
Young, Michael W. and Bown, Oliver

Clap-along: A negotiation strategy for creative musical interaction with computational systems

Originally published in Proceedings of the International Conference on Computational Creativity 2010 Copyright Departament of Informatics Engineering University of Coimbra. The publisher's version is available at: http://creative-systems.dei.uc.pt/icccx Please cite the publisher's version.

Clap-along: A negotiation strategy for creative musical interaction with computational systems

Michael Young and Oliver Bown

[1] Music Department, Goldsmiths, University of London, New Cross, London SE14 6NW, UK
m.young@gold.ac.uk
[2] Centre for Electronic Media Art, Monash University, Clayton 3800, Australia
oliver.bown@infotech.monash.edu.au

Abstract. This paper describes *Clap-along*, an interactive system for theorising about creativity in improvised musical performance. It explores the potential for negotiation between human and computer participants in a cyclical rhythmic duet. Negotiation is seen as one of a set of potential interactive strategies, but one that ensures the most equitable correspondence between human and machine. Through mutual negotiation (involving listening/feature extraction and adaptation) the two participants attempt to satisfy their own and each other's target outcome, without knowing the other's goal. Each iteration is evaluated by both participants and compared to their target. In this model of negotiation, we query the notion of 'flow' as an objective of creative human-computer collaboration. This investigation suggests the potential for sophisticated applications for real-time creative computational systems.

1 Introduction

Music performance is a creative, 'real-time' process that often entails collaborative activity. Creative potential in a performance – i.e. the opportunity for individual or collective actions that appear innovative – is contingent on context and style, the presence of *a priori* agreements (whether explicit or tacit), and other cultural and procedural elements [1]. Various aspects of performance practice constrain or afford opportunities for immediate creative input, such as recourse to pre-existing materials, the relative emphasis on individual responsibility and the means by which information is exchanged while performing. In human-computer performance the capacities of software to exchange information with other participants and to take responsibility (i.e. act autonomously) are highly significant factors. Any given approach to these factors impacts greatly on the performance practice as a whole.

Free collective improvisation is an effective testing ground for computer-based creativity. The computer must be able to produce sonic events that appear intrinsically valid, and it must be able to collaborate appropriately (responsively and proactively) with one or more human musicians. Both properties satisfy a working definition of computational creativity, such as the ability to exhibit behaviour "which would be deemed creative if exhibited by humans" [2] as well as being "skillful, appreciative and imaginative" [3]. These criteria are only truly satisfied if the system is not directly reliant on a human 'user'; there must be an equitable correspondence of ideas between collaborators. Such challenges are not easily met, but at least apply equally well to human-only improvisation, where the behaviour of one performer would never be expected to depend entirely on another's contribution, or depend on rules agreed in advance. Ideally at least, group improvisation avoids organisational procedures that determine or influence musical content, structure or the interactions and mutual dependencies between performers. Implicit procedures may develop through a process of negotiation in performance. As in other forms of process-orientated art, this process may not be directed towards a known outcome. Rather, the process itself forms both the central problem and focus of interest that both enables and constitutes the performance.

We investigate negotiation as a musical process in a wider context of interaction strategies that can demonstrate performance-based computational creativity. We devise a simple system, *Clap-along*, defined by a number of constraints, which we believe demonstrates the challenges of performative, computational creativity, and offers a promising model for future, more elaborate, computational systems for interactive musical performance.

2 Interaction strategies

We regard negotiation as a specific strategy for human-machine interaction, a member of a larger set that includes the following list. (The terms 'source' and 'result' are used to refer to actions in an asymmetric relationship, and may apply equally to human and machine depending on context):

Shadowing: The source and result move together. There is a clear temporal simultaneity that produces layering within a coordinated motion. The coordination of motion between a body and its shadow is simultaneous but also distorted, because the shadow is projected into a different 'geometrical' space. Various musical strategies for textural organisation (homophony, micro-polyphony) entail shadowing. Real-time digital effects and simple interaction systems commonly exhibit shadowing methods. Timbral matching techniques can be thought to employ this strategy, as found in the Cata-RT [4] and Soundspotter systems [5]. The system as a whole may be weakly or strongly integrated, but in general interactivity is likely to be readily verifiable to participants and listeners.

Mirroring: The source is reflected in the result. Synchronicity is not required. There is a more elaborate re-interpretation of information received from the source than in shadowing, and this is more telling in the absence of temporal synchrony. Innumerable compositional and improvisational approaches are analogous to mirroring, to be found in structural repetition, motivic development and call-and-response strategies. Delay effects are elementary mirrors, projecting back an image of the musical source. Systems that seek to analyse an existing style to generate music (e.g. by Markov modelling) are mirroring at a high level of musical organisation, such as in the Continuator system [6], even though the method may be sub-symbolic. Mirroring can establish a cohesive musical identity between source and result, and may also be readily verifiable to participants as an interactive process.

