Artificial Intelligence 175 (2011) 1697–1721

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Towards a model of musical interaction and communication

Dave Murray-Rust *, Alan Smaill

Centre for Intelligent Systems and their Applications, School of Informatics, University of Edinburgh, Crichton Street, Edinburgh, EH8 9AB, United Kingdom

ARTICLE INFO

Article history:
Received 4 December 2008
Received in revised form 5 January 2011
Accepted 5 January 2011
Available online 11 January 2011

Keywords:
Multi-agent systems
Musical agents
Communicative acts
Human computer interaction

ABSTRACT

In this paper we argue there is a need for a formalised description of musical interaction, which allows reasoning about the musical decisions of human and computational players. To this end, we define a simple model of musical transmission which is amenable to distribution among several musical agents. On top of this, we construct a model of musical perception, based on analysis functions from the musical surface to values on lattices. These values are then used to construct a musical context, allowing for a music-oriented version of concepts such as common ground. This context allows for the interpretation of individual musical output as a stream of discrete actions, with the possibility of constructing sets of performative actions, analogous to those used in Speech Act Theory. This allows musical agent systems to construct output in terms of a communicative dialogue, and should enable more responsive, intelligent participation from these virtual musicians. Finally, we discuss a prototype system which implements these concepts in order to perform piano duets with human performers, and discuss how this theory can be seen as a better defined extension of previous theories.

1. Introduction

Musical interaction deals with the various relations between agents who are producing music. Often, music is conceived of as sequences of musical events (or notes) in time, and analysis focuses on the relations holding between parts of this "musical surface".[1] At this level, there are the traditional musicological questions about features of the music, for example deriving the key of a passage from the notes in it or extracting the underlying metrical structure from a series of note onsets.

Another level of analysis considers the relation between agents and the notes they are producing or hearing. This includes traditionally includes psychomusical questions such as why certain sets of notes are pleasing, and music performance questions such as what a particular rendering of a set of notes can tell us about the higher level structures a performer is using, or whether we can relate performances to particular agents and identify a style.

The questions addressed by this work are about the way in which the playing of one musical agents is related to that of another — to focus on the communicative aspects of the musical experience rather than those which can be directly derived from the musical surface or understood by analysing the relationship between a defined musical structure and a listener.

This is achieved by casting music in terms of intentional, communicative actions; firstly, we must define what is meant by an action in the context of playing music, and secondly we must provide a manner to relate these actions together. When a group of musicians play together, they will each output a stream of notes and silences. Some of this output is of particular

* Corresponding author.
E-mail address: D.Murray-Rust@ed.ac.uk (D. Murray-Rust).

[1] This is a term from Lerdahl and Jackendoff [1], and indicates the "lowest level of representation which has musical significance" [2, p. 219], and is used here in the sense of "the level of music representation of most use to the task at hand".

0004-3702/$ – see front matter
doi:10.1016/j.artint.2011.01.002

interest, and can be said to have intention behind it — producing new musical ideas for others to take up, creating shared rhythmic and harmonic structures, marking structurally salient points in time and so on. This paper gives a model which can serve as the basis for recognising these actions and understanding their interactive function.

1.1. Motivation

The notion of communicative actions can be traced back to Searle and Austin [3,4], and was originally used to describe the manner in which words could be used to change the world. The use of communicative acts in multi-agent systems is widespread — see for example [5] or [6] — since defining formal semantics for communicative actions allows the modelling of patterns of interaction as protocols based on the exchange of communicative actions, and allows agents to reason about the beliefs and intentions of other agents (e.g. [7]). Here we take a top-down approach to exploring what an analogous formulation for musical interactions could look like.

The issue which prompted the creation of Speech Act Theory — a reaction to the Positivist position that everything of value in language could be represented as true or false logical statements, and hence the need to talk about communication rather than truth — can be seen even more strongly when studying music, where there is no commonly accepted notion of truth or semantic grounding.

The analogy between communicative actions and musical performance is suggested by the musical literature which refers to musical improvisation and performance in terms of "conversation" or "discussion" between the musicians [8–10]; in particular Coventry and Blackwell [8] draw heavily on Grice [11] to characterise musical playing in *pragmatic* terms (as opposed to syntactic or semantic).

Murray-Rust et al. [12–14] are previous attempts by the authors to formalise communicative musical acts in the context of a multi-agent system.

Given the usefulness of Speech Act Theory and the notion of Communicative Actions to the AI and Multi-Agent Systems communities, it is reasonable to suppose that applying these ideas to the study of computer music will yield interesting results. However, despite the similarities between traditional Communicative Acts and musical communication, there are some structural differences which indicate that a new model should be created.

Firstly there is the question of semantics in music — there is not the sense of "literal meaning" of musical playing that might correspond to the literal meaning of a natural language statement, and many authors (e.g. Wiggins [15], Sloboda [16]) suggest that music is semiotic rather than semantic. It is possible in certain situations for the people playing music to have quite different ideas about what is happening, but for the interaction to still be considered "successful". Cross [17] discusses this *polysemic* nature of music at length.

Secondly, there is the aspect of temporality; musicians typically play simultaneously with other musicians, rather than alternating discrete actions — however, there is some support for the idea that even though the musicians are *playing* at the same time, they may not all be *communicating* at the same time [10]. The time at which actions happen must be taken into account, not simply the order in which they occur, and any system which designed to interact musically has to respect realtime constraints, which has implications both for performance, and for the richness of its temporal representation.

It is generally necessary to model interaction between more than two agents, which is the case usually treated in dialogue — with some exceptions, e.g. [18,19]. Although some existing agent protocols do this, musical work demands a different kind of flexibility and approach to collaboration.

There is also a range of levels of autonomy given to musicians while they play; for example, the strong constraints of a string quartet, where the correct notes must be played in the correct order stands in contrast to a free jazz group, where there is no such notion. This means that musical improvisation should be seen as a continuum, from the most rigidly structured to the most "free". Benson [20, pp. 26–30] gives a possible categorisation for these differences, e.g.

Improvisation$_1$ "Filling-in" certain details that are not notated in the score, e.g. tempi, timbre, attack, dynamics.
Improvisation$_2$ Addition of notes to the score that the performer is *expected* to perform, such as trills, and filling in figured bass parts.[2]
Improvisation$_5$ Addition or subtraction of complete measures, passages or scores.
Improvisation$_{10}$ The composer uses a particular work as a template, to produce a more complex (or simply different) work.

There are two points to take from this:

- on a musical level, even in the most rigorously notated work, there is scope for individual interpretation, which will occur with reference to the other members of the musical ensemble;
- in terms of building systems which communicate with humans, there is a continuum of situations to address, each of which can have quite precise bounds put on the level of autonomy required or allowed — it is hence a very rich problem set.

[2] *Improvisation*$_{1,2}$ are both common in Baroque scores, forming an expected part of Baroque musical practice.

1.2. Related work

A large amount of work has been performed in the direction of identifying the structures behind music; possibly the best known is the Generative Theory of Tonal Music (GTTM) [1], which provides a listening grammar for music. This comprises a set of four hierarchical descriptors that give progressive refinements of different aspects of the musical surface, which has informed a range of computational systems (e.g. Smoliar [21]). The GTTM gives a very detailed analysis of music, but only from the point of view of listening, and also from the position of *having listened* to an entire piece. Less well known, but more appropriate to the current work is the model proposed by Pressing [22], which deals with the processes by which a single improviser creates music. This works with music as it unfolds in time, and makes use of cognitive models to give an account of human musical performance.

The analysis of the structural and intentional elements of music which form part of a communicative dialogue between intelligent agents has not been addressed in a systematic manner. However, there are several computational systems which deal with the notion of musical agents, with origins in Minsky [23,24]. Ueda and Kon [25] details a framework for "Mobile Musical Agents", but leaves open the question of how the agents structure their playing. The Continuator [26,27] is a system which functions as a partner in musical improvisation, by completing sequences learnt from its human partners; Pachet's work on MusES is also relevant here [28,29], for developing the notion that some parts of musical output can be thought of as actions. Two more recent systems are described in Fonseka [30], which plays certain contemporary music pieces, and Wulfhorst et al. [31], which looks at beat and harmonic tracking for groups of agents.

1.3. Structure

The rest of this paper will develop the theory of Musical Acts, by:

- identifying the characteristics required of communicative acts in music;
- defining a mechanism for agents to extract features from the music they hear;
- using this feature extraction to allow reasoning about the musical beliefs of other agents;
- constructing musical actions based on this form of analysis and reasoning;
- demonstrating that this model can be implemented in a real computational system;
- showing that this model captures both an intuitive notion of musical action, and can be used to construct and extend similar previous formulations.

2. Musical acts

Given these differences and points of interest, we can start to frame what Musical Acts are, and what they are useful for. The rationale for developing this theory is to create a narrative "plan-view" of a musical interaction, which sheds light on the actions musicians take and the reasons for those actions.

Much like Communicative Actions, qualities which these acts must have are:

Embodiment through the production of music; much as speech acts are embodied through the production of utterances within certain contexts, musical acts must have a manifestation in music.

Intention is what differentiates a musical act from general musical playing. A musical act should have perlocutionary force — it should be an attempt to change the state of the world or the actions of others by its production, in a more significant sense than the mere fact that it has been produced.

Intelligibility is necessary for a successful act; if it is not understood, then it will fail to change the world, as other musicians will fail to react to it. So, the act must be conceptualised within the context of a certain musical situation, with a certain expectation of understanding from the other musicians.

So given these requirements, some questions are raised about the characteristics of Musical Acts, and how they may be determined:

- What are the effects of Musical Acts?
- How can Musical Acts be detected?
- How is the intention behind an act determined?
- Can acts occur in the context of scored music, or are they only applicable to improvised music?

To address some of these, we will first focus on the kinds of actions to be discussed. When human musicians play together, there are many things which take place; glances are exchanged, feet are tapped, music is played, particular phrases are played at particular times and so on. For the purposes of this analysis, we will first decide to ignore extramusical actions — those which occur outside of the musical surface; this includes nods, glances, the foot tapping and body movement used for synchronisation, hand signals for jazz chords, the conductor's baton and every other action which is not contained by the musical surface. This relates to the notion of *embodiment* from above — by working only with the musical surface, it is impossible to analyse actions which are not contained within the surface.

