VISUALISING CO-AGENCY AND AIDING LEGIBILITY OF SONIC GESTURES FOR PERFORMERS AND AUDIENCES OF MULTIPLAYER DIGITAL IMPROVISATION WITH UNISSON (UNITY SUPERCOLLIDER SOUND OBJECT NOTATION)

Jules Rawlinson
University of Edinburgh
Edinburgh, UK
jules.rawlinson@ed.ac.uk

Marcin Pietruszewski
University of Edinburgh
Edinburgh, UK
mpietrus@exseed.ed.ac.uk

ABSTRACT

This paper addresses interdisciplinary research into action, agency, aura and spectacle in electronic music performance, with a specific focus on small improvising ensembles. The aims of this research were to develop a framework for graphical representations of multiplayer sonic/musical interactions in performance by exploring data visualization and machine listening of shared musical gestures in performance. The main output from this research is *UniSSON*, a suite of software tools that presents a real-time multi-temporal and multi-resolution view of sonic data across a number of sound-based parameters in accessible ways, which contributes to audience engagement and collaborative performance. The research builds on existing strands of work in creative computing, computer music and musicology but seeks to make newly playful use of these techniques whilst also addressing accessibility issues by working with both widely adopted and open source software.

1. INTRODUCTION

In electronic music, especially by groups, it can be hard for audiences and performers to gauge who is doing what because physical movement and action is so decoupled from sonic results. If the audience and performers are not able to audibly or visibly (at a gestural level) perceive contribution, how might it otherwise be represented? On the other hand, mutual connectivity offers radical possibilities for making sonic outcomes that are fluid, shared and responsive through active layers of software with a capacity to act as a conduit and focus of interaction and exchange, what Bown, Eldridge and McCormack identify as "behavioural objects" [1].

Beyond issues of representation, there were other concerns regarding creative and cognitive extension/offloading raised by the symbolic nature of digital musical instruments, as proposed by Magnusson [2], that might fruitfully be examined. Could/would a visual representation of past/present activity exert its own agency and suggest future directions and possibilities in improvisation? In response to Magnusson, Green [3] describes his experience of "overspill" while coding, with detours, dislocations, deviations and disruptions creating a space or distance for reflection on structure and execution that might also be an ambition for practice-led enterprise.

This research aligns itself with Shafer's views of permutative interactions that arise from real-time (an)notation in research of formal, temporal interaction and morphological interaction [4]. In summary, the research began with a general set of questions at the intersection of the technical and the aesthetic for exploring real-time graphical display of multiplayer activity that would be both practice-based and practice-led. How to visualize the relationship between a digital musician's physical gestures and sonic gestures, given that these can be radically decoupled? How can relationships between a group of musicians' gestures be visualized in a way that is musically suggestive and provocative, not just descriptive, and serve as a foundation for an unfolding collaborative performance? In addition, can these visualizations contribute to audience engagement without merely signposting?

2. BACKGROUND

The specific starting point for this research grew out of collective experience of performing with an established laptop trio who blend styles found in dub, hip-hop, electronica, noise and beyond. A founding principle of the trio when it formed in 2008 was to try to make a virtue of the confusion of agency that collective laptop music can exhibit. To this end, an important aspect of our improvising approach is to be constantly sampling and transforming each other, in pursuit of an organic, shifting sound mass. We have regularly embellished what Waters might call an "ecosystemic" [5] approach with various technical mediations that, for instance, use machine listening technologies to aid and impede the consonance and counterpoint of the three humans on stage.

As a result of this continual co-tuning, there's opportunities for confusion to emerge about who exactly is contributing what, not least on the part of the audience, but also from the perspective of performers. While the trio uses aspects of embodied and gestural practice, each member uses roughly similar tools and interfaces, and there might not be a direct correlation between sounding action and sounding outcome (i.e. midi 'notes' produced as a result of interaction with controllers, will not necessarily result in pitched output, a single short action might result in a long stream of sonic material, one player might be processing another) impacting on awareness of visible rather than audible contribution.

The foundations of the research also drew on the experience of performing as part of an established improvising laptop and modular synthesis duo, where, again, sonic outcome is broadly similar and located in polyphonies of monophonic generative patterns, but in this case the differences between performer interfaces are markedly so. The laptop performer is exactly that, making use of text-based tools and a minimal graphical user interface with no use of controllers, and the modular synthesis performer having many visible controls. As with the laptop trio, the audience and laptop performer often struggled to differentiate material and contribution, especially where more stream based and textural sonic content was concerned.

