Experience Driven Design of Creative Systems

Matthew Yee-King and Mark d'Inverno
Department of Computing, Goldsmiths, London
m.yee-king@gold.ac.uk,dinverno@gold.ac.uk

Abstract

The key contribution of this paper is to describe and demonstrate a novel application of grounded theory to the analysis of a human/machine music performance. Rather than attempting to measure the 'creativity' of our machine improviser, we instead proposed an investigation of the *experiences* of humans - in this case the designer, the performer and the listener. We report the design of an AI system chosen to perform in a specific *creative context* - a jazz-inflected musical performance in this case - and explore the specific experiences of these human actors through the performance itself. The performance is one which is a commonplace one where a single human musician interacts and performs with a single autonomous system. We describe this system which improvises by training pitch and event sequence models in real time from a live audio input and then uses a riffing behaviour to generate output in the form of note sequences with varying timbre. However, the main thrust of this paper is to propose a new methodology for understanding the role of the system through the interplay of experiences of audience, designer and performer throughout the performance, and describe how our time based media annotation system can be used to support that methodology. We present the results of this grounded ontology methodology applied to the text-based commentaries between system engineer, performer and listener. We argue that by developing an understanding of these inter-related experiences we can understand the desired and potential role of computational systems in creative contexts which can help in the design of new systems and help us curate new kinds of performance scenarios.

Introduction

The field of computational creativity has exploded into life in the last five to ten years with a whole range of work that reaches across theories, computational architectures and systems. It is an important field for a number of reasons not least because it throws up a number of issues around understanding the human creative process, understanding how we can support that process with new systems, and how any such understanding can help us in novel approaches to the design of these systems. Moreover, it is an important field because it allows for non-traditional, perhaps more playful AI approaches to be considered. When the AI world is increasingly populated by big data, and deep learning seems to be conquering all, it provides an important counterfoil to the mainstream.

However, there are clearly issues with the word "creativity" and the multitude of definitions which currently exist (Still and d'Inverno 2016). These can refer to the output (such as the work of Boden who categorised different forms of creativity based on the resulting value and novelty (Boden 2004)), can refer to the nature of the specific person who is disposed to producing creative acts (Guilford 1957)), and can refer to the nature of the process undertaken to produce specific kinds of outputs (Csikszentmihalyi 2009). Just like the concepts of "agent" and "agency" that predominated the 1990s when it was almost impossible to write a paper without giving one's own definition of agency (Luck and d'Inverno 1995; d'Inverno and Luck 2003; Wooldridge and Jennings 1995), the positive side is that it allows for a whole array of innovative work. The negative is that - just as with agents - there is room for everyone and everything. In response, there is recognised need within the research community for clear methodological approaches that can evaluate autonomous computational systems which interact or collaborate in creative contexts with humans (Bown 2015).

In this paper, we respond to this need for appropriate methodologies by demonstrating how a grounded theory approach can be used to reflect upon human experiences around a new autonomous music improviser (AMI) called SpeakeSystem. The AMI was commissioned by the BBC's 'Jazz Line Up' programme in 2015 for a one off live performance at the Wellcome Trust in London with British saxophonist Martin Speake and which was also broadcast live on national UK radio.

Since we are interested in examining the human perspective and response to autonomous music systems, we align our work with d'Inverno and McCormack's promotion of 'collaborative AI' over 'heroic AI' (d'Inverno and McCormack 2015) and the interest of researchers such as Bown and Banerji in investigating the experience of musicians who play with these types of creative systems, and how they might be used as ethnomusicological probes (Bown 2015; Banerji 2012). This view is perhaps most in line with John Dewey who in his seminal work "Art as Experience" (Dewey 1934) looked to move the focus of thinking about art away from the object and towards the experience that takes place when we are making and experiencing art. Making and listening to music is a celebration of life, and it is through the experiences of making and listening where music - and all art - has its meaning.

The contributions of this paper are as follows:

1. Description of a methodology that can be used to inform (interactive music) system design based on analysis of precise discourse around time based media from the perspectives of performer, listener and algorithm designer.
2. A grounded ontology of time tagged comments made from the perspective of the human instrumentalist, a listener and the system designer that can inform the design of future systems and concert curation.
3. Description of a system for enabling shared annotation of time based media that supports the methodology and grounded ontology approach.
4. Documentation, source code and analysis data for an autonomous music improviser which was commissioned by the BBC 'Jazz Line Up' programme and which performed live in a high profile concert in 2015 (Yee-King 2016).