Secondly, we make a distinction between *conventionalised* and *free* musical actions:

Conventionalised Musical Actions are uses of particular musical phrases with particular conventionalised meanings in certain styles of music at certain points. For example, the "change rhythm" used in many African drumming traditions, which signals that the next section of the piece should begin, and would be known and understood by all the players.

Free Musical Actions are parts of the musical surface which can be seen as actions by their relation to the musical context surrounding them. Their *meaning* or import can only be inferred as part of a complete musical interaction. For instance, these are the moments when the drummer introduces a new rhythm, and the bassist starts filling in the gaps; when a new melody gets passed around all the members of an improvising ensemble; when the backing section adopt ideas from a soloist's playing.

In some cases, the line between musical actions which are understood through convention and those which are understood through some more general musical understanding may be poorly defined. The desired distinction is that conventionalised actions are defined by particular configurations of notes — specific forms of playing which can, in general, be made *explicit*. We ignore these, and instead concentrate on free musical actions are understood in the context of a particular musical interaction, and deal with the relationships between the playing of the players rather than any specific phrases or material.

2.1. Musical acts: an intuitive formalisation

When musicians play together, there is generally some set of musical structures which all of the musicians would agree on. For example tempo, chord structure, song structure and so on. In different musical styles, this shared understanding may take different forms: in an orchestra it could be a printed score along with a set of rehearsed directions about how to play it, in a free jazz ensemble, it may be a diffuse, implicit knowledge of what is *currently* happening on several different levels. This shared representation comes from two sources — pre-existing structures, such as scores or musical traditions, and extemporaneously created material which arises as part of the interaction between the musicians.

During the course of the musical interaction, musical output can be roughly divided into two sorts: that which is expected according to the shared representation that the agents have, and that which is unexpected or novel. This distinction needs more analysis, as it may be that only certain aspects of the music are predicted, while others are novel — for instance, playing the expected melody with a different rhythm, or unexpected expressive character, playing the expected rhythm, but accenting different notes. Certain features of the music played can be seen as novel, and hence can be seen as a vehicle for musical change. This means that all of the repetitive and commonplace aspects of the music can be ignored — the drummer keeping time, the fact that the string quartet are playing the right notes — in favour of analysing the novel occurrences, such as the phrasing structure the drummer uses, and the way that the expressive aspects of the violin's playing change in response to the cello.

Playing something which differs from the shared representation is necessary in order for an action to be *intentional*, and act as an attempt to alter the shared representation of music which the musicians have.

2.2. An example: Little Blue Frog

"Little Blue Frog" [32], is used as an example for what a Musical Act analysis of a piece of music looks like. The analysis is carried out from a personal listening perspective — that is, without recourse to transcriptions or other theoretical explanations, using only my individual musical competences, and to offer a view of what the first author personally found to be the most interesting parts of the interaction. This is fully in keeping with the philosophical thrust of Musical Act Theory — that meanings are extracted and used by individuals, who have their own capabilities, idiosyncrasies and deficiencies, and any model must take this into account. The full analysis, along with definitions of the performatives used is given in Appendix B.

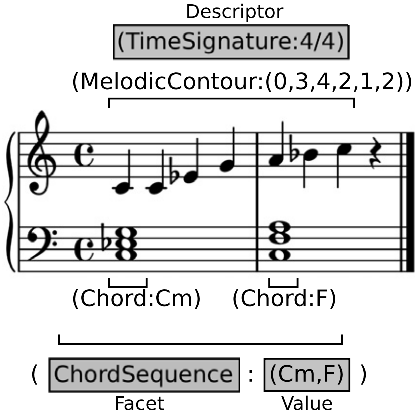
This is an intuitive overview of what musical acts should look like, and provides a motivation for the theoretical work carried out in the next section. In Section 6.3 we will give more formal definitions in terms of the logic presented in the rest of this paper.

3. Describing music

The first part of this model is concerned with the ways in which higher level structures may be derived from the flow of music, as this is felt to be a prerequisite for understanding action and intention, in much the same way as concepts and structure can be derived from natural language sentences.

For the purposes of this model we assume a musical surface which has been segmented into discrete events such as notes, percussive strikes, glissandi, trills, etc. We are interested in building up a model for expressing higher level musical

This figure shows a short passage of music, along with some potential descriptors attached to it. Rectangular brackets denote the source of the descriptors where appropriate.



Facet An aspect of music to be analysed. Examples: TimeSignature, Chord, Chord Sequence, Melodic Contour

Value A value for a facet, e.g. for Time Signature 4/4, 3/4, 6/8 . . .

Descriptor A mapping from facet to value, i.e. (TimeSignature: 4/4)

Analyser A procedure for generating descriptors from the musical surface

Fig. 1. Facets, values, analysers and descriptors.

structures, which can serve as the building blocks for an understanding of the communicative processes occurring while music is being made.

These constructs must allow for the following properties of musical descriptions:

- there will be several descriptions which could be applied to the same piece of music — for example a single note may well have rhythmic, harmonic and melodic functions; this polysemic nature is one of the properties of music which motivates the current work;
- the set of descriptions used will vary with the type of music and the skills and interests of the performer and listener, so not only may different agents have different capabilities, the range of potential capabilities cannot be known ahead of time.

It is also necessary to be able to calculate these descriptions and that they should be susceptible to automated reasoning methods in order to construct appropriate output.

In order to capture the notion of multiple structures being attached to a piece of music, the notion of *facets* will be introduced; these are descriptive classifications for certain aspects of the musical surface. Any particular piece of music can be analysed simultaneously along many different facets, depending on the capabilities of the agent (or person) performing the analysis, the style of the music and the results required.

An analysis procedure carries out the task of attaching a specific *value* to a *facet*; this mapping will be termed a *descriptor*, and the process an *analyser* (see Fig. 1).

An informal notation will be used here to represent these *descriptors* as facet–value pairs; as an example, a very simple *descriptor* would be (Time-Signature: 4/4), which delimits the set of possible musical excerpts where the signature is 4/4.

Before progressing further, it should be stated that:

- descriptors are essentially perceptual objects; a given set of descriptors only accounts for a *possible* analysis of the music; they are subjective and not necessarily complete;
- in the coming examples, the descriptors used may not seem to capture the essence of musical knowledge about the subject; this should be taken as a property of those particular descriptors, rather than an issue with the theory — simplistic descriptions of music have been purposefully used to make the logical structure as clear as possible;
- facets will be considered independently, although in general there may be some relation between the values assigned to certain facets.

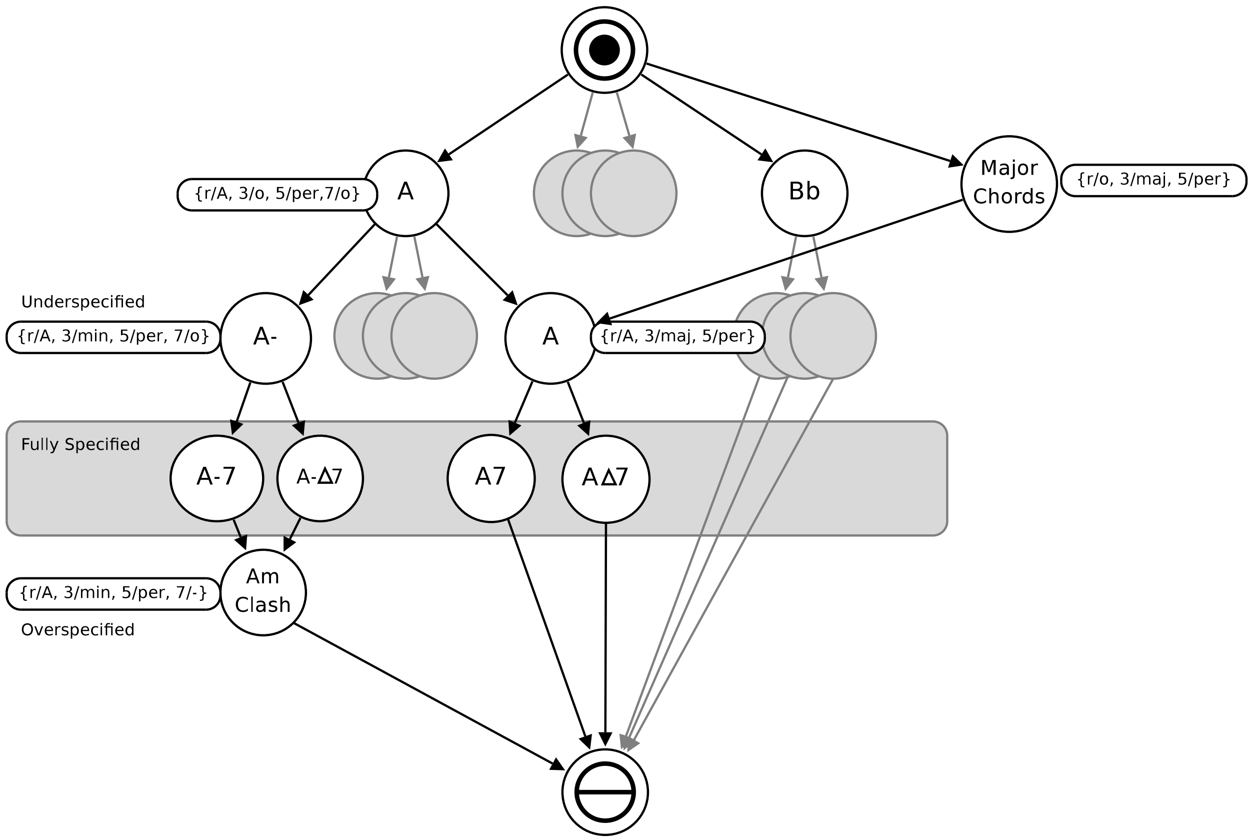


Fig. 2. Value system for simple jazz chords. The chords are composed of a root note, a third, which may be major or minor, with minor chords having a '–' after the root note, a fifth, and a seventh which may be major or dominant; major sevenths are notated with a '7', while dominant sevenths are notated '7Δ'.