3. RELATED WORK

There's a rich seam of work relating to visualization of sonic and/or musical data that has informed the development of the tools, either through adoption/adaptation of method or by noting where the approach is not particularly suited to a generalist implementation.

There have been different explorations of using visualizations prescriptively to structure improvisation, such as Justin Yang's *Webwork* (2010)[1] or Rodrigo Constanzo's *Dfscore* (2014)[2] as well as the author's own work on multiplayer notation for finely threaded interactions, which makes use of a modification of the approach outlined by Thoreson [6]. Within the field of live coding for music, Thor Magnusson's *Threnoscope* (2013)[3] is a specialized descriptive visualization designed to give the audience an insight into unfolding musical structure [7].

Couprie [8] has outlined a comprehensive method for transcription and representation that includes stages of causal analysis, morphological analysis, functional analysis and formal analysis, and notes the challenges of symbolic and iconic representations before discussing experiments with visual correspondence and providing an example of *EAnalysis* software for analysis, representation and transcription of nonwritten music.

Temporality and on-screen motion feature in Vickery's research [9,10], which highlights the necessity of sequential presentation of notation in time, observing that in many scrolling scores that motion implies a specific visual focus or notational 'now' [9]. Vickery goes on to explore the opportunities afforded by colour and shape to accommodate representation of sonic parameter through cross-modal similarity and equivalence [10].

Audience experience of electronic audio-visual performance is the focus of Olowe, Moro and Barthet's *resid-UUm* tool [11] which proposed a shape-based particle system orchestrated through sonification as an attempt to implement an interface in which all actions by a laptop performer are able to be explicitly interpreted by the audience. Their research notes challenges in regard of clarity in multilayered sound-image relationships. Graphic attributes such as shape, colour, fill, size and location define the sonic characteristics of events and material but viewers were not always able to detect or determine causality of outcome beyond the primary mechanic of shape creation.

More recently, Tadej Droljc's *Capillaries Capillaries* (2017)[4] makes use of a parameter led approach where control data for sonic events shapes visual material, and inversely, control data for visual material shapes sonic events. The final product is visually and sonically rich and engaging, but abstract and instantaneous. William Field's A/V improvisations are tightly correlated and integrated, and it's easy to perceive component parts (different sounds are represented by different visual elements)[5], and he has also shown work-in-progress on automatic transcription of melodic and percussive material which is result of both midi data and audio/envelope following[6]. Jason Levine's research explores form from sound using real-time 3D representations, including some analysis/representation of spectral brightness/noisiness[7], and Johannes Lampert's Anatomy of a Track non-real-time animation series uses shape and pattern to indicate developing musical structure[8].

4. METHODS

The main focus of this research is located in the visual representation of sonic material and form but before this process can begin a range of sonic data must be acquired. Analysis and evaluation rely on real-time audio feature extraction rather than other forms of controller data (e.g. MIDI), allowing for wide, flexible use. In addition to analysis and visualization tools, new techniques and tools for pulsar synthesis were explored and developed as a way to test the effectiveness of, and opportunities for, analysis and visualization.

A set of functional criteria for the software were set out before any practical work began, and included:

- Does the software make it easy to try different algorithms/approaches to machine listening and visualization to aid experimentation?
- Can the software be integrated into a standard computer music workflow with minimal configuration?
- Does the software run with minimal impact on computational resources?
- Does the software allow players to continue to use the base environment in their normal way, that is, does it respect their craft?
- Can the software be shared with third-parties reliably?

[1] See https://www.youtube.com/watch?v=O2F7M1Wh8n4
[2] See http://www.rodrigoconstanzo.com/dfscore/
[3] See https://vimeo.com/100148087
[4] See http://www.tadej-droljc.org/portfolio/capillaries-capillaries/
[5] See https://vimeo.com/268453148
[6] See https://vimeo.com/278033475
[7] See https://www.instagram.com/p/BlOYyKmHyXa/
[8] See https://www.youtube.com/watch?v=BKqTudojfhw

Alongside the functional criteria, a set of musical metrics for assessing the success of the research were also proposed before any practical work took place. These metrics included:

- Do players feel as if they have an improved sense of who is doing what in comparison to practice with an orthodox interface?
- Do players feel as if the visualization helps prompt decisions about what to do next?
- Do players feel that the interface both reflects their contribution to the interactive web and helps structure musical flow?
- Do audience members find the visualization aids their attention to the sound?
- Do audience members get a sense of who is contributing particular types of sound to the overall flow?
- Do audience members find that there is still room for surprise, even delight, with the visualizations?

