Designing Interaction for Co-creation

Oded Ben-Tal,

Kingston University
o.ben-tal@kingston.ac.uk

Abstract. This paper describes several compositions of live, interactive electronics where mutual listening between the performer and the computer form the basis for the interaction. The electronics combines algorithmically defined musical-logic with inputs from the performer. These become a musical partner to the performer mixing system creativity with the composer's and the performer's. The paper places this approach within the context of computational creativity on the one hand and live electronics on the other.

Keywords: algorithmic composition, live-electronics, interaction, machine listening.

1 Introduction

This paper uses the trajectory of my own creative work in the area of live, interactive electronics as a basis for discussions about computers as creative partners in this domain. Both 'live' [1], [2] and 'interactive' [3] are somewhat contentious terms. Nevertheless, for the purpose of this paper I will use these to mean music where signal from a performer on an acoustic instrument is routed to a computer. This signal is used by the computer to contribute electronic sounds to the music. Central to my approach to this domain is mutual listening. I use techniques borrowed from machine listening to analyse the performer's input. This information is used to shape the computer's responses. The score for the piece, together with the programmed electronics, setup a scenario whereby the performer needs to listen to the electronic sound and respond to it as well.

Working experimentally with live electronics creates specific technical and practical challenges. The composer (as well as the performer in many cases) need to navigate between the desire to create inventive and musically interesting electronics and the need for stability and reliability in the same. While some sound processing can be applied very reliably and in real time, a more creative approach to music making often ventures into less well-trodden territories. The result is that technical challenges and aesthetic/compositional issues are intertwined in the development of a piece.

Another challenge, at least for my own approach to composition, is about the relationship between the material and the electronics. As a composer that does not start from a concept or a structure but from a concrete sonic idea (in either notated form or rendered as audio), composing live electronics becomes a chicken and egg problem. Until I know what the musician will be performing it is difficult for me to discover what transformation or reaction from the electronics would work. By that I mean both from the technical point of view: what I am able to program, and from the musical point of view: what will integrate and complement the performer in interesting ways. My approach to solving this problem has been an iterative design process. I often ask performers to record some preliminary sketches which I then use to start developing the electronics. I adjust and refine both of those in tandem, sometimes with the intended performer over several months. The results is a creative system made of human and artificial components working together in a co-creative situation. The creativity that is in focus here is what takes place in performance.

This paper is organised as follows: first I will discuss the design of the electronics, then explain about the scores and how they allow the performer scope to listen and respond to the electronics. The final part will discuss two pieces in some detail to illustrate how these ideas play out in implementation.

2 Machine Listening

Pitch [4] and onset [5] detection are the attributes I use most frequently. Both are fairly robust for monophonic signals though octave errors are relatively common in pitch detection. It is also notable that like similar music information retrieval (MIR) techniques these were developed with studio recorded popular music as a prototype. Applying them in live concerts of experimental music is not 'covered by the warranty'. I also use some timbral feature such as spectral centroid, spectral flatness, or the degree of pitchness (a byproduct of pitch detection). But while it is reasonable to expect the performers I work with to exert control over pitch-time aspects of music with high degree of precision and virtuosity. The same cannot be said about spectral features. Performers can play brighter and darker tones but translating extracted features into meaningful relationship between the performer and the system requires more control.

This is one example where technical and aesthetic considerations interact. The mutual listening between human and computer I am after is based on pitch-time because both the performer and the system can handle those effectively. This in turn shapes the scores I produce and the electronic process I design. To me this is analogous to the technical limitations imposed by instruments themselves. In that sense MIR techniques are just another piece of music technology like the flute. And like the flute it is changing over time from the ancient bone flute to the modern orchestral one. And like many other artists over this long history I am interested in exploring new technology as well as pushing it further.