3.1. Values and their relations

In order for this system of description to be of use, the form of values should be constrained by the minimum amount necessary to allow the kinds of reasoning we are going to perform — there are many ways in which music is analysed, and the easier it is to implement and use these analyses within the model, the more useful it will be.

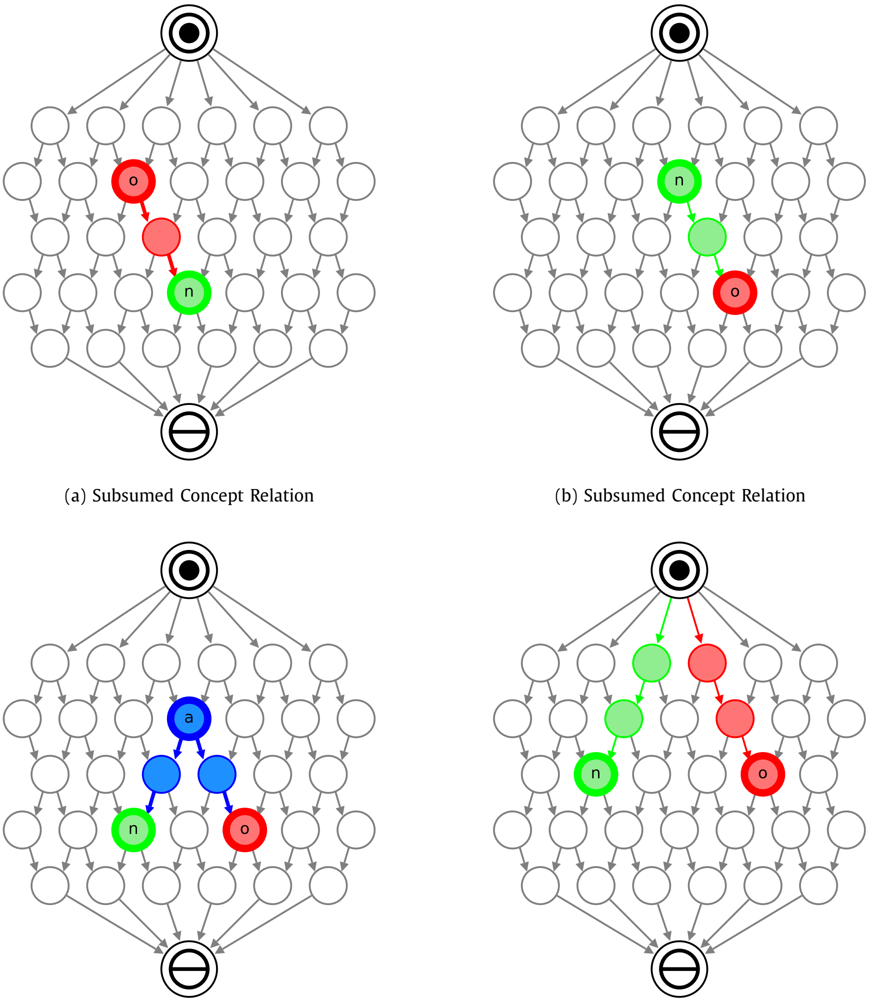
A value for a facet should be thought of in set-theoretic terms; if we take the set of all possible pieces of music, a value for a facet defines a certain subset of these possible pieces of music.

In order to allow for a rich range of reasoning about the relations between these values, we add the restriction that these subsets are ordered, so that they form a lattice — a similar formulation to Concept Lattices (e.g., [33]). This is designed to capture the notion that many descriptions of music have various levels of specialisation and interrelation — see for example Fig. 2, which gives a sketch of how different musical ideas (the root of the chord, major/minor chords, "extensions") can be used simultaneously to define different chords. The ordering restriction means that for any two values, it is possible to calculate their *join* — analogous to the union of the two subsets — and their *meet* — analogous to the intersection between the two subsets.

These lattices can be represented as directed graphs, where each node is a particular value, with specialisation increasing as one descends the graph. For instance, in Fig. 2, the second row contains concepts which specify a single part of the chord (either the root node or the relation between the root and the third), the next row contains values which specify two parts of the chord, etc. In this form, the *join* of two nodes is their first common ancestor, and the *meet* is their first common descendant.

In addition to this, two special symbols are used:

- $\odot$ is the top element of the lattice, and is used to signify that no value is specified for that facet. This corresponds to the set of all possible musical fragments;
- $\ominus$ is the bottom of the lattice, and is encountered when values are combined such that no pieces of music can satisfy the resulting condition — in the jazz chord example (Fig. 2, it is impossible to specify more than 4 parts of the chord, so values in the "Fully Specified" row are connected downwards to $\ominus$.



(a) Subsumed Concept Relation (b) Subsumed Concept Relation

(c) Altered Concept Relation (a is shared ancestor) (d) Disjoint Concept Relation

Fig. 3. Illustration of different concept relations between two values o and n.

Using only the *join* relationship given by the lattice structure, it is possible to create a variety of relations which can be extracted from any pair of values. These are designed to capture the important parts of the relations between these values, without being specific about exactly what the values are.

In order to talk about what is happening musically, we would like to be able to take two values for a single facet and compute a relation between them — for instance to compare the chords on agent is playing with those another is playing, or to compare the rhythm an agent is playing with the one it was playing previously.

Looking at two values n and o, the relations we define between them are as follows (see Fig. 3):

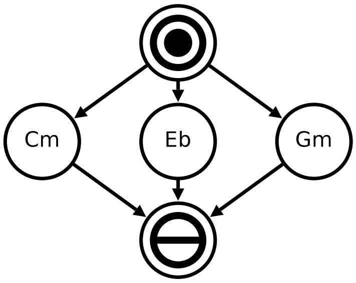
Same $n = o$.

Subsumed indicates that n is a specialisation of o — that n is subsumed by o, or that $n \subset o$. On the graph n is a descendant of o, and describes a smaller set of musical values; alternatively, the *join* of o and n is o. Looking at Fig. 2, we could say that "A7 is subsumed by A".

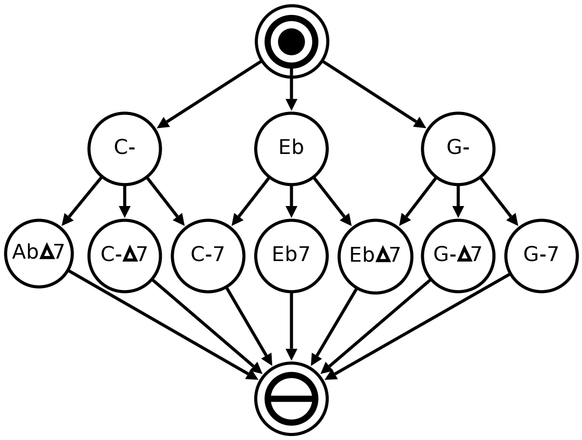
Subsumes indicates that n is less specialised than o, or that $o \subset n$. In graph terms, this means the new value is an ancestor of the old; in lattice terms, the *join* of o and n is n. Referring again to Fig. 2, we could say that "Major chords subsumes A major".

Alter indicates that n is an alteration of o — they have some elements in common. In graph terms they share a common ancestor a, or in lattice terms their *join* is non-empty, i.e. in Fig. 2 "A- is an alteration of A".

Disjoint indicates there is no commonality between n and v. Their only common ancestor is $\odot$ (which is an ancestor of every node), or their *join* is an empty set. In Fig. 2, "A is disjoint from B$\flat$".



(a) Nursery Rhyme chord analysis



(b) Jazz chord analysis

Fig. 4. Two different lattices for analysing chords showing chords related to C minor, E♭ major and G minor (see Fig. 2 for an explanation of notation).

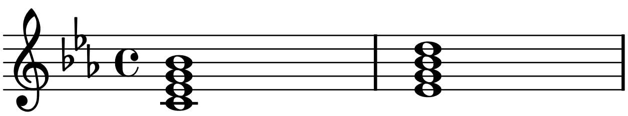


Fig. 5. Example chords for analysis.

This set of relations will be labelled *Rel*:

$$Rel \underset{def}{=} \{\texttt{SAME}, \texttt{SUBSUMED}, \texttt{SUBSUMES}, \texttt{ALTER}, \texttt{DISJOINT}\}$$

3.2. Analysis procedures

An analysis procedure A_f for a particular facet f can be defined as a function which produces *descriptors* — facet–value tuples. That is, when given a fragment of music M as input, an analyser for facet f maps it into the lattice f (notated L_f):

$$A_f : M \to L_f$$

It is generally assumed that an agent will have a set of analysers which it can apply to new music; the set of facets over which these analysers act is notated F. As noted before, there is no universal set of analysers, and different agents may have different analysers for the same facets. In order to make this clear, two chord analysers are detailed below. The lattices used by these analysers are shown in Fig. 4.

The first analyser — $A_{nursery}$ — is a "Nursery Rhyme" treatment of chords, where the lattice simply comprises a node for each major and minor triad (Fig. 4(a)). A_{jazz} takes a richer, jazz oriented approach. Here again, we start with a node for each triad, but this is now augmented with nodes for all the major and minor seventh variations of these chords, and extra relations are created between the nodes (Fig. 4(b)).

Now consider the reaction of these two analysers to the piece of music shown in Fig. 5, which contains two chords: C, E♭, G, B♭ followed by E♭, G, B♭, D. The analysers are asked to calculate the values for each of the chords, and the relation between them.

$A_{nursery}$ cannot fully represent the C minor 7 (C-7) chord, given that its universe consists only of three note triads. It has two possibilities — Cm or E♭ — and would need some way to determine which of these was more appropriate. There is a third possibility, which is that it would refuse to classify the chord, but it will be assumed that it attempts to locate every input *somewhere* on its lattice, even if the fit is not exact. It is possible that a well designed analyser would choose Cm in this case. The following chord is analysed in a similar manner as E♭. When computing the relation between them, it can simply note that they are different nodes, and say that they are disjoint.

On the other hand, A_{jazz} can classify both of these chords exactly on its lattice, as C-7 and E♭7 respectively. Furthermore, when computing a relation between them, it can determine that E♭ is a parent to both chords — since adding a C to an E♭ chord gives C-7 and adding a D to an E♭ chord gives E♭7 — and hence that moving from C-7 to E♭7 is an alteration.