The research developed through an iterative process of rapid prototyping and revision following critical practical sessions with both the laptop trio and laptop/modular duo. There were formal and informal presentations of work in progress to colleagues and peers at a range of institutions. The research culminated in a small number of public presentations and performances where audience feedback was canvassed following the questions posed above.

5. TECHNICAL OVERVIEW

The main focus of this research is located in the visual representation of sonic material and form and its impact on perception and practice, but before this process can begin a range of sonic data must be acquired. In this project analysis and evaluation rely on audio feature extraction rather than other forms of controller data (e.g. MIDI), allowing for wide, flexible use. In addition to the analysis and visualization tools, new techniques and tools for pulsar synthesis were explored and developed as a way to test the effectiveness of the analysis and visualization.

Extraction of audio features in the standalone tool is focused chiefly on low-level spectral, parametric, and temporal data. Preference was given towards computationally cheap processes using a range of SuperCollider analysis unit generators[9] that would be able to run on multichannel input in real-time on a single machine, in order that the tools could be easily shared at some later point.

Other options were evaluated, such as the use of dedicated networked low-latency Bela[10] devices for analysis. It's acknowledged that as the processor performance increases, more sophisticated solutions to feature extraction such as those reviewed by Ghalehjegh [12], including correlation, source separation and similarity might become more feasible for real-time multichannel use.

The analysis data is encoded to an OSC stream to be passed/parsed for visualization in a compiled Unity application. Early visual prototypes were developed in Processing using two-dimensional representations before Unity was settled on for its ability to easily accommodate a wider range of visually engaging and efficient three-dimensional renders and allowed for straightforward experimentation with physics (i.e. collisions), particle effects and game-like mechanics.

6. AUDIO ANALYSIS

The standalone analysis app (Figure 1) is currently fixed at six audio channels, with intended use as three stereo pairs and input source chosen through a drop-down selector. Pairs of input channels are summed to mono for analysis since there was no perceived benefit to analysis and subsequent representation of stereo data. While the standalone app is fixed in respect of channels, the underlying SuperCollider implementation allows for rapid array-based expansion should more channels be required (and the computer is capable of coping with the analysis).

Figure 1. UniSSON SuperCollider Analysis UI

Extracted features include an FFT for spectrogram (which is also the basis for other spectral measures), pitch, loudness, onsets, spectral flatness (essentially a measure from pure tone to noise), spectral centroid (the centre of mass of a sound and linked with perceived brightness), sensory dissonance (roughly a relationship between timbre and tuning), chromagram (an octave chroma band based measure of energy in a signal that attempts to correspond to traditional western musical notes), and lastly Mel frequency cepstral coefficients and Bark scale analysis for more reduced representations of energy in critical frequency bands. At the time of writing, analysis is not normalised.

Analysis data is formatted in the standalone extraction tool as OSC messages with a stereo pair prefix, data type prefix and either a single value or array of values as appropriate (e.g. /stream_1/frequency 440). Frames of data are sent at 60fps, so control rate rather than audio rate. Frame

[9] See http://doc.sccode.org/Browse.html#UGens>Analysis

[10] See https://bela.io/

sending of extracted features are enabled/disabled via toggles per stereo pair as a way to filter out data on demand and thin out transmission of OSC packets. The OSC stream is received[11] by the compiled (or editable) Unity based visualization app and additional data processing and manipulation takes place there. The OSC data could of course be read by other software for a variety of purposes.

Inside of Unity, the incoming values are smoothed relative to prior input and stored in shift register arrays of variable length (the default array length is 64 values, but this is easily changed in the Unity editor) allowing for a (short-term) historical view of data (which is further impacted by visual effects and design). Additionally, data was clamped and scaled as appropriate for effective display, sorted using nearest neighbour algorithms to aid subsequent legibility in image (rather than fidelity), and smoothed current minimum, maximum and average values were also stored. Some experiment was made in sending these derived values back out of Unity as OSC frames as a feedback process.