2.1 Signal to Data

The relationship between the acoustic performer and the electronics is a central question in live electronic music. Croft [1] speaks for many practitioners when he argues for the instrumental paradigm where the electronics become an extensions of the performer just as a musical instrument is. His focus is on the immediate and embodied experience of listeners and performers and the goal seems to be similar to Csikszentmihalyi's [6] concept of flow. Implicit in this discussion is the importance of performance practice. Croft's instrumental paradigm envisions the performer achieving the same fluency with the electronics as that obtained through many years of learning an instrument. This fluency is defined by Croft at the micro-level of sound production and control. In some ways my approach is also concerned with the performer's hard earned musical training. But the focus of the interaction is at a slightly higher level, at the level of notes and above: phrases and sections. Shaping the music at this meso-level is also part of musicians' training and their performance practice. It also engages (and challenges) their musical creativity in different ways.

Shifting the focus of the interaction from the event level of pitch (and onset) detection to larger context adds latency to the system. In that sense my approach runs contrary to one of the primary concerns of practitioners in a field where the question 'how low is the latency on your system?' is near the top of many conversations. But since I am interested in a musical (not just sonic) dialog between the performer and the computer this is a price worth paying.

The first way in which this interaction at a higher level is achieved is by including more then just the current sound when determining the system behaviour. This can take different shapes - from estimating average inter-onset intervals over the last n-attacks, to generating material taken from the last x-pitches detected, to collecting melodic contour and resynthessing them with different sounds. As these examples illustrate the problem of segmentation complicates matters considerably. I have been using either arbitrary decisions (keep last 4 notes) or simplistic assumptions (long enough silence, large pitch gap) to overcome this problem. Neither are very musically informed mechanisms but the problem of segmentation is complex even in off-line applications.

2.2 Decision Making

The data extracted from the performer is used by the system to shape the electronic sounds produced. The system includes some analysis of the data coupled with decision making. These decisions could be as simple as binary choices e.g. apply different processing to high/low notes. Some of the parameters in the electronics are also chosen randomly from an experimentally determined list or range. The electronic sounds include both processing of the microphone input but also synthesised sounds. For performers who already worked with electronic processing before, the first encounter with my systems is a bit confusing. Unlike fixed effects were players quickly learn to anticipate the sound they will hear, the system is too dynamic. However, during the rehearsal process and with the help of explanations and guidance from me about what the electronics are doing, they learn to listen in the right way. The system is consistent in a statistical way – anticipating specific events is rarely possible but the overall behaviour is predictable.

The end result of this learning process is a creative system, encompassing both human and computer components together, with a shared creative responsibility [7]. The computer component of that system conditionally responds to outputs from the human player based on analysis and evaluation. This is a simple form of self evaluation which many (including Colton and Wiggins [7]) consider as critical component of a creative system. Needless to say, the human player also engages in evaluating the computer generated outputs and responding to them. Obviously, the two components in this system – the human and the machine – are not equal. The computer is operating at a much more basic level with a more limited range of responses. The human player definitely bears much more responsibility to the overall aesthetics of the performance. Nevertheless, as the performers themselves attest, the computer exhibits enough independent behaviour to be considered a partner.

At the end of the composition/development process I arrive at a set of independent electronic processes. Some of these are variations on ones I have used in an earlier piece and some are new. Since I tailor the electronics with recordings from the player, even when starting from existing instances I adapt and change aspects to work better with the different sound qualities, articulations, range, and the musical material I compose. Since most of the pieces discussed here are for a solo performer, giving her the ability to control the electronics maintains the nature of the piece as a *solo*.

3 Scores

When the electronics are conceived as an extension of instrumental techniques, performance nuances (such as articulation, vibrato, etc.) allow the player to shape her sound in line with the electronics. For the relationship between the performer and the electronics to be heard as a musical dialogue, the human performer needs more freedom to react to material generated by the computer.