Research questions

The work is framed with the following research questions:

1. How can we design a methodology based around collaborative annotation of video or audio recordings of performances which can effectively inform the design of autonomous music improvisers?
2. How does this methodology validate and expand upon previous research around autonomous music improvisers?
3. Which aspects of live human/machine improvisation performances are of particular interest to listeners, performers and algorithm designers?
4. How can understanding the interplay between the experiences of designer, performer and audience help in the design of future systems (and curated concerts)?

Structure

In the following section, we discuss related work before discussing the implementation of the system itself. In the section entitled Evaluation Method, we describe our methodology for evaluating human experiences with our system. In the section called 'Results', we present our ontology and further information about our categories. In 'Analysis', we reflect upon our results and compare them to those of other researchers. In 'Concluding Thoughts' we re-state our research questions and how we have addressed them.

Related work

First, we shall consider the evaluation of systems designed to be used in creative contexts with humans. Bown reflects upon the state of affairs in creative systems evaluation, noting that the lack of empirical grounding for evaluations might be preventing the kind of iterated improvement seen in other areas of AI research (Bown 2014). As a solution, he promotes user based analysis in real creative contexts. Eigenfeldt noted that "some attempts have been made at evaluation" but that many systems are "idiosyncratic ... specific to the artist's musical intention" (and thus presumably difficult to compare to eachother) (Eigenfeldt 2015). We address these issues - we describe and demonstrate a specific, transferable methodology which explicitly aims to develop knowledge that can inform future iterations of AMIs. Whilst we agree with Bown's appraisal, we acknowledge that other researchers have made significant attempts to specify evaluation methodologies. Collins proposed three areas in which AMIs can be evaluated: technically, aesthetically (audience reaction) and in the sense of interaction for the musicians (Collins and D'Escriván 2007) and Stowell et al. described a range of techniques that are suitable for evaluating live human-computer improvisation systems, including Turing Tests, audience surveys and task analyses (Stowell et al. 2009). Both schemes include aspects of human experience, but it is not the main focus. Hsu and Sosnick describe an HCI framework that directly considers human experience, where usability for the musician and musical interest for the audience of AMIs are evaluated using survey instruments (Hsu and Sosnick 2009). Subsequently to the work above, Bown provided a qualitative, thematic analysis (Clarke and Braun 2006) of musicians' experiences with his Zamyatin system (Bown 2015). Finally, Banerji reported an ethnographic approach to analysing how musicians changed their playing in response to an AMI, placing the system in a kind of socio-cultural map (Banerji 2012). We will contextualise our work by relating it directly to some of this previous work on the evaluation of AMIs

Human experience is also considered in non-music specific evaluation methodologies. It appears in one leg of Colton's "Creative Tripod", (skill, *appreciation* and imagination), but only the audience is considered, since Colton's work is focused on machine only creation (Colton 2008). Jordanous' Standardised Procedure for Evaluating Creative Systems (SPECS) provides a set of components of creativity, several of which relate to human-in-the-loop type interaction and experience, e.g. component 10, Social Interaction and Communication (Jordanous 2012). We shall revisit SPECS in our analysis later.

Considering the specific methodology used in this paper, we conduct a qualitative discourse analysis with a grounded theory method (Glaser and Strauss 1967). Grounded theory is chosen as it is suitable for the extraction of an ontology that can describe a discourse (Stern 2007). Our grounded theory approach consists of iterated data collection and categorisation followed by theory construction, in the form of an ontology (Birks and Mills 2011, p10). Whilst it is widely used, particularly in the social sciences, for qualitative analysis we note that grounded theory is not a panacea and that since its development in 1967, it has split into dialects and has been criticised for being overly dogmatic in its insistence upon emergent analysis as opposed to mapping analysis to existing theory (Goldkuhl and Cronholm 2010). Goldkuhl et al's Multi-Grounded Theory provides a solution to this, wherein pre-existing theory is mapped back onto the emergent theory (Goldkuhl and Cronholm 2010). We take this into account in our analysis, connecting our grounded ontology to Bown's thematic analysis and Jordanous' SPECS components.

To conduct the data collection phase of our analysis, we made use of a collaborative media annotation system called MusicCircle which allowed the participants to discuss very specific parts of the performance in a solitary, then a collaborative phase (Brenton et al. 2014). In a sense, we wanted the annotators to become ethnographers, where they were thinking about what the musician and the system were doing, as they were doing it and the use of the annotation system allowed them to focus on specific aspects of the 'exhibited behaviour'. For a reference point, consider the ethnographic approach described in (Barthet and Dixon 2011).