It is open to debate whether this ability to capture extra relationships between values and give richer descriptions of music makes A_{jazz} a better analyser, and the response is that it is up to a particular agent to decide which analysers are most useful for the style of music which it is playing — whether C-7 is seen as an alteration of E♭7 is dependent on the theory within which one is working, and different approaches will be appropriate for different musics. This is important to the generality of this model, as it means that any musical theory can be used so long as it can be represented as a lattice of values and an analysis function from the musical surface to these values.

4. Playing with other musicians

This section expands the model to take into account the different agents involved in an interaction, and what they can deduce about each other's musical beliefs.

We develop two concepts related to this: the musical context and musical common ground.

4.1. Listening to others

As previously mentioned, agents are likely to have a range of analysers to apply to any music which they hear — each analyser will produce its own descriptor, e.g. (Tonality: C minor), (TimeSignature: 4/4), etc. The symbol $\Rightarrow_A$ is used to mean "A can extract values for the facets…", so if A has an analysis procedure for facets $F = \{f_1, \ldots, f_n\}$:

$$m \Rightarrow_A \big\{(f_0, v_{0m}), \ldots, (f_n, v_{nm})\big\}$$

In other words, for each facet $f_i \in F$, A can extract a value v_{im}.

The formulation is that when agent B hears agent A play music, B believes that A has expressed[3] all the properties that B can derive from the music[4]:

$$Heard_B\big(Played_A(m)\big) \wedge (m \Rightarrow_B D) \rightarrow Bel_B\big(Expressed_A(D)\big)$$

where $D = \{(f_0, v_{0m}), \ldots, (f_n, v_{nm})\}$ for the set of descriptors which B can extract from m. It should be noted that this set of descriptions may be entirely different from whatever representation (or process) A used to create the music, and that A and B might have entirely different sets of analysers.

4.2. Current values and musical context

The concept of musical context, or the "musical now", is intended to capture the idea of "what all the musicians are doing at the moment". The extraction of values discussed above is outside of any temporal framework; it is simply the method of determining what features are embodied in a fragment of music. Since music happens within its particular temporality, this extraction process must similarly be embedded in a temporal structure.

In order to hide some of the temporal complexity from our analysis, a state based description of the musical interaction is introduced. We continue the assumption that everything of interest in the interaction is captured in the output of the agent's analysers (this is discussed more fully in Section 5). We can then talk about the interaction as a series of states, so that any change of a value leads to a change of state.

Each value for each facet in the current state is labelled the current value, so $CV_a(f, v_f)$ states that the current value of a's playing when analysed on facet f is v_f.

More formally, for an agent x in a group of agents A, with a set F of facets it can analyse, the context is defined as the conjunction of all relevant facet–value statements:

$$Context_x = \bigwedge_{a \in A} \bigwedge_{f \in F} CV_a(f, v_f)$$

[3] *Expressed* is used in a weak sense here, to mean that those properties could be perceived in the music — it does not necessarily mean that they were *intended*.

[4] $Heard_B(Played_A(m))$ is used to indicate that B has heard A play the musical excerpt m, and $Bel_B(\phi)$ is used to indicate that B believes the proposition ϕ.

4.3. Common knowledge and musical common ground

These formulations work for a single agent analysing the output of its peers; in order to reason about actions to take, it is necessary to discuss the beliefs which the agent can reasonably hold about what the other agents are doing. Here we draw on the notions of common knowledge and common ground. For common knowledge, notated $CK(\phi)$:

> for a proposition ϕ, it is common knowledge that ϕ among a group of agents iff all know that ϕ, all know that all know that ϕ, all know that all know that all know that ϕ, etc.

This is a property of the system as a whole, which can prove problematic; hence the use of "common ground" — the more pragmatic notion of that which a speaker presupposes to be common knowledge [34], which is then a property of an individual.

In terms of the musical agent system, using the assumption that every agent can hear the output of every other, and that this fact is common knowledge, it is clear that the playing of any agent is common knowledge:

$$Played_A(m) \rightarrow CK\big(Played_A(m)\big)$$

as is the fact that this has been heard by every other agent (for all $B \neq A$):

$$Played(A, m) \rightarrow CK\big(Heard_B\big(Played_A(m)\big)\big)$$

It is possible to imagine situations in which it is not the case that all of the musicians hear each other, but this covers a wide range of "normal" musical situations, and is in most cases a state which is desirable even if not attained.

If an agent believes that another agent can analyse a certain facet of music, and will produce the same values for that facet, it will believe that this agent will derive the same values from the music it hears:

$$\begin{aligned}
&Bel_A\big(CanAnalyse_B(f)\big)\\
&\quad\wedge Heard_A\big(Played_C(m)\big)\\
&\quad\wedge Bel_A\big(CV_C(f, v_f)\big)\\
&\quad\rightarrow Bel_A\big(Bel_B\big(CV_C(f, v_f)\big)\big)
\end{aligned}$$

Hence, given a set F_s of facets which A believes B will analyse in the same way:

$$Bel_A\big(CV_c(f, v_f)\big) \rightarrow Bel_A\big(Bel_B\big(CV_c(f, v_f)\big)\big) \quad (\text{for } f \in F_s)$$

Now, suppose that A believes there is a set F_c of facets which every agent can analyse, that all agents will arrive at the same values for these facets, and also that these facts are common knowledge. It follows then that A believes the values which can be deduced for these facets are common knowledge:

$$Bel_A\big(CV_x(f, v_f)\big) \rightarrow CG_A\big(CV_x(f, v_f)\big) \quad (\text{for } f \in F_s,\ x \in Agents)$$

We define this as *musical common ground* — the set of values which an agent reasonably believes to have been extracted by every other agent, and hence to be common knowledge.

This expectation of understanding is supported by the fact that much of music education, particularly music theory, is concerned with creating a common vocabulary for musical occurrences; in general, within any particular style, there are aspects of music which it would be taken as read that any skilled practitioner would understand. For example, it would be assumed when playing jazz that people will agree on the time signature a piece is in, or what the chord structure is.

There are many situations where these assumptions turn out to be false: people may disagree about what chords are being used, about whether a piece is in 6/8 or 3/4 time, etc., but they are reasonable assumptions to make within a given cultural context. Similarly, when musicians with different backgrounds get together, the shared set of features may be smaller; for these reasons, the set of features which any agent shares with another may dynamically change, both within a single interaction and over the course of many interactions as expectations are adjusted, new capabilities are discovered and deficiencies are exposed.

It is also sometimes useful to think of common ground among a subset of the players: the beginning bassist may be happy to know the root of each chord being played, while the pianist and guitarist share an understanding of the extensions and passing notes they are using, and the drummer might have little knowledge of the actual chords used — see Appendix C for a worked example of this.

5. Actions within musical context

The focus of this work is to develop a framework within which to analyse musical action; up to this point, we have developed a model of musical description and belief structures which allows the discussion of what is happening at any given time, and what the agents involved can deduce from this. The transition to looking at *actions* is made by casting the musical interaction as a series of states (defined by their musical descriptors), and analysing the transitions between the states. In order to do this, an assumption is made:

$A_{context}$: everything which is of interest to an agent about the musical interaction is contained in the *context* which it maintains.

In other words, the agent is only interested in those aspects of music for which it has analysers which can produce descriptors — everything else is assumed to be unintelligible. This relates to the requirements in Section 2, that actions must be *intelligible* (and also *embodied* in the input to the agent).

Given this, in order for anything to constitute a free musical action from the point of view of an agent hearing it, it must involve a change to that agent's context, and hence must involve a new state in the state description:

A_{action}: all free musical actions consist of a change of state.

Turning this around, the converse assumption is that:

A_{state}: every change of state constitutes a musical action on the part of the agent whose value changed the state.

Both of these assumptions are discussed further in Section 5.4, but accepting them leads to the conclusion that analysing the transitions between states gives a picture of the intentional actions present in the interaction. Since we define the states in terms of values for which certain relationships (from *Rel*) may be calculated, we can start to use these relations to characterise the changes.

5.1. Actions and relations

In order to extract actions from the musical surface, new playing is related to the context in which it occurred — in other words, it is compared to the current values of the whole group at the time the action happens. Every time a new value is expressed by an agent, a new state is created. The action is then a combination of:

- the musical surface so far, including the part which constitutes the action;
- the state description of the musical interaction;
- previous actions;
- the agents involved in the interaction;
- the new value which has just been played.

This refers to a specific action, by a specific agent at a certain point in time in a specific interaction. Since this is such a specific event, it is not easily generalisable to other situations. Instead, a *musical action signature* is constructed, which consists of:

- the relations between the new value being played and the current musical context;
- relevant relations between values in the current musical context.

5.2. Action signatures for two agents

In order to simplify this, consider the case of two agents playing together; agent A has introduced a new value for a particular facet, so the values which are available to characterise the act are:

- A's new value (a_{new}),
- A's old value (a_{old}),
- B's old value (b_{old}).

Since relationships are directional, there are six possible relationships to consider. However, since special significance is given to the new value played by agent a, relations are considered between (see Fig. 6 for details):

- a_{new} and a_{old},
- a_{new} and b,
- a_{old} and b.

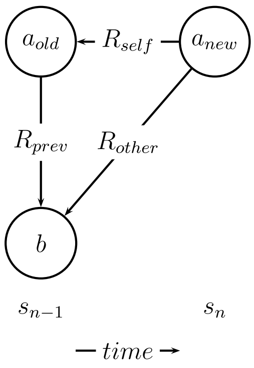


Fig. 6. Relations between the values of two agents *a* and *b*, from the point of view of *a* constructing a new value. R_{self} is the relation between *a*'s previous and new playing, R_{other} is the relation between *a*'s new playing and *b*'s previous playing, and R_{prev} is the relation which used to hold between *a* and *b*'s playing.

Put together, this means an action signature is defined as a triplet of relations — from *Rel* — for values of a certain facet.

$$ActionSignature \underset{def}{=} (R_{self}, R_{other}, R_{prev}) \quad \text{where } R_{self}, R_{other}, R_{prev} \in Rel.$$

In the rest of the text, it will often be important to know which agent produced an action signature. In this case, it will be written *AgentName*: $(R_{self}, R_{other}, R_{prev})$, e.g. A: $(\texttt{SAME}, \texttt{ALTER}, \texttt{DISJOINT})$.