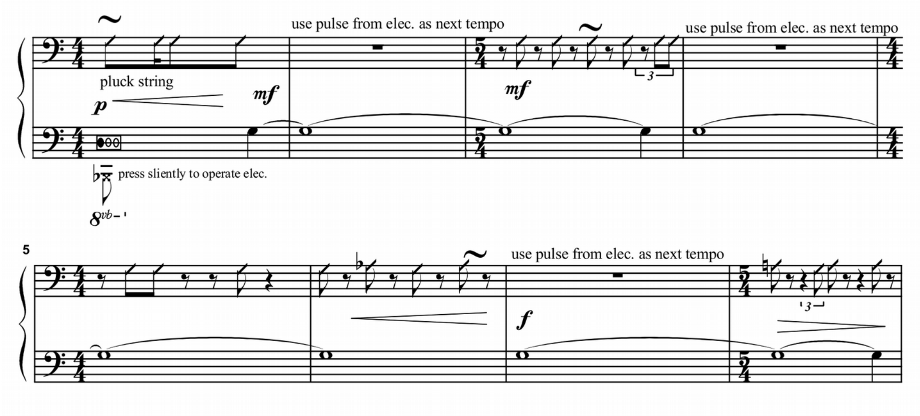

Fig. 1. Opening bars of Non Sequitur for Piano and PnoScan. Pianist getting a performance cue from the electronics.

In the opening section of *Non Sequitur* notes from the pianist (converted to MIDI messages using the PnoScan device[1]) trigger regular pulsed notes from the

1 http://www.qrsmusic.com/PNOScan.asp

electronics. The tempo of these pulses is mapped from the velocity of the note (louder sounds result in a slower pulse). The score (Fig. 1) instructs the pianist to listen to the generated pulse and use it as the tempo of the next phrase. This is an example of setting up an explicit mechanism requiring the performer to listen and respond to the electronics.

The first movement of *Anemoi* for flute and live-electronics, asks the player to shape the movement through listening and controlling the electronics. The electronic processing uses a network of delay-feedback loops. When the flautist plays high notes the amount of feedback is increased while low notes decreases it. The player is asked to locate the delicate balance point in the feedback resulting in a slow crescendo – before the feedback gets out of control. The score leaves moments (see beginning of 2nd stave in Fig. 2) for the player to continue playing in a manner similar to the preceding music but leaves the pitch height free to allow the player to increase/decrease the feedback using high/low notes.

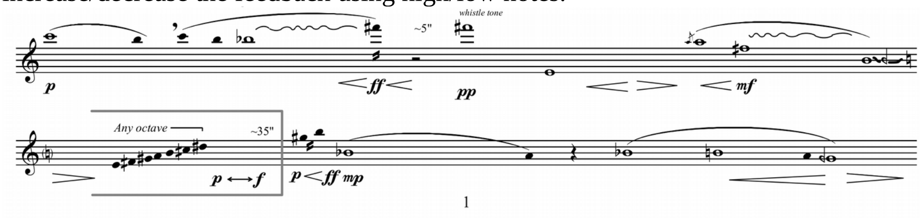

Fig. 2. extract from *Anemoi* for flute and live electronics. Proportional notation and constrained improvisation.

In practice we found that visual monitoring (through an on-stage screen) is needed to perform this task.
The score for *Metaphors of Space and of Time* (see 4.2) leaves much more room for improvisation as it was composed for a keen improviser. The amount of information the score provides varies between movements (e.g. Fig. 4 & Fig. 5) but always leaves a wide scope for the performer to shape the music.

4 Case Studies

Two examples will serve as illustrations for the issues discussed above: Use of machine listening to extract information from the performer; Parsing that data to build selective response mechanisms; generating electronic responses from a combination of internal processes and performer's input; Designing compositional strategies for mutual listening between computer and performer.

4.1 Zuam: Beyond Mind

Zaum: Beyond Mind is a sound theatre piece developed over a period of 3 years with fellow composer/performer Caroline Wilkins [8]. Caroline is performing vocally and on bandoneon both of them routed into my laptop. In the performance I am also on stage and we position the computer as another musical instrument. This is achieved both through the design of the interaction with the acoustic sources but also through staging. Movement, lighting, and video projection are an integral aspect of the piece.

The design of the piece is modular in two respects. First, each unique performance of this piece is created through structuring of loosely connected scenes[4]. The duration of the piece can range from 15 to 50 minutes. The choice of modules to perform and the order takes into account the context including possible other performances in the concert, the space, and the occasion.