7. VISUAL MAPPING

The use of Unity as a graphics engine opens up a range of opportunities for data visualization, not least of which is the ease in which a variety of perspective can be swiftly rendered and manipulated. It also offered fast, flexible and scalable approaches to developing an object-oriented workflow for visualizing the analysis data in engaging ways. This section of the paper will avoid overly technical discussion of the Unity development, and focus instead on the qualities of the representation of data.

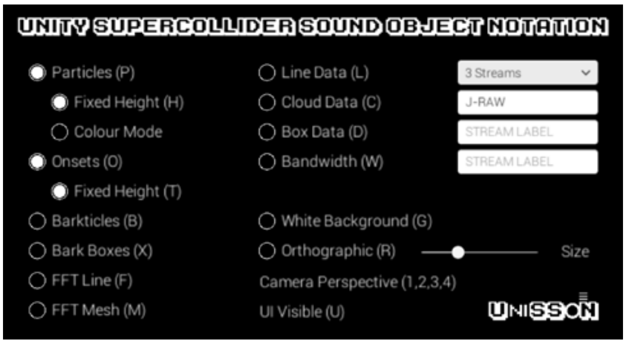

Figure 2. UniSSON Unity UI Settings Menu

The visualization follows some key principles outlined by Tufte [13], namely that the approach taken should encourage visual comparison, reveal data at multiple layers of resolution from coarse to fine detail, and be closely integrated with verbal descriptions of the features. The visualization also implements aspects of Shneiderman's [14] framework for the design of visualization systems by providing an overview of the data, allowing users to change focus, enabling detail on demand, presenting the ability to filter out data, revealing relationships between data, showing a history of activity and supporting extraction of collections. Different aspects/elements of the visualization can be controlled through keystrokes and user-interface (Figure 2), and include presentational preferences for background colour suitable for different settings and use cases (e.g. projection onto walls).

The overall appearance of the visualization application is aesthetically indebted to early video games, and makes use of simple geometric forms and fluorescent dayglo colours, but these colours (cyan, magenta, yellow) are easily distinguished from each other as part of comparison and visible against both black and white backgrounds. Due to the use of transparency and colour blending, these colours form additional mixed colours when sonic forms overlap, highlighting areas for consideration.

Single colours per stream were used as a way to track individual streams rather than the use of variable HSL across amplitude intensity or frequency as is commonly found, as in this case it would be very difficult to discern which visual feature belonged to which stream. In terms of domain specifics, conditions for acceptance of sound-image relationships include perceived temporal correlation of changes in sound and image [15] and structural linkage through mapped characteristics of appearance such as position, size, shape and opacity [16].

The main visual features are:

- A line-based display of the most recent incoming FFT values.
- A scrolling 3-dimensional waterfall spectrogram, with amplitude mapped to both height and opacity giving rise to a temporal sonic landscape (Figure 3).
- Scrolling particles showing a view of onset, monophonic pitch, amplitude and spectral flatness. Pitch is mapped on the x-axis, amplitude on the y-axis and scale, and flatness mapped to iconic representation (Figure 3). Where it's not possible to detect pitch, spectral centroid can also be used. These particles can be fixed in on the x- and y-axis to allow for different relationships and details to be extracted.
- Bar display of Mel or Bark based values of critical frequencies, providing a coarse overview of spectral energy.
- Scrolling particles (aka 'barkticles') showing Mel or Bark based critical frequencies, which give a useful representation of more percussive activity.
- A particle display of mid-term activity based on pitch/centroid, amplitude and spectral flatness. Pitch is mapped on the x-axis, amplitude on the y-axis, scale and opacity, and spectral flatness to the z-axis. This provides a useful indication of morphological gesture.
- A line display and/or particle cloud of longer-term activity based on pitch/centroid, amplitude and