This does not entirely define the action — it characterises the relational aspects of it independent of the context it occurred in. It does not include the musical surface, or even the part of it deemed to constitute the action, but is designed as a compact, transferable representation of the components of the action which are useful from the point of view of analysing interactions. The interaction can now be described using a series of these action signatures, as will be demonstrated in the following paragraphs. Finally, due to structural constraints, not all combinations of relations are possible — see Appendix A for a discussion of this.

5.3. A worked example

At this point a worked example is useful, to demonstrate these concepts in action. This involves two agents, *A* and *B*, who are both jazz musicians, and looks at a chordal analysis of their output. The analysis is going to be conducted from the point of view of *A*, who is using the simple jazz chord analyser (A_{jazz}) from Section 3.2 to extract chords and their relationships.

At the start of the example, *A* is playing a C chord (i.e. the notes C, E, G), while *B* is playing C7 (C, E, G, B♭), so *A*'s value SUBSUMES *B*'s. At some point, *B* decides to start playing $C\Delta 7$ instead (C, E, G, B). This allows the calculation of three relations:

- R_{self} is the relation between *B*'s new and old values. Even though *A* is conducting the analysis, when analysing the actions of others, R_{self} describes the relations between the *other*'s values. The relation is from CΔ7 to C7, and is hence ALTER — they have a common ancestor in *C*.
- R_{other} is the relation between *B*'s new playing and *A*'s current playing, so between CΔ7 and C, and is hence SUBSUMED.
- R_{prev} is the relation which held between *B*'s old value and *A*'s current value, i.e. between C7 and C, and is again SUBSUMED.

This means that from *A*'s point of view, *B* has executed a musical action with the signature B: $(\texttt{ALTER}, \texttt{SUBSUMED}, \texttt{SUBSUMED})$. It should be noted that *B* might have been using an entirely different method of analysis and generation — this analysis depends on *A*'s perceptions and capabilities, which may be entirely different to *B*'s. The exchange can be seen in the first two states of Fig. 7 while Table 1 represents a larger excerpt of this interaction in tabular form.

5.4. Reasonableness of $A_{context}$ and A_{state}

There are three assumptions (from Section 5) which are central to this form of analysis, which must be discussed. To recap, they are:

- $A_{context}$: everything which is of interest to an agent about the musical interaction is contained in the context which it maintains.
- A_{action}: all free musical actions consist of a change of state.
- A_{state}: every change of state constitutes a musical action on the part of the agent whose value changed the state.

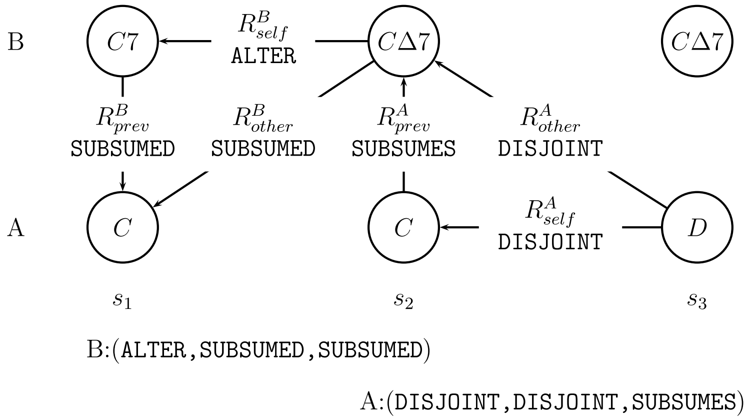


Fig. 7. Example of two action signatures extracted from the playing of agents.

Table 1
Example of representing a musical interaction as a series of actions.

State	A	B	Action signature
s_0	C	⊙	A: (SUBSUMED, SUBSUMED, SAME)
s_1	C	C7	B: (SUBSUMED, SUBSUMED, SUBSUMES)
s_2	C	CΔ7	B: (ALTER, SUBSUMED, SUBSUMED)
s_3	D	CΔ7	A: (DISJOINT, DISJOINT, SUBSUMES)
s_4	D	D	B: (DISJOINT, SAME, DISJOINT)

There is a question as to how reasonable these assumptions are, and whether they hold for all musical situations. Firstly, $A_{context}$; within the bounds of this model, this is a safe assumption — agents are modelled with a set of analysis routines for all of the features that they understand. Given this, it is fair to say that everything the agent is capable of understanding is in the context. In a more general situation, it becomes somewhat tautological: the context *is* everything which the agent is interested in. Carrying on from this, A_{action} seems reasonable, as if the context represents everything an agent knows about the interaction, then anything which does not alter the context is invisible to the agent.

A_{state} needs more defence, however, on several counts:

- an agent may alter its values unintentionally — for instance, it may be attempting to play a constant rhythm, but be unable to keep a steady pulse, or it may play the wrong note in a chord;
- values may change in response to other processes. The major example of this would be if the agents are playing with some form of score, which dictates certain aspects of the musical surface. In this case, the agents would need a way to differentiate between intentional actions, and those which come about from following the referent, which are effectively conventionalised actions, as they are pre-arranged between the group;
- lastly, not all actions are likely to be of equal significance — for instance changing one note in the drum rhythm is likely to be less important that an entirely new harmonic pattern. However, in certain styles of music, the single note change could have a profound effect.

Given these concerns, it appears that some measure of the importance or interest of an action would be useful, so that expected or trivial actions are noted as such, leaving more room to react to the truly novel and unexpected.

6. Extending actions

We have laid out a simple formulation of for Musical Actions, which by design has ignored many aspects of the theory. In this section we discuss:

- how Musical Actions may be used to inform playing;
- how to deal with multiple musicians simultaneously;
- how these Musical Actions relate to the intuitive formulation of Musical Acts given in Section 2.1.

6.1. Choices

This theory of musical actions has a dual purpose: to describe the music played by agents in terms of high level actions, and to give agents a framework for making decisions about what to play.

If an agent A is engaged in a musical interaction, at many points A must make a decision about what to play next. Assuming that A is using musical actions as a basis for playing music, there are two high level components to this decision:

- what musical action signature to use;
- what high level feature to use to embody it.

Firstly, A must decide on an action signature to produce. This can be seen as choosing values for R_{self}, R_{other}, R_{prev}, or action signatures can be seen as discrete entities with some kind of performative *meaning* as discussed later in this paper. The only constraint on this choice is that R_{prev} is already determined — it is the value between A's current playing and B's new playing, in this case SUBSUMES.[5]

Suppose now that A decides to do something radical, and play some music which does not match what is currently happening — that is to say, a new value which is DISJOINT from both A and B's current playing. A would then decide to use an action signature of A: (DISJOINT, DISJOINT, SUBSUMES) for the next piece of output.

Once A has chosen this action signature, it is necessary to find some musical values which embody it. The choice now is out of all the values which satisfy the R_{self} and R_{other} relationships — in this case, a value must be found which is DISJOINT both from C and C$\triangle$7. If A were to choose to play a D chord, the interaction would be as shown in Fig. 7.

6.2. Extension to multiple agents

The simple formulation of Musical Actions used a pair of agents playing together. In a more general case, it is necessary to look at the actions which occur between several musicians at once. An easy solution for this would be to suggest that the interactions between each pair of agents be analysed separately. This has the benefit of not introducing any more theoretical complexity, but does not capture the fact that in most musical situations, the group will be in agreement on the majority of features.

To rectify this, the notion of *common acceptance* is introduced, based on the *musical common ground* which was introduced earlier. This is designed to capture the set of musical values which the group agrees on through their playing.

Recall that there is a common ground of values which is the set of all values for each facet that an agent reasonably expects the others to understand. Looking at a single facet from this common ground gives the set of values which the agents are producing. Taking the *join* of this set of values gives the most specific value for that particular facet which contains the playing of every agent — this is the *commonly accepted* value for that facet.

This formulation allows for an individual agent to construct groups of agents whose playing is in some way similar, and construct responses to the group, rather than each agent individually. Appendix C gives an example of this.

6.3. Relations between Musical Acts and musical action signatures

Section 2.1 gave a textual, top-down description of a method for interpreting musical interactions in terms of intentional actions, which provided the impetus for creating this model of musical interaction. One of the prerequisites for using Musical Acts computationally is a semantics for the conditions of expressing a musical act, and constraints on the form in which it is expressed. In order to do this, we now revisit the performatives given in Appendix B, and attempt to formulate them in terms of musical action signatures. The situation is taken to be: A is performing an action on facet f by emitting v_{new} after previously emitting v_{prev}; v_{old} is used to indicate the value played by the other agents — if differentiation is needed, then v_{all} refers to the set of individual values, and v_{common} refers to their common musical ground.[6] The seven performatives can then be formalised as follows:

Propose requires that there was previously no accepted value for a particular facet, i.e. that $v_{old} = \odot$; hence, R_{self}, R_{other} must be SUBSUMED (since $\odot$ subsumes everything). Different formulations could be created around whether $v_{old} = \odot$ uses v_{all} and hence means that none of the agents have a value for f (Propose-New), or that v_{common} is used, so there is no common value for f (Propose-Discussion).

Confirm conveys an acceptance of an idea proposed by another; so, R_{other} should be SAME (although SUBSUMED might be allowed). It also requires that the new value was not previously contained in v_{prev}, so $R_{prev} \in$ {SUBSUMES, ALTER, DISJOINT}. Additionally, since $R_{other} =$ SAME, R_{self} must be the inverse of R_{prev}.

Reject indicates that the new playing is not accepted, so it must be different from it; hence $R_{other} \in$ {DISJOINT, ALTER}. It could also be argued that there should be no commonly accepted value for this facet, i.e. $v_{common} = \odot$.

Extend is the extension of currently accepted material, so R_{other} must be SUBSUMED. Since it is an extension, R_{self} cannot be SUBSUMES, as that would indicate a withdrawing from the current position.

Alter involves altering a value which is already being used; so R_{other} must be ALTER.