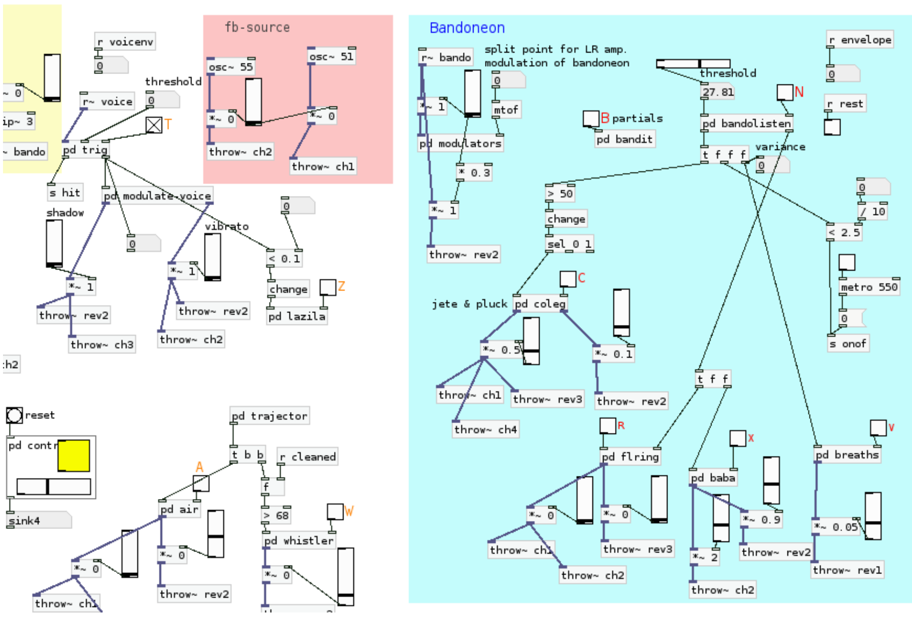

Fig. 3. top level pure-data patch of *Zaum: Beyond Mind*.

The design of the electronics is also modular – independent processes which I mix live in performance. There are separate sets of processes for the voice and for the bandoneon all implemented in pure data. The long development process of this piece allowed us to adapt our performances to each other's musical ideas. We met every month or so for an extended rehearsal/development session. In between those meeting we would each develop our own side of the dialogue – Caroline on her vocal and instrumental material me on electronic material (which includes fixed soundfiles generated from recordings of Caroline's performances as well as live, interactive processing of her sounds).

The first stage in the patch is pitch tracking and some 'cleaning' of the data. Gaps in time and frequency are used as a very crude approximations for weeding out misfiring of pitch detection and for segmentation. Figure 3 shows the top level pd patch of the piece. The letters (A, B, C, R, W, etc.) are keyboard shortcuts that turn processing modules on and off. Both voice and bandoneon have some modulation applied to them ([pd modulators] and [pd modulate-voice] objects in fig. 3). In both cases the detected pitch determines the exact nature of the modulation. The main purpose of this effect is to enhance the integration of the acoustic and electronic components, bridging the perceptual and spatial gaps between these sources. The parameters of the modulation in relation to incoming pitch were determined experimentally. The effect is often subtle in relation to the rest of the electronic sounds present but is nevertheless important.

One of the scenes has Caroline performing the bandoneon on the floor.[2] The modulation of the bandoneon sound is audible as a sharpening of the timbre on some sustained notes (~0:07 for example). The other electronic sounds are synthesised reactions. The noise bursts are filtered white noise. The set of filters is mapped from spectral peaks of the bandoneon. The trigger of the noises and the individual durations are generated from the variance of the detected pitch. I used a semi-random selection from this data stream to thin-out this layer which was too dominant. The percussive sounds, most audible starting 0:28, echo the key clicks and tapping of the body Caroline is using in this section. The sounds are generated by two mechanisms. The first is a Karplus-Strong plucked-string model with pitch and velocity mapped from the bandoneon signal. The other sound originated from a recording of a cello bouncing the wood of the bow on the string (col legno battuto). This sound is then ring modulated based on the pitch detected from the bandoneon. While each of these individual components are fairly simple the combined result is a rather complex relationship between the performer and the electronics.