[11] OSC handled by http://thomasfredericks.github.io/UnityOSC/

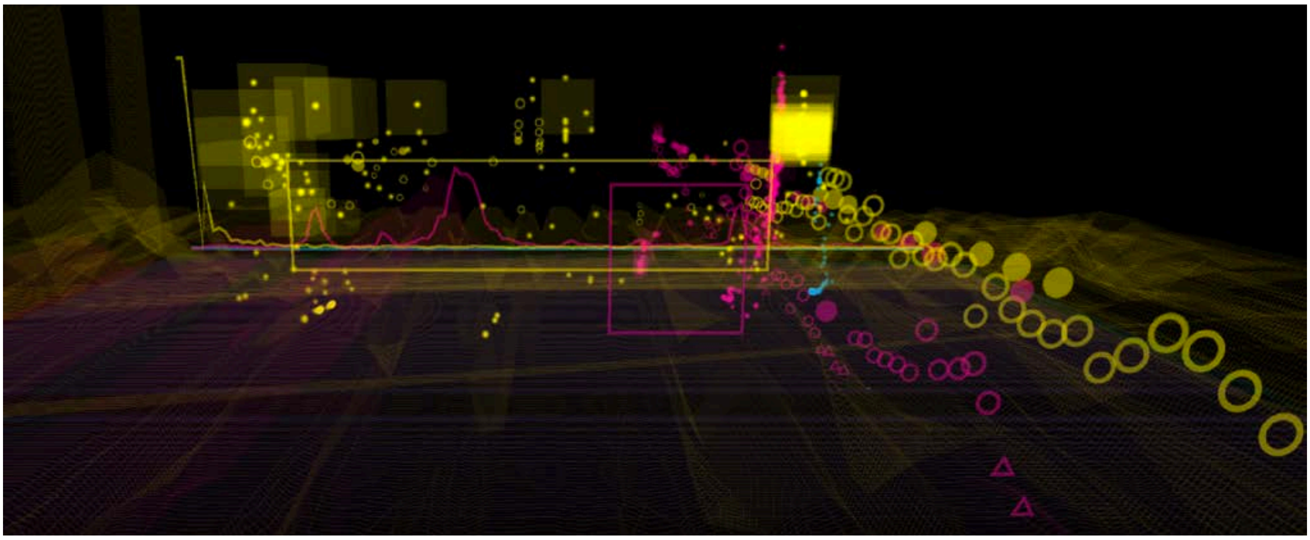

Figure 3. UniSSON Unity UI Waterfall display and particles showing pitch/centroid, amplitude and spectral flatness

spectral flatness. Pitch is mapped on the x-axis, amplitude on the y-axis, scale and opacity, and spectral flatness to the z-axis. This serves as a useful histogram of distribution of activity.

- An outline of smoothed average minimum and maximum frequency (pitch/centroid) and amplitude values. In part this offers a view of recent bandwidth, but also proved useful as aid to perception and clarity/definition when viewing data from an oblique angle as it acts as a bounding box for particles.
- Control of display of data from four different positions (front, back, oblique, above), giving a variety of views and insights. Control of background and projection method (e.g. perspective/orthographic).

Unlike the analysis tool, the compiled standalone visual implementation does not support per stream selection of features, although this is possible inside of the editable Unity project. Chromagram and sensory dissonance values are not currently represented visually.

In respect of scrolling data, some time was spent evaluating different scroll speeds, or more accurately interval between data points (how 'far' would each data point move per frame). A smaller number allowed for more temporal resolution, but the visualization did not seem to be perceptually coupled in terms of the audible experience of ephemeral and transient material. As this value is a variable, it can be updated as part of the user preference, although there's an additional computational impact on rendering speed given more objects. Studies were also made of different scroll speeds and directions for different simultaneously visible elements (e.g. spectral waterfall display vs. pitch/amplitude), to assess the challenges and opportunities, and consonance and dissonance of differences, which sometimes provided more clarity at the expense of correlation. Currently scroll speeds are fixed such that different elements (appear to) move at the same rates.

8. USAGE / EVALUATION

The research plan incorporated regular practical use sessions alongside code sprints as a way to trial the effectiveness of particular approaches, and some visualizations, such as a rotating envelope-following display, were discounted. In the case of the rotating envelope following display, it was a compelling method for showing synchrony and phase in repetitive (e.g. looping) structures, but its mandala-like qualities drew attention away too strongly from other features.

Outside of 'regular' use, a small number of investigations into game-like mechanisms occurred, by, for example, making use of analysis data to drive forms of games based on particle collisions, but this created additional visual material that did not derive directly from sound, and, while it was fun, took the practice in extra-musical directions that were not sustainably satisfying. Similarly, sonic data was used to build 'blockages' (restrictions) that fed back into a player's ability to make sound with particular properties (through gating), but specifics of setup were felt to be counter to wider use in line with the aims of the research.

While some of the approaches adopted are common (e.g. waterfall display), the strength of the visual tool lies in the combined and comparative approach in both time and spectral domains, in particular the way that various types of visualizations produce heat maps of activity. The visual analysis provided answers to the questions of 'is that me?' and 'who is that?' during sections of intense playing, and provided a stronger sense of shape and (over)contribution.