Coupling: Two sources are distinct but connected. There is – or appears to be – mutual influence, but the roles of 'source' and 'result' may be unstable and unequally balanced. In live performance, coupling can constitute "procedural liveness" and/or "aesthetic liveness" [7]. For instance, coupling can be trivial and procedural, when one system controls another and receives appropriate feedback, as in the laptop-as-instrument paradigm. Or, it can be illusory and "aesthetic", as in the (increasingly rare) genre of music for live instrument with tape, where binding and apparently causal links between the performer and tape are in reality entirely controlled and pre-determined. More abstract couplings that use virtual modelling and dynamical systems offer a more open-ended and genuine relationship between sources (agents), potentially integrating the procedural and the aesthetic. Examples include music systems where sources share a virtual environment, as in the Kinetic Engine [8] and Swarm Granulator [9]. In coupling, equal relationships are possible, but verification of the degree of true interaction is problematic.

Negotiating: The roles of 'source' and 'result' are conflated. Participating elements have equal status and are engaged in a series of transactions. Negotiation treats performer and computational system as equal in status and overall approach to the interaction. It can be seen as a unified system based on equivalence, and this contrasts with the categories above that suggest an unequal architecture of source and result, in which the most likely scenario is a musician (source), acting upon a computational system (result).

According to OED definitions[3] 'negotiation' refers to transactions directed towards an objective:

1. *To communicate or confer (with another or others) for the purpose of arranging some matter by mutual agreement; to discuss a matter with a view to some compromise or settlement*
2. *To arrange for, achieve, obtain, or bring about (something) by negotiation*
3. *To find a way through, round, or over (an obstacle, a difficult path, etc.)*

To negotiate, participants engage in a series of transactions that are guided by local goals directed towards an individual desired outcome (expectation). The transactions can be understood to involve two mutually informing strategies, one externalised, the other internal; "action" and "description" [10]. Actions executed within a system may instigate further changes of state in the system. An assessment of these changes (especially in relationship to anticipated or desired changes) forms a description of action-outcome in the system. This empirical description may inform the next action. If so, a cyclical process of experiential accumulation develops, as the total description becomes more detailed and complex. In a pre-determined and constrained context, this accumulation might be

[3] Accessed online, 18th September 2009

understood as a straightforward 'making sense of it'. But in a more process-orientated context there may be an emergent formulation of knowledge that is not external to the system (i.e. the system is not pre-determined). Hamman [10] uses Foucoult's term 'episteme' to describe interactive processes that are "immanent in the very particularity of the thoughts, actions, and descriptions made with respect to a hypothesised object of interaction" (p. 95). This is an open-ended process of negotiation, orientated by variable or uncertain expectations. Local goals (intentions) might change as a product of the interactive transactions underway. Desired outcomes (expectations) need not be static either, so the OED characterisation of a "compromise or settlement" may remain theoretical and notional.

A reciprocal negotiation entails a degree of equivalence between human and machine capacities to act and formulate descriptions. Both participants must form their own descriptions of the system that incorporates the other participant. Both must be able to modify their actions, short-term goals and intentions, given new information. In other words, they should also be able to formulate an expectation about the overall musical output and modify their contribution, given the other's, in order to best satisfy the expectation. Verification that transactions are underway is, in itself, a part of this process, but the accuracy or efficacy of any "description" is not relevant to the fact that negotiation occurs.

We regard "optimal flow" [11] as a directly relevant but problematic concept. Flow is the human enjoyment derived in undertaking a task that becomes autotelic, achieved when there is an optimal level of difficulty relative to the skills of the subject. This balance requires the subject to form an internal description of the task's demands and an assessment of his/her skills in meeting them. Flow has been explored in human-computer interaction [12] and in the creative process [13] including group-based creativity [1]. Particularly in the case of creativity, Csikszentmihalyi notes a number of factors contributing to flow, some of which could be modelled (clarity of goals, availability of immediate feedback) and others perhaps not (no worries of failure, no distractions etc.) [13]. Whereas this might describe a positive and productive psychological state, flow perhaps does not take fully into account other facets of creativity, or for that matter the experience of negotiating; the role of randomness, unpredictability, happenstance [14], the use of haphazard trial and error, periods of incubation and, subsequently, innovative "behavioural mutations" [15]. More emotively, consider the pursuit of the impossible, the thin borderline between absorption and obsessive compulsion [15], and, perhaps resultant periods of boredom or frustration. Flow describes a settled state that may be too effortless in itself to be central in establishing creativity. So we attempt to avoid an easily achievable sense of 'flow' in designing the *Clap-along* system.