A recording from our performance at the Sonorities festival[3] illustrates some of the vocal moments of the piece. Starting from 5:20 in the video Caroline is performing English translations of Zaum poems (originally Russian). There is a background layer of a fixed soundfile - fairly high wobbly sounds. Bursts of bandoneon clusters are triggered primarily by sharp changes in amplitude including both sharp attacks and abrupt endings. At the same time Pd is attempting to trace melodic contours from the voice ([pd trajector] object lower left side of Fig. 3). When one is captured – above a threshold of number of detected pitches in fairly close proximity – this is mapped to the frequency of a sine-tone. A feedback delay, with the velocity of the voice controlling the amount of feedback colours the resulting sonic gestures ([pd whistler] in Fig 3).

The last example[4] illustrates a moment in the performance that came out of an accidental discovery. The electronics used in this section were originally intended for the bandoneon. I was struck by the rich spectrum of the instrument and built a process that enriches that spectrum through the gradual addition of partials over a detected fundamental. While testing aspects of this I was using a microphone which picked up the generated overtones. Since the microphone was still running into the pitch detection mechanism the system began to feedback on itself. But the signal was not routed directly in-out but through and analysis-to-synthesis process. The results was a rich, evolving sound as partials became fundamentals for new partials. I added mapping from the amplitude of the signal to a vibrato of the partials. Moving the microphone in front of the speaker changed the vibrato but also resulted in different pitch being identified as the most likely fundamental. The result (most visible starting around 1:40 in the video) is to turn the system – microphone, speaker and computer - into a kind of musical instrument.

2 This extract is available at https://www.youtube.com/watch?v=PxgeuCRkiAs .
3 https://www.youtube.com/watch?v=_rqr58OP0jc
4 https://www.youtube.com/watch?v=1xry7us_JvM

4.2 Metaphors of Space and of Time

The composition process here was also collaborative though not as lengthy and involved as in *Zaum*. A set of pedals arrayed on the floor in front of the player allow him to trigger and control levels of individual elements in the electronics. These, programmed in Supercollider, are in the form of 6 independent elements which include both processing of trombone signal and sounds synthesised in response to analysis of that signal. Each of the four movements of the piece – points, surface, lines, volume – has a main process matched to it, but the performer is free to add others as well. Over several performances he developed ideas about shaping each movement and the piece as a whole between the considerable freedom offered by the score and his ability to anticipate the sounds generated by the electronics and use them effectively.

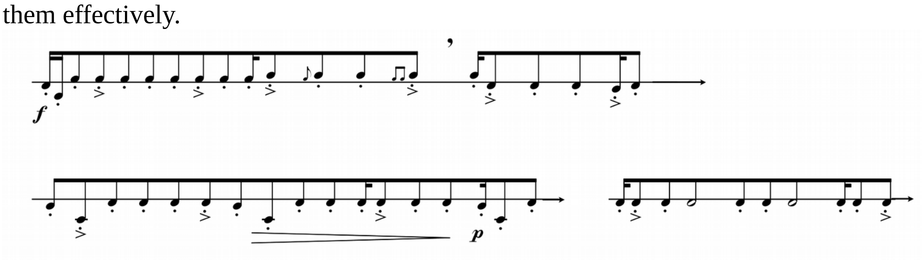

Fig. 4. extract from the first movement of *Metaphors of Space and of Time*. Guided improvisation.

Figure 4 shows part of the score for the first movement.[5] Both rhythm and pitch are suggested rather then specified and his interpretation is fairly free. The electronics consists of two simple delays (without feedback) with the delay time derived from the average inter-onset interval (IOI) of the trombone playing. One delay is set to the average IOI (over the last 5 attacks) while the other is set to the reciprocal of that value. That is, one delay follows the player the other delay counters him. At 60 bpm all should be synchronised – a tempo Torbjorn is able to find towards the end of the movement (from ~1:12 in the video). There is also some pitch shifting in the electronics first from the dynamic changes in the delay time but there is also a pitch-shifter in the process. Supercollider counts the number of attacks and gradually drifts the pitch away from the original either up or down.