Importantly, rather than, or as well as, a head-down focus on a laptop screen or attention to physical gesture/action between playing, the head-up shift in focus to an external screen disrupted or diverted musical flow in what were judged to be useful ways. For example, sonic features were visible that might not have been heard, which re-focused listening as an active part of playing. Overall playing became more speculative as performers had an

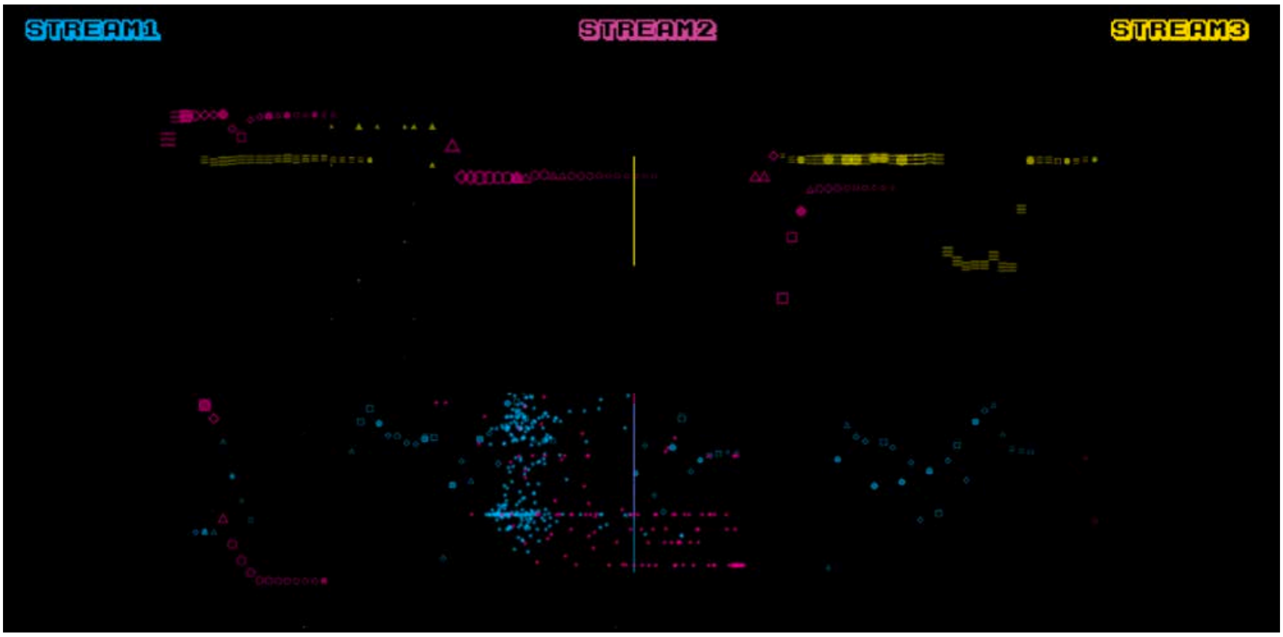

Figure 4. UniSSON Unity UI ‘Piano-roll’ view of particles showing pitch/centroid, amplitude and spectral flatness

additional set of information that caused them to deliberate over past, present and future form and content.

Audience members reported that they had found it interesting to try to connect the sounds to the visuals, but wanted more of a frame of reference in terms of mapping. Most found the scrolling displays viewed from above which result in something like a piano-roll (Figure 4) to be the most rewarding in linking sound and image and overall suggested that trying to tie sight and sound together made the experience more interactive and more engaging. In particular audience members enjoyed the visual patterns that resulted from generative processes employed by the laptop/modular duo.

While the tool is generally successful in meeting the criteria that had been set out at the start of the research, there are other questions that arose in and out of practice that are yet to be resolved. Most significant is the limitations of the analysis, as players might be contributing multiple components that are by necessity summed to a stereo output. In practice this muddies the audio analysis, and in particular effects the pitch/spectral centroid analysis.

When one performer is processing another player’s output without making additional contributions, similarities and divergencies can generally be observed, but if additional material is contributed, then this masks source similarity. A related issue was masking of onsets and terminations. Furthermore, some of the available analysis tools focuses on fairly ‘western’ notions of musicality (e.g. chromagram), where the laptop trio focus on more textural material. Side issues/aspirations that were not able to be addressed during the time available for the research included the urge to trace where a particular set of processes/transformations began, continued and ended (essentially ‘threading’), and to makes connections between sequences of gestures by performers.