Now, we shall consider the second area of previous work: AMIs that are technically similar to SpeakeSystem. SpeakeSystem uses an hierarchical Markov model which is trained in real time from an audio stream. Pachet's Continuator built Markov models from MIDI input in real time and used them to generate stylistically related, MIDI output (Pachet 2002). Yee-King has reported a series of systems that carry out timbral and symbolic sequence analysis and mimicry, including a 'Matt Yee-King simulator' that was based on Markov modelling of MIDI input (Yee-king 2011; Yee-King 2007; Nort 2014). Hsu's timbral improvisation systems built non-Markovian, hierarchical models of timbral features from a live audio input (Hsu 2008). Collins' FinnSystem used a pre-trained model of saxophonist and flautist Finn Peters combined with realtime audio analysis to control the output of the model and improvise (Nort 2014). Bown's Zamyatin system used an evolved decision tree to move between target behaviours during live improvisation (Bown 2015). All of these systems were designed to operate in a human-machine creative context.

Implementation of the SpeakeSystem

The system was developed in the SuperCollider environment, and consists of essentially 3 modules: input, modelling and output. The input module shown in Figure 1 is responsible for generating a stream of labelled events and a stream of pitches from an audio signal obtained from a microphone. Event labels consist of *event type*, either note or silence and the *quantised length*. For example, note_500 would be a 500 ms long note. It was designed to work with monophonic instruments, but could be adapted to polyphonic instruments, given a sufficiently reliable polyphonic pitch tracker (or MIDI input).

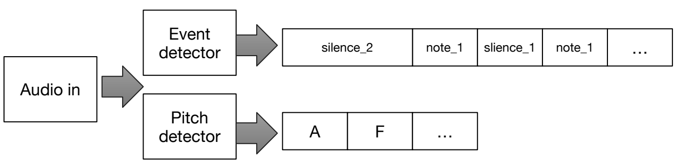

Figure 1: The input module analyses audio into a series of silences and note events and a series of detected pitches.

The modelling module consists of two multi-order Markov chains, one for pitches and one for events. As an example, the sequence of pitch labels a, b, b, d would result in several, different order entries to the pitch chain:

- $a \rightarrow b$
- $b \rightarrow b$ and $b \rightarrow d$ would be combined to make $b \rightarrow [d, b]$.
- $ab \rightarrow b$
- $bb \rightarrow d$
- $abb \rightarrow d$

The resulting chain is visualised in Figure 2.

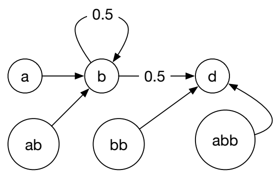

Figure 2: The pitch chain resulting from the input *abbd*.

Pitches and events could have been stored in a single chain but the output of the system was more varied when it was able to model pitch and event sequences separately, as it could generate similar rhythms with different notes to those played and vice versa. The output module ran the Markov chains in generative mode to make a sequence of events. The length and type of event was taken from the event chain and the pitch of note events was taken from the pitch chain. Considering the above input, and just the pitch chain, the initial note would be a, b, b or d, with 25% chance of a, 50% chance of b and 25% chance of d. If b was chosen, the generator state would be b, so there would then be 2 options: b or d, with equal chance. The system would always choose the highest order option that had at least two possible next steps; if only one option was available for *bb*, it would shorten its state description from *bb* to b and look up the options following state b, if only one option was available for b, it would pick from the distribution of all observed single notes. This combined accurate modelling with an interesting level of variation in the output.

The chosen pitch and duration would be used to generate MIDI note on and note off messages which were sent to an Access Virus C hardware synthesizer running in monophonic mode. The synthesizer was programmed with a sound which combined subtractive synthesis with some frequency modulation. There was no technical reason for choosing a hardware synthesizer over synthesis inside SuperCollider, but the Access Virus C is considered to have a very distinctive and powerful sound palette. The choice of a note based system as opposed to a more timbral system was made based on a discussion with the producer of the radio programme who commissioned the work, who pointed out that the performance was to be broadcast on the mainstream jazz show 'Jazz Line Up', as opposed to its more experimental counterpart 'Jazz on 3'. The output module had some additional features which were designed to make it a more interesting improviser:

1. *Riffing with diminishing energy.* The system plays varying length sequences of notes wherein the modulation index of the FM synthesis was reduced in variably sized steps.

2. *Leaky models.* The system 'forgets' the training data leading to temporarily naive output. The aim was to provide a more structured feel to the piece.
3. *Separate timing and pitch models.* The system stored separate models of event and pitch sequences, so it could combine separate timing and pitch structures from the audio input. We aimed to provide more interesting and varied output over simply mimicing the performer.