For the third movement (starts at 5:00) we decided to use a practice mute which he fitted with a microphone. The result is that very little acoustic sound is present from the instrument instead we hear the amplified trombone blended with the electronics. The main effect is ring-modulation with the modulation frequency taken from the pitch detection but through a slow ramp. Since the trombone material is also primarily glissandi the result are constantly shifting tones. At 5:30 Torbjorn is turning on a second effect which is based on fm-synthesis (with configuration derived from emulation of brass instruments). These are also mostly glissandi tones which are chosen semi-randomly based on the pitch detection (melodic intervals from the trombone note).

5 The following discussion refers to time point in this video: https://www.youtube.com/watch?v=RXG__euYtcY&t=87s .

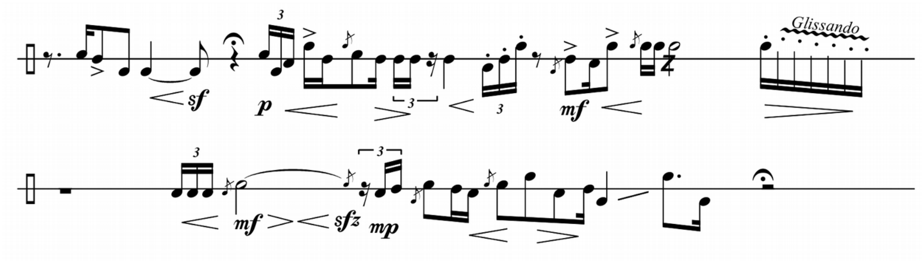

Fig. 5. extract from the fourth movement of *Metaphors of Space and of Time*. Player needs to find several percussive sounds with the instrument.

The final movement (starting 6:25) asks the player to produce noisy almost pitch-less sounds (see Fig 5). The analysis is based on onset detection and some timbral parameters such as spectral centroid, spectral flatness, and spectral entropy. These are mapped onto bandpass filters with centroid controlling the centre frequency, flatness and entropy effecting the bandwidth. The sound source is noise. Controlling the electronics here proved the most difficult. Triggering the onset detection required very heavy emphasis on lip articulation putting great demands on the player.

5 Discussion

In this paper I argued that when interactive electronics exhibit enough independent behaviour, we can consider those as a creative system. As Pearce, Meredith, and Wiggins [9] observe, the evaluation of systems that produce musical outputs needs to reflect the purpose for which it was designed. Since central to my approach is the musical relationship that performers can form with the electronics, observations from the performers themselves are a pertinent form of evaluating the success of this approach.

Torbjorn considers *Metaphors of Space and of Time* "one of the most unique and generally useful pieces of music&technology ever written for me. […] The work combines a great degree of free interpretation within a very strict artistic framework, something that has made it possible to successfully perform what might in some circles appear as 'difficult' music to audiences of all backgrounds." He specifically commented about the unique nature of the electronics: "These are very effective and useful patches - useful well beyond my own use and of *Metaphors of Space and of Time."* He now regularly incorporates these electronics into his performances improvising with them or using them when performing other pieces. In a sense, these became his accompanist which he described as "a somewhat wilful partner". Perhaps not every performer enjoys working with a wilful partner but Torbjorn does and even considers these as "one of the most liberating artistic devices in my already very extensive technology-based equipment."

After improvising with my electronics in a concert, Cellist Matthew Barley remarked: "The system enabled me to improvise with the computer which gave the impression of being a 'live' **improvising partner** which is something I have been wanting to achieve for over 10 years" (emphasis added). After another concert he added: "I did an improvisation of around 20' and it felt great - very organic, and it was possible to **'compose' on the spot** without the technology getting in the way, but feeling it as an extension of the instrument **and my own imagination**" (emphasis added). In some ways, an expert musicians (such as Matthew) describing a computer as an improvising partner, is closer to a musical Turing test then the discrimination tasks that are often labelled as such [10].

The main attraction for these musicians is the balance of novelty and predictability that comes about from the combination of algorithmically defined musical processes with machine listening input. As both of them (and others) observed – performing into a fixed 'effects box' even if the effect creates rich and beautifully crafted sounds, easily becomes too predictable. The right amount of randomness, some of it programmed some of it resulting from the changing nature of the input, coupled with enough predictability allows the musicians to develop a creative dialogue with the system. The predictability is not at the lowest level – specific notes or sounds are rarely anticipated. But the system does exhibit enough statistical predictability that the performers can learn. The result, in the phrasing of Agres, Forth and Wiggins [11], is that the performer *advises* the system rather then controls it. It can be seen as a matter of degree but such shared creative responsibility means that the system is more then simple reactive electronics or a purely generative system.