9. CONCLUSIONS & FUTURE WORK

This work investigates ways of exploring productively the tension between legibility and co-agency in laptop and electronic performance. The tool stimulates what qualities of what Green [3] calls ‘agility and playfulness’ as a mediation between sonic fabric and performer and follows Bowers’ aesthetic of making a feature of accessible display of material, interactivity and collective activity [17].

The current state of the software gives a clear view of which audio belongs to which player, and indicates relationships between events/streams and gestures, but complexities arise in both analysis and display when a performer is both transforming another stream and adding new material. Again, here there are limitations of technology and visual perception which suggest humanised algorithmic listening.

Future developments could include corpus-based and machine-learning approaches that would extended the capacity and capability of the system in capturing and cataloguing sound character and formal structures. More ‘playful’ elements of the work (sonic gaming) remain in an exploratory and prototype state as in practice they seemed a distraction away from the more ‘informative’ use of the visualization tools.

The main successes of the analysis and visualization tools are in presenting a multitemporal and multiresolution view of sonic data allowing the simultaneous display of ‘instant’, ‘recent’ and ‘long-term’ data across a number of important sound-based parameters in accessible ways, which contributes to audience engagement and collaborative performance.

Acknowledgments

This research was made possible with funding from the University of Edinburgh's Challenge Investment Fund, and thanks Owen Green, Dave Murray-Rust, Lauren Sarah Hayes and Martin Parker for valuable advice and feedback.

10. REFERENCES

[1] O. Bown, A. Eldridge, and J. McCormack, "Understanding Interaction in Contemporary Digital Music: from instruments to behavioural objects," *Organised Sound*, vol. 14, no. 2, pp. 188–196, 2009.

[2] T. Magnusson, "Of Epistemic Tools: musical instruments as cognitive extensions," *Organised Sound*, vol. 14, no. 2, pp. 168–176, 2009.

[3] O. Green, "Agility and Playfulness: Technology and skill in the performance ecosystem," *Organised Sound*, vol. 16, no. 2, pp. 134–144, 2011.

[4] S. Schafer, "Performance Practice of Realtime Notation," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR 2016)*, Anglia Ruskin University, 2016, pp. 65-71.

[5] S. Waters, "Performance Ecosystems: Ecological Approaches to Musical Interaction," *Proceedings of the Electroacoustic Music Studies Network 2007.*

[6] L. Thoreson, "Spectromorphological Analysis of Sound Objects: An Adaptation of Pierre Schaeffer's Typomorphology," *Organised Sound*, vol.12, no.2, pp. 129–141, 2007.

[7] T. Magnusson, "Improvising with the Threnoscope: integrating code, hardware, GUI, network, and graphic scores," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, London, United Kingdom, 2014.

[8] P. Couprie, "Methods and Tools for Transcribing Electroacoustic Music", in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR 2018)*, Montréal, Université Concordia, 2018, pp. 7-16.

[9] L. Vickery, "The Limitations of Representing Sound and Notation on Screen," *Organised Sound*, vol. 19, no. 3, pp. 215–227, 2014.

[10] L.Vickery, "Some Approaches to Representing Sound with Colour and Shape," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR 2018)*, Montréal, Université Concordia, 2018, pp. 165-173.

[11] I. Olowe, G. Moro, and M. Barthet, "Residuum: user mapping and performance strategies for multilayered live audiovisual generation," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Brisbane, Australia, 2016, pp. 271–276.

[12] N. Hamidi Ghalehjegh, "Visualizing temporality in music: music perception – feature extraction," PhD thesis, University of Iowa, 2017.

[13] E. Tufte, "Visual Display of Quantitative Information" Graphics Press, 1983

[14] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualisations", in *Symposium on Visual Languages*, Boulder, Colorado, 1996, pp.336-343

[15] J. Schacher, "Live audiovisual performance as a cinematic practice," *The Cinematic Experience*, Sonic Acts XII, 2008.

[16] S. Callear. Audiovisual particles: parameter mapping as a framework for audiovisual composition," PhD thesis, Bath Spa University, 2012.

[17] J. Bowers, "Improvising Machines: Ethnographically Informed Design for Improvised Electro-acoustic Music," MA Thesis, University of East Anglia, 2002