The system was developed against a recording of a saxophone improvisation provided by Martin Speake prior to the performance. The various characteristics described above were hand tuned to maximise the musicality of the system when it was playing against the fixed recording.

Evaluation Method

In this section, the method by which the system was evaluated is described. In summary, a live performance was recorded and uploaded to a collaborative annotation system called MusicCircle. The performer, system designer and a listener annotated the recording, then a grounded theory approach was used to analyse the annotations they made. A DOI'd github repository providing a recording of the performance, the system source code and the annotation dataset can be found at (Yee-King 2016).

Performance

The piece was performed by Martin Speake, an experienced British jazz saxophonist, playing alto saxophone and the system, as specified in the previous section. A photograph of the 'performers' is shown in Figure 3, but we note that the computer operator was simply there to execute the autonomous system and to set the output level. Martin had not previously performed with the system or any other autonomous improviser, aside from a short technical test in the sound check on the night. He knew that he was performing with an autonomous system but he was given minimal insight into its design. The performance was recorded live at the Wellcome Trust on 26th September 2015 and simultaneously broadcast on BBC Radio 3.

Figure 3: Matthew Yee-King and Martin Speake at the live performance. The system was autonomous but Matthew had to adjust the volume level at the start of the performance.

The annotation system

The recording of the improvisation was then annotated using a system we have developed called MusicCircle. MusicCircle was developed within a European research project and is available through www.museifi.com.

It has certain key features which were not available in other systems and which make it an appropriate tool for a range of applications including research and education. It can be classified as a scalable, web based, collaborative, time based media annotation tool and it has been used by several thousand students and researchers. For a more in depth discussion of MusicCircle and how it was developed, we refer to (Brenton et al. 2014) and (Yee-King et al. 2014). For the purposes of this work, MusicCircle allows its users to select regions of an audio or video file and to enter text comments which are then attached to the regions. Each user's annotations are displayed along a 'social timeline', which shows each person's commentary as a series of coloured blocks. Clicking on a coloured block reveals the comment and allows replies to be added. Each annotated region can then become a separate discussion thread. Figure 4 shows the user interface of the annotation system, where the recording of the improvisation has been annotated by 3 different people, as described in the next section.

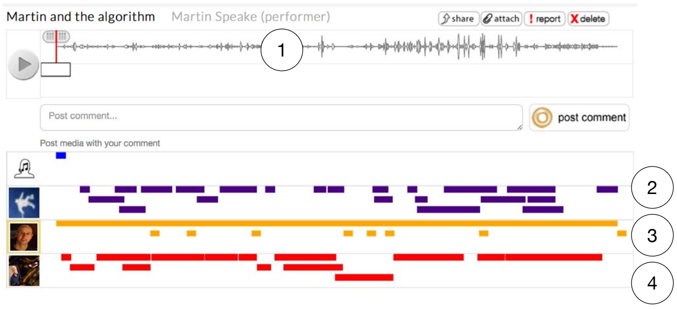

Figure 4: The annotation system, showing the time series of the recording (1) and 3 sets of annotations below, from the algorithm designer (2), performer (3) and the listener (4).

Annotation and tagging protocol

The concept of the annotation and tagging protocol was to obtain 3 independent perspectives on the improvisation in the form of time linked annotations, then to use an iterated grounded theory approach to create a set of tags categorising the annotations. The algorithm designer, the performer and a listener carried out the annotation. The listener did not attend the live concert and is a jazz music and autonomous agent expert, so is not a typical listener. In a future study, we would gather annotations from a wider range of listeners. The following protocol was followed:

1. A recording of the complete performance which lasted 3m 35s, was obtained and uploaded to the annotation system.
2. Each person was provided with a login for the system and their own copy of the recording for annotation.
3. They were asked to select regions of the recording that were interesting to them and to explain in the comment attached to the region why that region was interesting. They could not see each others' annotations at this stage.

4. The annotations were combined onto a single timeline, as shown in Figure 4.
5. The annotators were asked to read each other's annotations and type replies if they wished. This marked the end of the annotation phase.
6. In the tagging phase, the algorithm designer read through the comments and replies, assigning tags to each.
7. The process of reading comments and adding tags was repeated until no new tags were needed and no comments needed to have any more of the existing tags added to them.

The production of the set of tags through the above protocol represented the initial and intermediate coding stage of grounded theory (Birks and Mills 2011, p9). Following this stage, tags were organised into a hierarchy of categories. This was achieved by considering each tag in turn and identifying whether that tag could be placed as a sub tag of any of the other tags. A constraint that each tag could only have one parent tag was imposed to simplify the process but it was found this did not induce excess 'stress' in the structure; each tag either stood alone or fit well beneath another. After this stage, we refer to the tags as categories, and the overall set of categories as a grounded ontology.