3 Implementation of Clap-along

Our aims in *Clap-along* are to explore process, expectation and verification in a negotiation-based system, and to consider how these elements might ultimately be extended to produce more aesthetically complex and musically valid results. Negotiation occurs both in a feature space and in the foreground surface of actual rhythms.

Clap-along is a duet system for human-computer interaction. Both participants produce continuous, synchronised 4-bar clapping patterns in 4/4. The musical context is as minimal as we can conceive: a fixed tempo, a fixed metrical structure, and single sound events quantised to beats.

In any loop instance n there is a human clapping pattern, H_n, a computer pattern, C_n, and a composite of the two patterns, R_n. A feature set, F_n, is extracted from R_n and compared by the system to a target feature set T_n, and this comparison is the basis for the next iteration. R_n is the reality of the current state, and T_n represents an expectation that is unknown to the human performer.

The computer maintains patterns as a sequence of 0s and 1s, representing either a rest or a clap on each beat. The initial pattern C_0 is generated randomly. Machine claps are generated from a sample bank of human clap recordings that have some natural variation in sound; this offers some semblance of human clapping. The human performer claps into a microphone, allowing the system to build a second binary sequence that represents the human's clapping rhythm, H_n. Human claps are quantised by rounding down to the previous beat, unless within 200ms of the following beat, in which case they are rounded up. The initial expectation, T_0, is obtained from a randomly created composite rhythm R_{target} that is immediately discarded.

At the end of each 4-bar pattern, the system takes the composite of the two rhythms, R_n, and calculates a feature set F_n that forms a minimal internal representation of the musical output. The four features used in this version are:

- density: the total number of claps as a fraction of the maximum possible number.
- homophony: the number of coincident claps as a fraction of the maximum possible value.
- position weighting: the normalised average position in the cycle over all claps.
- clumping: the average size of continuous clap streams as a fraction of the maximum possible value.

In this multi-dimensional feature space, the system calculates the Euclidean distance between F_n and the target feature set T_n. If the distance exceeds a pre-defined threshold, this is deemed to indicate a significant musical difference between reality and the expectation. We use a threshold (s) of 0.001 for satisfaction of expectation, measured in the feature space where each feature was normalised to the range [0,1].

To create C_{n+1} the system runs a generate-and-test loop, producing 20 variations of C_n. Variations were generated by flipping each bit in the rhythmic representation with a probability of 0.1. Each variation is combined with the previous human pattern H_n to produce a candidate composite rhythm R'_n, with features F'_n. The pattern with the nearest features to the target is chosen as C_{n+1}.

The human performer is invited to negotiate with the system in a comparable way. As each loop occurs, the performer contributes H_n to the total pattern R_n and assesses the machine contribution. He/she may introduce a modification to the next contribution H_{n+1}. Any modification might be experimental and pseudo-random. Alternatively, it may constitute an intentional action based upon his/her internal description of how the two contributions are co-dependent, and so contribute to developing a better understanding of the expectation, the target point T_n.

This implementation could in theory allow the performer and machine to quickly settle upon a rhythm that satisfies the target T_n, so any further changes would need to be entirely elective; this might be likened to a state of 'optimal flow'. To avoid this settled state, and ensure a continuing creative negotiation, we introduce an additional device that ensures the open-ended nature of the negotiation. If the distance between F_n and T_n is under the threshold (s) – i.e. if reality is sufficiently close to the expectation – the system introduces random variation to its expectation to produce T_{n+1}, with mutation along each feature dimension drawn from a Gaussian distribution. In other words, as the contributing rhythms approach the target, and as the human performer has developed a relatively accurate set of descriptions about the system (based only on the musical surface), the expectation changes. As the feature description F_n is obtained from the composite of both rhythms, both the machine and human performers have a potential role in initiating this change of expectation, whether deliberately or unwittingly, requiring a change of descriptions and resultant actions. A continually diverging system is possible, fostering a mutual creative negotiation that avoids 'optimal flow'.

4 Evaluation in performance

The system awaits testing with a number of human collaborators, to build upon informal, proof-of-concept tests undertaken by the designers. It is evident that with care, a performer can induce a stable and sustainable behaviour in the computer. For the human performer, there are a number of common sense actions to be attempted. For any rhythmic cycle R_n, possible actions include:

1. varying: an intuitive rhythmic variation of a previous pattern.
2. matching: attempting to follow the C_n homo-rhythmically.
3. repeating: so $H_{n+1} = H_n$
4. complementing: an attempt to insert events or gaps that mirror the system pattern in C_n or remembered from C_{n-1}
5. parsing: rhythmic patterns c, where c is a substring of C, that are either complementary or matching parts of R_{n-1}

For each of these possible actions, outcomes are heard in the next outputted rhythm. So the performer attempts to form a description of the system based on how this new pattern deviates from the last. Upsetting the system with a marked variation of pattern (action 1) can cause unstable changes in the output. In this event the system's attempt to update its behaviour in any progressive way is frustrated. Consequently, the performer struggles to verify logical interaction. Clapping the exact same pattern repeatedly (including not clapping at all) and matching patterns (action 2 or 3) can cause the system to slowly evolve its output towards the expectation, allowing a more accurate description of the system to develop, ultimately initiating a change in expectation (feature target point). However, since the features specified are not independent, it cannot be guaranteed that a given expectation can actually be achieved in the musical foreground. In this case the system is stuck, and, to move on, requires a more radical intervention from the performer.

