2012. *Sound Composition*. York: Trevor Wishart.