Results

Figure 4 shows all of the annotations as they appear in the user interface of the annotation system. There were 46 comments and 23 replies which were placed into 51 categories. An annotation could belong to several categories, and the number of categories assigned to an annotation varied between 1 and 8 with a rounded average of 4 categories per annotation. The number of annotations per category varied between 1 and 24, with a rounded average of 3 annotations per category. Tables 1, 2 and 3 show category frequencies for each of the three annotators. The frequency value is relative to the total number of categories assigned to annotations by that person, to make the numbers more comparable by compensating for the fact that different people left different numbers of annotations.

Frequency	Category
0.08	interaction
0.08	algorithm leading
0.07	autonomy
0.05	space
0.04	real
0.04	conversation
0.04	musician leading
0.04	structure
0.04	collaboration
0.04	roles

Table 1: Most popular categories for the listener

Figure 5 shows the grounded ontology that was derived from the process described in the previous section. Each category has a number next to it which is the number of annotations that were assigned to that category (this value does not include sub categories). The thickness of the border around the categories indicates this information visually,

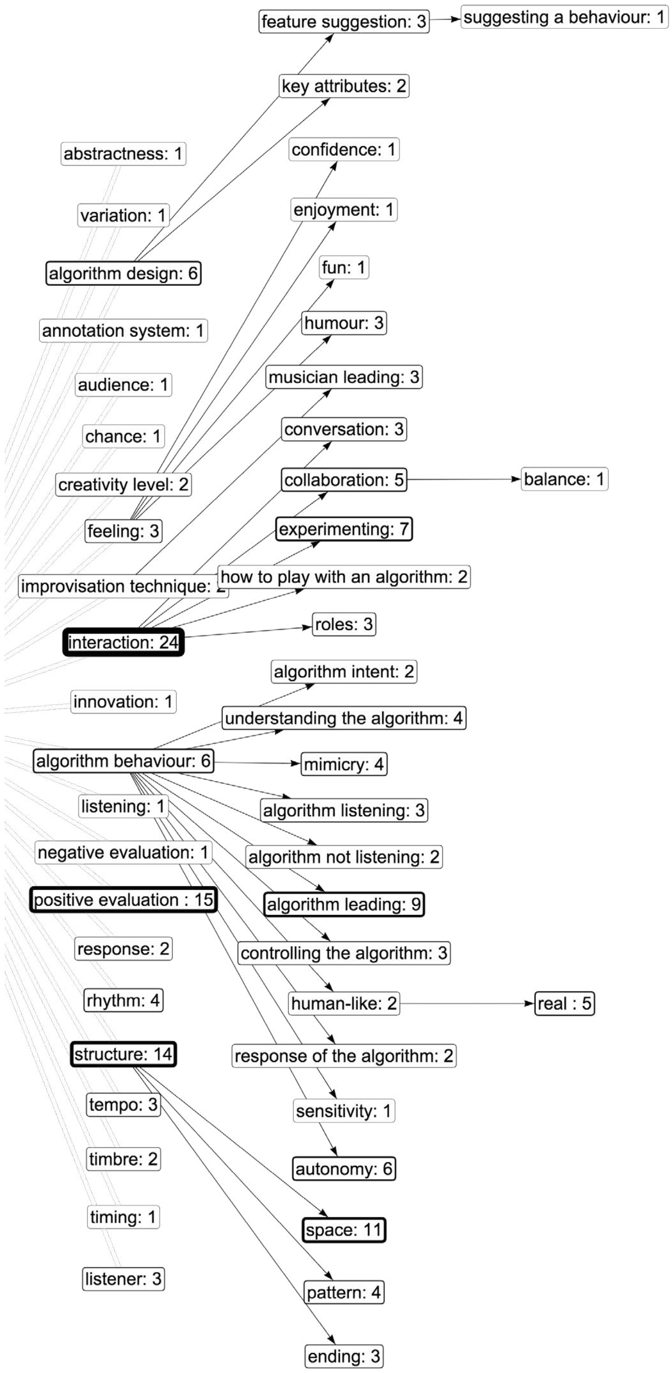

Figure 5: The ontology that was derived from analysis of the annotations left by the three participants. The number of annotations which were assigned to a category is indicated after the category name. The thickness of the border for a category visually indicates the number of annotations in that category. Counts are for that category only, not the category and its sub categories.