[5] To be clear, R_{prev} for this signature is between the same two values as R_{other} from the previous signature, but the direction of the relationship is opposite, since A is now the one carrying out an action.

[6] See Table 3 for a reminder of the symbol meanings; in particular $\odot$, which means no value has yet been set for that facet.

Table 2
Formulation of example set of performatives using action signatures.

Performative	R_{self}	R_{other}	R_{prev}	Additional
Propose	SUBSUMED	SUBSUMED	any	$v_{old} = \odot$
Confirm	R_{prev}^{-1}	SAME	SUBSUMES ALTER DISJOINT	
Reject		DISJOINT ALTER		
Extend	¬SUBSUMES	SUBSUMED		
Alter		ALTER		

Argue was an example of a composite act, where several people play without accepting each other's values; this could be modelled as a stream of acts where R_{other} was continually either ALTER or DISJOINT.

Request is a conventionalised action, and so it cannot be modelled with action signatures — it is an expected response to a certain pattern of playing.

It can be seen that a particular musical action could be used in several different performatives; for example, (SUBSUMED, SUBSUMED, SAME) could be either Propose or Alter, depending on the musical context and the interpretation of any particular agent, which is as desired — attribution of intention should not be a formally derived process at this level. However, in the context of building a musical agent system, the question arises of the computability of these Musical Acts. In several agent communication languages (for example [35,36]), performative intentions are explicitly represented. This is an approach which could be used *within* a multi-agent system — each agent could label parts of the musical surface with performatives to indicate intentions. However, this would not then extend to an understanding of human music playing. It would hence be necessary to build a model of intentional musical behaviour, which could be trained on a large corpus of human data, that was combined with the semantics for Musical Acts to give a possible interpretation of musical interactions.

7. Case study

To test the workability of this theory in practice, MAMA was constructed — a prototype multi-agent musical system built on the principles of Musical Acts. The system provides for the agents to have individual sets of analysis procedures, with which they analyse the playing of others to produce a stream of Musical Action Signatures. These Musical Action Signatures are then fed into a decision making module, which has been trained on a corpus of human playing to have a notion of how human players respond in an improvisatory situation.

Appendix D details the functioning of this system, and a pilot experiment carried out using it. A more comprehensive overview of the experiment and other case studies using the system can be found in Murray-Rust [14, Sections 10.2, 11].

The experiment was designed to measure the perceived performance of the system by comparing the subjects' evaluation of different musical partners — live and recorded humans, and the system in different configurations. By comparing the rating of the humans for the different configurations, it was hoped to be able to determine whether the addition of components based on Musical Acts improves the performance of the system.

"Canto Ostinato" (Simeon ten Holt) was chosen as the target piece for a variety of reasons, but the most important one was that it illustrates a form of music where the notes are fixed, but the players have a large amount of freedom to choose the way the piece sounds. Within the given notes, the manner in which the notes are played — in particular which notes are stressed — allows the player to choose certain melodies implicit in the sets of notes to pick out as interesting. This makes it a useful testbed for interaction as it is constrained enough to allow for comparisons, but also open enough to allow the players a large degree of musical freedom.

The system used a set of analyses which examined levels, slopes and patterns for note timing, length and loudness, which were selected as appropriate for the piece at hand. An underlying assumption of this assumption is that an intelligent musical agent would have a range of configurable analysis routines which could be brought to bear on the musical material, and appropriate ones would be chosen, although this is beyond the current scope of the system.

The experimental results were unfortunately not significant, although this is at least partially due to the small sample size. However, it demonstrated the implementability of the theory.[7]

8. Discussion

From a theoretical perspective, the examples given in Section 5.3 show how this technique can be applied to aspects of musical analysis, and produce a compact description of the communicative actions undertaken. Similarly, the discussion in

[7] Selected audio and MIDI recordings from the experiment can be found at http://www.mo-seph.com/MusicalActExperiment.

Section 6.3 shows that Musical Actions can be tied into a more intentional level of analysis. Finally, the experimental section examined the question of whether or not the theory improves the performance of musical agents in their interactions with humans. Next, we discuss in more depth the questions of whether the model is implementable, how it relates to previous work and what improvements could be made in the future.

8.1. Implementation

Section 7 (and Appendix D) detailed a proof of concept system which embodies the basic mechanisms of Musical Act Theory. The components of the theory which were implemented were:

- A collection of basic analysis and generation routines, which could derive values for *facets* from an incoming audio stream, and calculate the requisite relations between them.
- A simple mechanism which could generate new Musical Action Signatures based on a combination of the values found in playing and a database extracted from previous analysis of human playing.
- A routine for generating new facet values from the given MASs, which could then be used by the generation routines to create real time output.

All of these operations could be carried out on a standard laptop (Pentium 1.3 GHz, 2 GB RAM, without problems. This shows that at a basic level, all of the components of the framework can be implemented — analysing music to extract values, extracting MASs, deciding on MASs and then finding based on these to use when generating output. In the future, we envisage fuller implementations, allowing more musical versions of each of these components: more advanced analysis and generation routines, a more musically advanced method for generating MASs and potentially higher level mechanisms for managing the operation of all these components in the context of more general improvisational situations.

8.2. Relation to previous work

The most similar formulation of actions taken in a musical context is given in Pelz-Sherman [10]. Musicians are treated as intelligent agents, who are capable of producing and interpreting musical signals, and using intelligence to generate new plans and alter behaviour to optimise the performance of the group. The interaction between these agents is modelled in communication theory terms: each agent may be either *sending* or *receiving* musical information, termed *i-events*. Agents are *sending* when they play music which has a high rate of change, or a lot of "musical information".[8] Several types of *i-events* are discussed, relating to different circumstances in which musical information is exchanged between two participants. It would be useful to define these in terms of musical act signatures, as they would then provide a computational implementation of an already existing theory about the nature of musical interaction.

The following list details all of the possible *i-events* from the formulation, and means for them to be translated into musical action signatures:

Imitation is where one feature from an agent's playing is adopted by another. In terms of musical actions, the important fact is that R_{other} is `SAME`. However, it is also important that the previous playing did not contain this feature, so R_{self} and R_{prev} must be `ALTER` or `DISJOINT`.[9]

Question and Answer events consist of some form of response to a cue, but the response need not use any features of the cue. Some of these are stylised, and hence not detailed here, but the general definition would be that R_{other} is either `ALTER` or `DISJOINT`. At this point, it might be useful to look at formulations across multiple values, such that some of the R_{others} are the same, while some are `DISJOINT`.

Completion/Punctuation occurs when one agent initiates a "directed movement", which can be predicted to complete at a certain time, and another agent completes this gesture. This can be modelled with R_{other} becoming `SAME`, on a facet which looks at these kinds of directed gestures.

Interruption involves one agent playing in an undirected manner, which is decisively responded to by another. This could be modelled as a specific case of R_{other} being `ALTER` or `DISJOINT` on a feature which tracks some sense of musical direction.

So, musical action signatures can be used to provide a computational formulation for all of these *i-events*, conditional on there being the necessary analysers to produce "musical direction" values. This is a fairly reasonable constraint, as to be a high quality musical performer, an agent would need to have some ideas about the directions that individual agents and the performance as a whole are taking.

Also, Pelz-Sherman [10, p. 130] mentions the idea of agent systems — subgroups of musicians whose playing is very closely aligned. The notion presented here of overlapping regions of common ground within a group of agents gives a way to analyse this computationally, and allow software agents to join these human agent systems.

---

[8] The exact nature of musical information is not defined by Pelz–Sherman, but the intent should be clear.

[9] It should be noted that R_{self} must have the same value as R_{prev} if R_{other} is `SAME`.

Finally, a similar distinction is made between free and conventionalised musical actions: *i-events* cannot be part of some prearranged schema, and must occur in the moment, just as musical actions.

Another theory of musical improvisation is that provided by Pressing [37]. This work is very cognitively oriented, and seeks to claim cognitive validity for all of the process it describes. The process of improvisation is modelled as:

$$\left(\{E\}, C, R, \mathcal{G}, M\right)_{i_k} \rightarrow E_{i_k+1}, \quad k = 1 \ldots K$$

where $\{E\}$ is the previous playing of the improviser, C is the cognitive representation the improviser holds of the playing of other members of the group, R is a referent or "score", $\mathcal{G}$ is the improviser's goals and M is the improviser's long term memory. The musical events have three components: *objects* are the lowest level representation of music dealt with, i.e. the musical surface, *features* are properties of those objects, and *processes* are the parameters and processes which give rise to those objects and their features. The improviser has three potential routes for generating new material:

Similar associative generation involves creating new events based on the current processes and parameters with minimal changes;
Contrastive associative generation involves changing some but not all of the major parameters used;
Interrupt generation sets all of the processes and their parameters to new values.

The improvisation is seen as a series of groups of event clusters, with the first cluster in each group being produced by interrupt generation, and subsequent clusters by associative generation.

This can translated into relationships from *Rel*: interrupt generation is `DISJOINT` with previous playing, while associative generation is one of the others. Since no formal distinction is made between similar and contrastive association, it might be reasonable to suggest that contrastive association is typified by `ALTER` and similar by `SAME`, `SUBSUMES`, `SUBSUMED`.

Once this is done, the contrast between Pressing's work and Musical Acts is clearer: Pressing deals in detail with the relations between a musician's playing and what has gone before (and the processes underlying the playing), but only cursorily with the relations between a musician's playing and that of the other members of the group. Also, Musical Acts gives a way to distinguish between the different possible relationships which does not rely on knowledge of internal cognitive states. Apart from this, the two theories can be seen as broadly compatible, although Musical Acts makes no demands about the processes which give rise to the musical output.

It is argued that both Pelz-Sherman's and Pressing's work can be at least partially represented within the theory of Musical Acts, and certainly are not contradictory in any major dimension. Musical Acts is a computationally implementable system, which goes beyond existing work in specifying relations between the playing of different musicians, and hence represents an advance on previous theories.

8.3. Further work

There are some areas which have been noted as needing more theoretical development, particularly around coherence of analysis, importance of actions, the use of repetition and the roles agents take within an improvisation. Firstly, all the different facets of analysis are currently treated separately. Real values in musical playing are unlikely to be entirely separate, so it is necessary to develop some notion of an action which affected several values at once, especially since a single piece of playing may do this. A composite musical action could then involve the attachment of a set of musical action signatures to a certain part of the musical surface.