These scenarios are thought experiments as much as real tests. We intend to look at how different performers go about negotiating with the kinds of simple 'black-boxes' under different scenarios, and how strategies to deal with the computer's behaviour are developed. Further development of the system could involve additional, or more effective, feature extraction, to provide a richer feature space. Use of alternative methods, such as autocorrelation, would allow an open definition of the length of any given loop, currently defined as 16 units, and the system could be expanded to involve expressive timing and other more complex rhythmic features. Future systems could offer a more sophisticated and rich musical environment that incorporates many other elements of musical organisation beyond metrical rhythm. In all cases – however frustrating or rewarding – these procedures manipulate the expectations and actions of human and machine performers alike. An unresolvable negotiation is fostered. This points to more complex, and possibly creatively valid, negotiation processes that could produce real-time, computational performances of manifest interest and integrity.

5 Conclusion

We have outlined a minimal musical context for investigating computational creativity in improvised performance. We have offered a framework for interaction between human and machine that comprises four categories: shadowing, mirroring, coupling and negotiating. We adopt a critical approach to the notion of 'optimal flow' in creative interaction. We have developed a test system that explores a process of negotiation in practice, which uses an adaptive system with hidden expectations for varying rhythmic cycles. This creates a demanding context for interactive negotiation. This simple study suggests that the negotiation paradigm could be used to test the dimensions of musical interaction in greater detail (this includes comparing human-human, human-computer and computer-computer interactions using the same paradigm), and could be built up from this minimal form to a critical level of complexity where meaningful and verifiable interaction does occur.

Acknowledgements

This work was developed at the July 2009 Dagstuhl Seminar, 'Computational Creativity: An Interdisciplinary Approach'. We would like to thank the organisers, Margaret Boden, Jon McCormak and Mark d'Inverno, and the other member of our 'interactivity' discussion group: Iris Asaf, Rodney Berry, Daniel Jones, Francois Pachet and Benjamin Porter.

Oliver Bown's contribution to this research was funded by the Australian Research Council under Discovery Project grant DP0877320.

References

1. Sawyer, R.K.: Group creativity: Music, Theater, Collaboration. Lawrence (2003)
2. Wiggins, G.A.: Searching for computational creativity. New Generation Computing **24**(3) (2006) 209–222
3. Colton, S.: Creativity versus the perception of creativity in computational systems. In: Papers of the AAAI Spring Symposium on Creative Systems. (2008)
4. Schwarz, D., Beller, G., Verbrug, B., Britton, S.: Real-time corpus-based concatenative synthesis with catart. In: Proceedings of 9th International Conference on Digital Audio Effects. (2006)
5. Casey, M.: Soundspotting: A new kind of process? In Dean, R., ed.: The Oxford Handbook of Computer Music. Oxford University Press (2009)
6. Pachet, F.: Beyond the cybernetic jam fantasy: The continuator. IEEE Computer Graphics and Applications **24**(1) (2004) 31–35
7. Croft, J.: Theses on liveness. Organised Sound **12**(1) (2007) 59–66
8. Eigenfeldt, A.: Emergent rhythms through multi-agency in max/msp. In Kronland-Martinet, R., Ystad, S., Jensen, K., eds.: Proceedings of the 4th International Computer Music Modeling and Retrieval Symposium, CMMR 2007. (2008) 368–379
9. Blackwell, T., Young, M.: Self-organised music. Organised Sound **9**(2) (2004) 137–150
10. Hamman, M.: From symbol to semiotic: Representation, signification, and the composition of music interaction. Journal of New Music Research **28**(2) (1999) 90–104
11. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row (1990)
12. Ghani, J.A., Deshpande, S.P.: Task characteristics and the experience of optimal flow in human-computer interaction. Journal of Psychology **128**(4) (1994) 381–391
13. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention. Harper Collins, New York (1996)
14. Boden, M.: The Creative Mind. George Weidenfeld and Nicholson Ltd (1990)
15. Abra, J.: Skinner on creativity: A critical commentary. Leonardo **21**(4) (1988) 407–412