The other major factor in making the pieces work is the ability of performers to respond to what they hear. This relies on the performers themselves as well as on designing the score and electronics in a way that accommodates bi-directional adaptation. These works require involved learning process from the performer -they are not pieces that can be successfully performed after two rehearsals[6]. At the same time many of the skills required to navigate these pieces relate directly to the core musicianship skills of performers: nuanced and sophisticated control of an instrument; ability to listen and adjust their playing in response to immediate musical circumstances; Bring their interpretation to a score according to their musical judgement. The result is what Lubert [12] calls computer as a colleague where "technology is incorporated in the creative act as a support for the creator's musical expression."

It also echoes the enactive model for co-creation proposed by Nicholas et al[13]. The importance of listening between the human and machine (which together make up the creative system) matches the action-perception link at the heart of the enactive approach. Both sides change their output (action) in response to an evaluation of the input (perception). The computer agent is much less developed and sophisticated in that sense and the responses are not based on an explicitly aesthetic evaluation. The human player is the one who listens and controls the combined mix and therefore has the final creative control. Therefore, there is considerable scope for research on improving those aspects. The freedom written into the scores seen above matches the enactive model's focus on directives rather then goals: "A directive constrains and suggests potential actions that could yield productive changes in an emergent process of sensemaking"[13]. The score for *Metaphors of Space and of Time* (Fig. 4 & 5), for example, provided Torbjorn with constraints and suggested potential actions which he

6 Circumstances which are, unfortunately, all too common in concerts nowadays.

explored over a period of discovering 'the piece' through rehearsals and performances.

Finally, there is a degree of tension between developing live electronics as a musical instrument or as a creative system. We want musical instruments to be robustly predictable – sound reliably follow actions. In contrast, we expect to be surprised by creativity, at least occasionally. In other words, for a live electronics setup to be perceived as a creative system, we have to think beyond extending instrumental capabilities.

Acknowledgements

I would like to thank the reviewers for their useful suggestions

Reference

1. Croft, J. Theses on liveness. Organised Sound, 12(01), 59-66 (2007).
2. Emmerson, S. Living electronic music. Ashgate Publishing, Ltd. (2013).
3. Stroppa, M. Live electronics or… live music? Towards a critique of interaction. Contemporary Music Review, 18(3), pp. 41-77 (1999).
4. De La Cuadra, P., Master, A., & Sapp, C. Efficient pitch detection techniques for interactive music. In Proceedings of the 2001 international computer music conference pp. 403-406 (2001).
5. Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. A tutorial on onset detection in music signals. IEEE Transactions on speech and audio processing, 13(5), pp. 1035-1047 (2005).
6. Csikszentmihalyi, M. Flow and the psychology of discovery and invention. New York: Harper Collins. (1996).
7. Colton S, Wiggins GA. Computational creativity: The final frontier?. In Proceedings of the 20th European conference on artificial intelligence pp. 21-26. IOS Press. (2012).
8. Ben-Tal, O., & Wilkins, C. Improvisation as a creative dialogue. Perspectives of New Music, 51(1), pp. 21-39 (2013).
9. Pearce, M., Meredith, D., & Wiggins, G. Motivations and methodologies for automation of the compositional process. Musicae Scientiae, 6(2), pp. 119-147. (2002).
10. Ariza, C. The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems, Computer Music Journal 33(2), pp. 48–70, (2009).
11. Agres K, Forth J, Wiggins GA. Evaluation of musical creativity and musical metacreation systems. Computers in Entertainment (CIE) 14(3):3 (2016).
12. Lubart, T. How can computers be partners in the creative process: classification and commentary on the special issue. International Journal of Human-Computer Studies, 63(4), pp. 365-369. (2005).
13. Nicholas D, Hsiao C, Popova Y, and Magerko B. An enactive model of creativity for computational collaboration and co-creation. In Creativity in the Digital Age, pp. 109-133. Springer London, (2015).