Secondly, there is no built in mechanism for assigning different levels of importance to the different changes. This could, however, be supported both at the level of comparing changes within a facet (i.e. is this volume change more important than that one), and across facets. This could also link in with the notion of compound actions relating changes in several values, and would for a key input for systems when they decide what actions to respond to.

Next, the idea of repetition is not fully handled at the moment; in general terms, playing the same phrase repeatedly would generate actions on the first repetition, but then no new information would be added. However, this is not entirely in keeping with accounts of listening to music, especially where minimalist (and other repetition influenced) traditions are considered. This could potentially be handled with Analysers which look at repetition, or it could be addressed with a more fundamental extension to the theory.

Finally, nothing is said about the roles which agents take in interactions — soloist/backing, teacher/student, etc. This is intentional — it is a layer to be built on top of these structures. From the study of human interaction, e.g. Pelz-Sherman [10], there is a potential for roles to be abstracted and defined in terms of the type of musical action signature which that role is likely to produce. On a related note, the role of the listener has not been addressed in depth; it is assumed that a listener can use the same techniques as a performer, simply without the ability to join in.

One underlying assumption which needs more exploration is the idea that every interesting musical property can be represented sensibly using concept lattices. In response to this, it should be noted that the use of these structures was inspired as a more general version of the reduction hypothesis in [1]: their analyses of time-span and prolongational reduction depend on any given set of pitch events being seen as an elaboration of a simpler structure — this is fully in keeping with the idea of subsumption used in the lattice structures here. Although there are defined limits to the applications of this technique, it at least provides an example of representing complicated musical structure using lattice-like concepts.

Since the model is designed to be used in the context of an intelligent system, there are many possible avenues for further work, but two of the most pressing are learning, and the assignment of intentional performatives to musical action signatures.

Firstly, it is clear that musicians learn over time, both between performances and within the scope of a single interaction. In the context of the MAMA system, a simple form of learning was used to help the system understand common patterns in human interaction; this initial experiment could be widened and made more robust in a variety of ways. However the more interesting aspects of learning could be applied to the parts of the system which had to be specified for the particular piece being played — in particular, the use of certain analysis routines could be a more dynamic choice, as the system experimented with analysing incoming music in different ways, it until if found analysers (and their related output routines) which could both provoke and perceive change in the output of the other musicians.

Secondly, the formalisation of performative actions opens up the possibilities for virtual musicians to respond to the intentional aspects of human playing; with a given set of performatives, there is a limited set of labels which can be applied to a musical action, and some process could then be used to allow an agent its own strategy for choosing one of these labels — for example by modelling the other musicians, or referring to a database of past interactions. The formalisation also allows and encourages different sets of performatives to be experimented with, to find whatever set is most appropriate for a given circumstance. Desirable qualities of performative sets might be formal completeness — any action can only have one performative label, or all possible actions have *some* performative label — or resonance with natural language usage.

9. Conclusions

In this paper, we have developed a computational model of musical communication, using concept lattices. It allows for a dense, style-independent description of the communicative actions which occur when agents play music together, and provides the agents with a framework which can be used to create notions of musical common ground and to reason about the actions and beliefs of other agents. Furthermore, this model has been implemented in a real-time system, to play duets with human pianists. Finally, it was shown this model can be used to describe previous formulations of musical communication, but allows for richer analysis as well as computational implementation.

We believe that this model can be the starting point for a rich variety of systems which interact musically with humans. We hope to inspire the creation of musical systems which use Musical Acts as a framework for interaction, aided by the fact that the current formulation is computationally implemented and so can be added to existing systems where appropriate. Finally, understanding the relations between this formal system and human music making will drive the next level of richness and complexity in Musical Act Theory.

Appendix A. The universe of action signatures

An action signature consists of a triple of relations, where each relation can take one of five values in

$$Rel \underset{def}{=} \{\mathtt{SAME}, \mathtt{SUBSUMES}, \mathtt{SUBSUMED}, \mathtt{ALTER}, \mathtt{DISJOINT}\}$$

At first glance, this would appear to give 125 possible values for an action signature. On further inspection, however, this is an overestimate; for example, it is not possible to have an action where R_{self} is SAME — there would be no grounds for calling it an action. There are other constraints on which relations can hold between three values, and this is the subject of an investigation in [14, Appendix B].

R_{self}	Allowed values for R_{other}
	$R_{prev} = \mathtt{SAME}$
SUBSUMED	SUBSUMED
SUBSUMES	SUBSUMES
ALTER	ALTER
DISJOINT	DISJOINT
	$R_{prev} = \mathtt{SUBSUMED}$
SUBSUMED	SUBSUMED
SUBSUMES	SAME, SUBSUMED, SUBSUMES, ALTER, DISJOINT
ALTER	SUBSUMED, ALTER, DISJOINT
DISJOINT	DISJOINT

R_{self}	Allowed values for R_{other}
	R_{prev} = SUBSUMES
SUBSUMED	SAME, SUBSUMED, SUBSUMES, ALTER
SUBSUMES	SUBSUMES
ALTER	SUBSUMES, ALTER
DISJOINT	SUBSUMES, ALTER, DISJOINT
	R_{prev} = ALTER
SUBSUMED	SUBSUMED, ALTER
SUBSUMES	SUBSUMES, ALTER, DISJOINT
ALTER	SAME, SUBSUMED, SUBSUMES, ALTER, DISJOINT
DISJOINT	SUBSUMES, ALTER, DISJOINT
	R_{prev} = DISJOINT
SUBSUMED	SUBSUMED, ALTER, DISJOINT
SUBSUMES	DISJOINT
ALTER	SUBSUMED, ALTER, DISJOINT
DISJOINT	SAME, SUBSUMED, SUBSUMES, ALTER, DISJOINT

Appendix B. Little Blue Frog analysis

"Little Blue Frog" [32], on the Columbia/Legacy reissue of "Big Fun" is used as an example for what a Musical Act analysis of a piece of music looks like. Annotated audio for this can be found at http://www.mo-seph.com/academic/littlebluefrog. See also Table 3.

The analysis is carried out from a personal listening perspective — that is, without recourse to transcriptions or other theoretical explanations, using only the first author's individual musical competences, and to offer a view of what was personally found to be the most interesting parts of the interaction. For example, in the analysis, "scale" has been used to refer to a collection of pitch classes used when playing. Since the first author can distinguish between these scales but not give them their correct musical names, they have been assigned arbitrary numeric designators.

This analysis was carried out from the point of view of a listener who was not party to rehearsals or discussions prior to the performance of the piece; this means that initially, the author's representation of what would occur in the piece was empty, or contained general stylistic expectations of what the first author thinks a Miles Davis fusion piece is likely to sound like. To the musicians who were playing, a different set of musical features would be seen as novel and intentional, and a lot of the changes which were noticed by the first author would be expected due to whatever score had been agreed on beforehand.

The performative labels attached to the musical act analysis presented here were constructed in an ad-hoc manner. Hence, the terms which were used should be defined and discussed more thoroughly. However, this is not a formal or exclusive definition — more a sketch of the type of analysis which musical acts are aimed at creating.

The performatives used in this analysis, and their intended meanings were:

Propose occurs when an agent introduces a new musical idea; this should be an idea which does not conflict with any of the material which is already present — for example, introducing a harmonic structure when previously only percussive strikes were being played, or introducing a melody when previously only chords and rhythms were being played. In the analysis, the trumpet introduces a lyrical phrase based around a particular scale (0:34–0:55). The performative intention is that the idea being introduced becomes a part of the shared representation.

Confirm occurs when an agent (A) proposes an idea, and another agent (B) indicates amenability to working with this idea; a typical way to carry this out would be for (A) to adopt the idea in its own playing. Again at (0:34–0:55), the clarinet takes up the musical idea introduced by the trumpet. The performative intention is that the new idea becomes part of the shared representation for the agents.

Reject is used by B to indicate unwillingness to adopt A's suggestion — for example, by playing something different to A's idea, or emphatically not taking it up. The performative intention here is that the new idea is not taken up, and does not become part of the shared representation. (1:08–1:35) shows the clarinet suggesting an idea, and the trumpet rejecting it by returning to previous ideas.

Extend happens when an agent presents an elaboration of an already existing idea — for example adding extensions to a chord, playing an elaboration of melody. The intention again is the addition of the new idea to the agents' shared representation. At the beginning of the analysis (0:35–0:55), the xylophone uses the scale which is present, but adds additional, dissonant elements.

Alter is used to change an existing musical idea in some way which has a relationship to the existing material, but cannot be seen as an elaboration of it — for instance changing from a bossa-nova to a son rhythm, or substituting a chord in a chord sequence with one which is harmonically related. (This is a suggested act, not present in this particular analysis.)

Request has the intention of causing an action which is not musically related to the musical idea embodied in the action — in the analysis (3:35–4:11), the snare drum plays a crescendo which both provides a point of organisation and a steadily increasing tension which indicates a desire for a new musical direction, without specifying what that direction is.

Table 3
Musical act analysis of Little Blue Frog, by Miles Davis.