Frequency	Category
0.11	experimenting
0.09	structure
0.07	interaction
0.07	positive evaluation
0.07	space

Table 2: Most popular categories for the performer

Frequency	Category
0.15	interaction
0.10	positive evaluation
0.07	structure
0.05	algorithm design
0.04	pattern
0.04	space
0.04	algorithm behaviour

Table 3: Most popular categories for the algorithm designer

where thicker bordered categories had more annotations assigned to them.

Analysis

In this section, we will consider the results of the grounded analysis as they compare to other analyses. Jordanous empirically derived a set of 14 linguistic 'components' for the SPECS framework which were statistically more likely to be present in a corpus of research papers about creativity than in a corpus of research papers not about creativity (Jordanous 2012). Whilst the components were developed for the purposes of measuring the creativity of computational systems, we find them useful in framing our analysis.

Taking the category *Algorithm Leading* from the ontology which was assigned to 9 annotations, we can connect it to Active Involvement (SPECS component 1), Independence and Freedom (component 6), Social Interaction and Communication (component 10) and Value (component 13). So the system exhibited creative behaviour, but this is not the main focus of our work; perhaps we would prefer to consider if the human exhibited creative behaviour as a direct result of the actions of the system. The *Experimenting* category was assigned to 7 annotations, and looking at the commentary from the musician, they carried out three phases of deliberate experimentation to understand the behaviour of the algorithm. Experimenting links to Thinking and Evaluation (component 12), Variety, Divergence and Experimentation (component 14), Dealing with Uncertainty (component 2) and Social Interaction and Communication (component 10). The system seemed to encourage creative behaviour on the part of the human musician in a range of areas, and we can qualify this with reference to SPECS.

Next, we shall consider Bown's thematic analysis which derived four key themes from a focus group discussion with musicians who had played with an AMI (Bown 2015). We contrast our approach with Bown's approach in two key areas: 1) the method used to gather the data and 2) the method used to analyse the data. Our data gathering method was different in that our annotation system forced comments to be connected to a very specific region of a recording, so each comment came with explicit, musical evidence. Our data analysis method was different as it involved a categorical rather than a thematic analysis.

Bown's (paraphrased) themes were 1) interacting with the system gave a stronger sense of the nature of the interaction than watching someone else interact 2) there was an interest in the tangibility of the rules the system was using 3) participants did not refer to the system as a virtual musician, rather as an instrument or composition 4) they felt that long term structure was lacking.

Does our analysis support Bown's themes? The contrast between interacting with the system and listening to someone else doing it did not appear in our ontology; perhaps a comparison of the categories connected to the listener's annotations and those for the musician would shed some light here but we should note that Bown explicitly had the musicians listen to eachother performing, but we did not. Understanding and discussing the rules used by the system was very evident in our data, as represented in particular by the *Algorithm behaviour* category and its sub-categories, which were used 41 times in total. The grounded analysis did not pick up on different ways of referring to the system but with hindsight, each annotator did have a different way of referring to it - the listener, who was an autonomous agent expert, decided early on in their annotations how they would refer to it:

[Listener]... amazed that the CMA (short for computational music agent or algorithm - someone else can decide)

The Musician referred to it as 'the computer':

[Musician]... to see how the computer would respond

The algorithm designer used 'the algorithm' or 'it':

[Designer] The algorithm picks up well on the rhythm here ...

Regarding long term structure, we tagged 14 annotations (over four times the average per category) with the 'structure' category, suggesting this was a strong theme in our dataset, given the assumption that structure refers to the compositional structure.

In summary, we found evidence for three of the themes identified by Bown, though we had to retrospectively look at the annotations for the 'referring to the system' theme, and this theme was a necessity in a sense as the commenters had to refer to the system somehow. Despite this, the fact that these themes emerged from two quite different data gathering and analysis approaches, with different people and different systems supports Bown's findings and supports the validity of our findings.

Examining system design decisions

Our annotation methodology enables a very precise connection between the commentary, its derived ontology and specific sections in the recording of the performance. This allows us to consider the impact or otherwise of system design decisions upon the performance - when the system exhibits behaviour as a result of certain features, is this noticed by the annotators? As mentioned in the system description earlier there were three distinct features which aimed to produce more interesting output: *Riffing with diminishing energy*, *Leaky models* and *Separate timing and pitch models*.

We can map these features directly to comments such as the performer responding to the result of leaky models:

> *I slowed down my activity to the one long held note with the computer repeating it as separate notes until it seemed to give up realising I had finished playing! The audience felt like they were really with me/us in the moment too with their laughter at the end.*

We can then look in the ontology for the categories that are associated with this comment: 'humour', 'structure' and 'space'. Another example of the leaky models being noted (again, from the performer):

> *Yes i did wonder as sometimes it seemed to have logic in how it responded and then at other times it didn't make sense to me.*

Here is an example of the listener responding to a section of the performance where the separate model feature was prominent:

> *Here we hear there very high notes which the CAM seems to "hear" and then responds to them in different ways each time.*

In this way, we can consider key system design decisions and look for evidence that they had an impact upon the human experience, without the humans needing to understand how the system worked. This is similar to the unlocking of tacit knowledge made possible by user centred design.

A final potential application of this technique is that it might be used to inform the curation of concerts involving human/machine improvisation (McCormack and d'Inverno 2016). We can take the key items in the ontology and turn them into a set of challenges for algorithm designers - 'create a humorous algorithm which experiments with algorithm leading and human leading', 'create an algorithm which uses space to encourage experimentation on the part of the human', and so on.

Concluding thoughts ...

In this paper we have described and evaluated a new autonomous music improviser using a novel methodology. Here are the research questions stated at the start of the paper, with brief summaries of how we have addressed them. We start with the first two together.

1. *How can we design a methodology based around collaborative annotation of video or audio recordings of performances which can effectively inform the design of autonomous music improvisers?*
2. *How does this methodology validate and expand upon previous research around autonomous music improvisers?*

Response: Our methodology uses a social, time based media annotation system to enable focused annotation then discussion of human/machine performances. We have shown how this data can then be further analysed through grounded theory to yield an ontology that describes the resulting discourse. We have shown how the output of this method can be compared with that from other methodologies and that we are able to contrast and compare these results.

3. *Which aspects of live human/machine improvisation performances are of particular interest to listeners, performers and algorithm designers?*

Response: We derived and presented a grounded ontology describing the themes observed in a set of annotations left on a specific human/machine performance by a listener, a performer and an algorithm designer. We found that key themes included interaction, structure, space and algorithm behaviour. We were also able to verify our themes by mapping them to those described by previous, related research.

4. *How can understanding the interplay between the experiences of designer, performer and audience help in the design of future systems and concerts?*

Response: We have described how the kind of highly specific annotations and analysis enabled by our methodology can provide evidence for the impact of system design decisions upon the experience of listeners and performers. It is interesting to note that this can also inform the curation of concerts of such systems, where we can perhaps use our ontology to provide a list of challenges for system designers. In this way we can communicate interesting research themes in the field to the wide range of participants and investigate the themes through practice based activity.

We believe that key to designing systems that enable human/machine improvisation is starting from the perspective of the unfolding human experience, not just in music but in all forms of human creative activity.

.. and an Epilogue

A quote from the very beginning of Dewey's seminal book *Art and Experience* feels appropriate here. Back in the 1930s Dewey argued that it is experience that is key to understanding the nature of art and creative endeavour:

> *In common conception, the work of art is often identified with the building, book, painting or statue in its existence apart from human experience. When an art product attains classic status, it somehow becomes isolated from the human conditions under which it was brought into being and from the human consequences it engenders in actual life-experience. When artistic objects are separated from both conditions of origin and operation in experience, a wall is built around them that renders almost opaque their general significance ... The task is to restore continuity between the defined and the everyday events, doings and sufferings that are universally recognised to constitute experience.*

Our view is that it is of little practical interest to consider the amount, or system, of "creativity" contained within a computational system, but much more compelling to design systems that provide new kinds of creative experiences and opportunities for us all.

Acknowledgements.

We would like to thank the British saxophonist Martin Speake for generously taking part in this experiment and to Arthur Still for introducing us to the work of John Dewey.

References

Banerji, R. 2012. Maxine's Turing TestA Player-Program as Co-Ethnographer of Socio-Aesthetic Interaction in Improvised Music. In *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment of the Artificial Intelligence and Interactive Systems Conference*, 2–7.

Barthet, M., and Dixon, S. 2011. Ethnographic observations of musicologists at the British Library: Implications for music information retrieval. In *Proceedings of the 12th International Society for Music Information Retrieval conference (ISMIR 2011)*, number Ismir, 353–358.

Birks, M., and Mills, J. 2011. Essentials of grounded theory. *Grounded theory: a practical guide* 1–14.

Boden, M. 2004. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

Bown, O. 2014. Empirically Grounding the Evaluation of Creative Systems: Incorporating Interaction Design. In *Fifth International Conference on Computational Creativity*.

Bown, O. 2015. Player Responses to a Live Algorithm: Conceptualising computational creativity without recourse to human comparisons? In *Proceedings of the Sixth International Conference on Computational Creativity*, 126–133.

Brenton, H.; Yee-King, M.; Grimalt-Reynes, A.; Gillies, M.; Kirvenski, M.; and d'Inverno, M. 2014. A Social Timeline for Exchanging Feedback about Musical Performances. In *British HCI Conference*, 1–6.

Clarke, V., and Braun, V. 2006. Using thematic analysis in psychology. , 3(2):77101, Jan. 2006. [5]. *Qualitative Research in Psychology,* 3:77–101.

Collins, N., and D'Escriván, J. 2007. *The Cambridge companion to electronic music*. Cambridge University Press.

Colton, S. 2008. Creativity Versus the Perception of Creativity in Computational Systems. In *Proceedings of the AAAI Spring Symposium on Creative Systems*, 14–20.

Csikszentmihalyi, M. 2009. *Flow*. HarperCollins.

Dewey, J. 1934. *Art as Experience*. New York: Perigree Books.

d'Inverno, M., and Luck, M. 2003. *Understanding Agent Systems*. Springer.

d'Inverno, M., and McCormack, J. 2015. Heroic versus collaborative AI for the arts. In Yang, Q., and Wooldridge, M., eds., *IJCAI International Joint Conference on Artificial Intelligence*, 2438–2444. AAAI Press.

Eigenfeldt, A. 2015. Generative Music for Live Musicians: An Unnatural Selection Real-time Notation. In *Proceedings of the Sixth International Conference on Computational Creativity*, 142–149.

Glaser, B. G., and Strauss, A. L. 1967. The discovery of grounded theory. *International Journal of Qualitative Methods* 5:1–10.

Goldkuhl, G., and Cronholm, S. 2010. Adding Theoretical Grounding to Grounded Theory: Toward Multi-Grounded Theory. *IJOM: International Journal of Qualitative Methods* 9(2):187–206.

Guilford, J. P. 1957. Creative abilities in the arts. *Psychological Review* 64(2):110–118.

Hsu, W., and Sosnick, M. 2009. Evaluating Interactive Music Systems An HCI Approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 25–28.

Hsu, W. 2008. Timbre-aware improvisation systems. In *Proceedings of ICMC2008, International Computer Music Conference, Belfast, UK*.

Jordanous, A. 2012. A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative. *Cognitive Computation* 4(3):246–279.

Luck, M., and d'Inverno, M. 1995. A formal framework for agency and autonomy. In *Proceedings of the First International Conference on Multi-Agent Systems*, 254—260. AAAI Press/MIT Press.

McCormack, J., and d'Inverno, M. 2016. Designing improvisational interfaces. In Pachet, F.; Cardoso, A.; Corruble, V.; and Ghedini, F., eds., *Title: Proceedings of the 7th Computational Creativity Conference (ICCC 2016)*. Universite Pierre et Marie Curie.

Nort, D. V. 2014. Sound and Video Anthology: Program Notes. *Computer Music Journal* 38(4):119–127.

Pachet, F. 2002. The continuator: Musical interaction with style. In *Proceedings of the International Computer Music Conference, ICMA, Gotheborg*, 333–341.

Stern, P. 2007. Grounded theory methodology: Its uses and processes. *Journal of Nursing Scholarship* 12(1).

Still, A., and d'Inverno, M. 2016. A history of creativity for future AI research. In Pachet, F.; Cardoso, A.; Corruble, V.; and Ghedini, F., eds., *Title: Proceedings of the 7th Computational Creativity Conference (ICCC 2016)*. Universite Pierre et Marie Curie.

Stowell, D.; Robertson, a.; Bryan-Kinns, N.; and Plumbley, M. D. 2009. Evaluation of live human-computer music-making: Quantitative and qualitative approaches. *International Journal of Human Computer Studies* 67(11):960–975.

Wooldridge, M., and Jennings, N. 1995. Intelligent agents: Theory and practice. *Knowledge engineering review* 10(2):115–152.

Yee-King, M.; Krivenski, M.; Brenton, H.; and d'Inverno, M. 2014. Designing educational social machines for effective feedback. In *8th International Conference on e-learning*. Lisbon: IADIS.

Yee-King, M. J. 2007. An Automated Music Improviser Using a Genetic Algorithm Driven Synthesis Engine. volume 4448 of *Lecture Notes in Computer Science*, 567–576. Springer.

Yee-king, M. J. 2011. An Autonomous Timbre Matching Improviser. *Proceedings of ICMC2011, International Computer Music Conference, Huddersfield*.

Yee-King, M. 2016. speakesystem: ICCC ontology release http://zenodo.org/record/46232.