Time (s)	Instrument	Performative	Description
0:00–0:09	Bass, guitar	PROPOSE	Bass+electric guitar establish a tonal centre. Bass suggests tonic, and guitar replies with an extension
	Bass, guitar	CONFIRM	Repeated as a confirmation, and then the guitar suggests an alternative extension
0:09–0:34	Cuica	CONFIRM	Cuica confirms the current beat
	Tablas	PROPOSE	Tablas suggest a “1 2+” rhythm
0:34–0:55	Trumpet	PROPOSE	Trumpet suggests a scale 1 and a lyrical phrase
	Clarinet	CONFIRM	Clarinet echoes the same scale, in a similarly lyrical style
	Xylophone	EXTEND	Xylophone follows the given scale, but adds dissonant elements
	Backing	None	Backing abandons complex rhythm
0:55–1:08	Trumpet	EXTEND	Exploring scale 1
	Clarinet	CONFIRM	Follows trumpet’s exploration of scale 1
	Triangle	PROPOSE	Suggests a more explicit rhythm
	Trumpet	CONFIRM	Spiky notes agree with triangle’s suggestion
1:08–1:35	Triangle	REQUEST	A loud clang signals the start of a new section
	Clarinet	PROPOSE[a]	A new scale 2 still in a lyrical style
	Trumpet	REJECT	By sticking with scale 1, the trumpet rejects the clarinet’s proposal
1:35–1:47	Xylophone	CONFIRM	Xylophone confirms clarinet’s scale 2
	Trumpet	CONFIRM[b]	Trumpet gives in to the new scale, and leaves with a few parting blasts
1:47–2:00	Triangle	EXTEND	Many muted notes extend and emphasise the rhythmic ideas
	Trumpet	EXTEND	Using scale 2, the trumpet adds an increasing rhythmic element
	Clarinet	PROPOSE	Clarinet introduces a new scale 3 (more eastern sounding), ignores the trumpet’s rhythmic direction and continues lyrically
2:00–2:13	Trumpet	PROPOSE	A spiky, stabbing phrase, based on scale 2
	Clarinet	CONFIRM	Briefly seems to agree with the trumpet
2:13–2:29	Bass clarinet	CONFIRM	Confirms scale 3
	Trumpet	REJECT	Ignores bass clarinet, and continues with stabs
	Clarinet	REJECT	Ignores bass clarinet, and continues with lyricism in scale 2
2:29–2:43	Trumpet clarinet Bass clarinet	ARGUE	All play lyrically, with clarinet on scale 2, trumpet on 1 and bass clarinet in 3
2:43–3:08	Trumpet	PROPOSE	Proposes a resolution, by playing stabs which fit with any of the scales
3:03–3:08	E-PianoVibes	CONFIRM	Supports the trumpet’s resolution
3:08–3:17	Backing	REJECT	Increased dissonance and rhythmic confusion reject the proposed resolution
3:17–3:35	All	CONFUSION	
3:55–4:11	Snare	REQUEST	Snare enters to call for new section with crescendoing 8th notes
	Winds	CONFIRM	Winds join in with the 8th note idea
4:08–4:11	Triangle	CONFIRM	A bar of loud crotchets pinpoints the section change called for by the snare drum

[a] Repeated suggestion, so maybe followed by PROPOSE-AGAIN?
[b] Also withdraws — do we need a WITHDRAW?

Argue is a composite act; it happens when several musicians are presenting conflicting ideas at the same time — for example, at (2:29–2:43) where the melody instruments are all playing with different scales.

The performatives used in this analysis were: PROPOSE, CONFIRM, REJECT, EXTEND, ALTER, REQUEST, ARGUE.

Appendix C. Musical common ground example

To illustrate ideas about analysis and common ground, we use a caricature of a jazz trio:

- A is a drummer, with an advanced understanding of rhythm, but no knowledge of chords.
- B is a bass player, and has an understanding of chords and rhythmic patterns.
- C is a pianist, with an advanced understanding of chords, but little idea about rhythms.
- D is a guitarist, with the same capabilities as the pianist idea about rhythms.

They each then have their own set of analysers, as shown in the table — for instance, A has $A_{structure}$ which keeps track of where they are in the piece, $A_{time\text{-}signature}$, and A_{rhythm}.

See Fig. 8 for a diagrammatic version of this.

At a particular point in the piece, the outputs of these analysers are:

- (Time-Signature: 4/4),
- (Harmony: C) (A simple knowledge of the root of the chord),
- (Harmony-Adv: C7) (More refined harmonic analysis of the chord),
- (Rhythm: basic swing),
- (Chord: 7th).

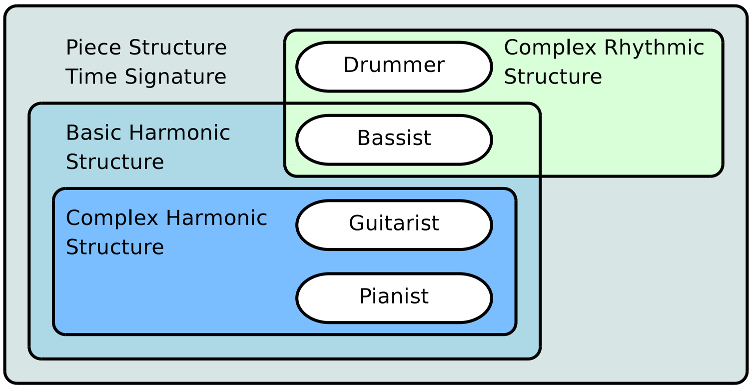


Fig. 8. Example of different common grounds for a group of agents playing jazz.

Here, there are several common grounds, assuming that each agent has perfect knowledge of the other's capabilities:

- between the entire group about what the time signature is, and what the structure of the piece is,
- between the bassist, guitarist and pianist about the basics of the chord sequence,
- between the guitarist and pianist of the complex harmonic material being used,
- between the bassist and drummer about the complex rhythmic material being used.

For example, it might be expected that the drummer had knowledge of chord sequences. In this case, the other players would suppose that the common ground included (Harmony: C), and be surprised if they learnt that this was not the case.

Agent	*CV*
A	(Time-Signature: 4/4), (Structure: Verse), (Rhythm: basic swing)
B	(Structure: Verse), (Time-Signature: 4/4), (Rhythm: basic swing), (Harmony: C)
C/D	(Structure: Verse), (Time-Signature: 4/4), (Rhythm: basic swing), (Harmony: C), (Harmony-Adv: C7)

Appendix D. Experiment description

An experiment was carried out to test the use of Musical Acts within an interactive situation. We give a brief description here, and a more thorough analysis can be found in Murray-Rust [14, Sections 10.2, 11].

This experiment was designed to measure the perceived performance of the system when interacting with human musicians. Pianists were asked to play a series of short duets with an unseen partner, who they might reasonably expect to be another human. After each duet, they were asked to rate their partner's performance by scoring their responses to a series of questions. In order to differentiate aspects of the musical experience, a series of situations were created which were hypothesised to have different amounts of *expressivity* and *interactivity*. These conditions were:

Human: another human (expressive and interactive),
Recording: a recording of a human (expressive but not interactive),
Straight: a completely mechanical rendering of the music (neither expressive nor interactive),
Mirroring: the system copying features from the human's input (expressive, and slightly interactive),
Reasoning: the system using the learnt interaction data to inform its responses to the human player (expressive, and more interactive).

By comparing the rating of the humans for the different configurations, it was hoped to be able to determine whether the addition of an understanding of Musical Acts to a system improves its performance.

The participants were asked to play "Canto Ostinato", by Simeon ten Holt. This is a piece for two or more pianos, where repetitive phrases are played for a length of time determined by the performers. In the experiment, a small range of bars were chosen, and then number of repeats defined, in order to simplify the pianists' task, and allow them to concentrate on the expressive side of the musical interaction.

The steps taken to configure the system for the experiment were:

- encode the score in an appropriate, machine readable form,
- define a set of features which could be used to inform analysis and generation of playing,
- learn sequences of actions within the space of features from example human playing.

D.1. Features and relations

A limited set of feature categories were used, representing those aspects of playing over which the performers were allowed control:

- note volume (velocity),
- onset times (relative to a metronome),
- note lengths (as proportions of their scored values).

For each of these categories, a feature was calculated describing:

- the average value,
- a trend in the average value,
- a *pattern value*, consisting of the residuals once the average and the trend have been removed, for each note position in a bar (based on the assumption that each bar consists of two groups of five equally spaced notes). This is designed to model repetitive patterns of accents, e.g., placing emphasis on the first and third notes of each five note group.

These were calculated first as numeric values, and then segmented to give symbolic values; for instance, note volumes were expressed with the standard musical markings such as *pp*, *mf* and *f*. In each case the lattice values are formed as follows:

- for the basic value (e.g. dynamics, onset timing), the medium value is the root of the lattice, with more extreme values making up the branches of the lattice (see Fig. 9). From this, we can say that e.g. *mf* subsumes *f*, and is disjoint from *pp*;
- for *pattern values*, the pairwise *meet* of each element in the pattern is calculated. If this is equal to value A, then A subsumes B (and vice versa). If each value is $\odot$, then they are disjoint, and otherwise A is an alteration of B.

This gives a total of 9 facets:

$$\{dynamics, onset, length\} \times \{average, trend, pattern\}$$

These features were tuned to the piece at hand — in particular, the use of five note patterns. The features analysed here are essentially numeric in nature, and were symbolised in a relatively crude manner — this is a particular style of analysis which was appropriate for working with this piece.

For all of these features, an analysis procedure was created which could extract their symbolic values from music, and derive relations between pairs of values. This provided the basis of the analysis section.

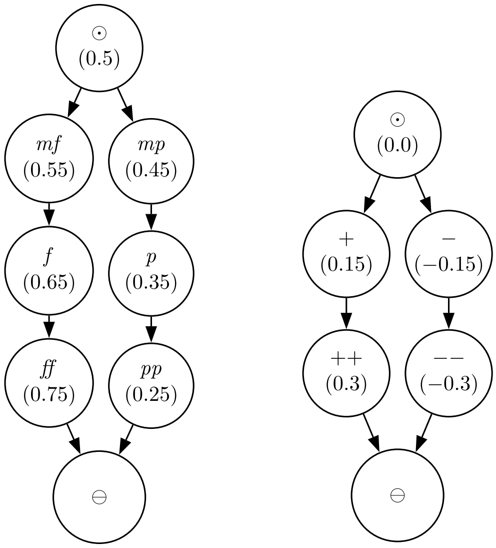
D.2. Deliberation

In order to use these features in a musical act context, a mechanism is needed to first generate new action signatures, and then find a realisation of those signatures. An initial phase of learning was performed to give the system an initial musical grounding, which was then used for deliberation during the experiment. The learning phase was as follows:

- Recordings were made of a pair of pianists playing the excerpts which would be used in the experiment.
- These recordings were annotated using the analysis procedures described above, to produce symbolic values for all the different facets.
- Based on the assumptions made in Section 5, and the relationships between values discussed previously, Musical Act Signatures were extracted from these symbolic values. A separate stream of MASs was produced for each facet under analysis.
- All of the streams were the used to build up a tree representing all of the sequences of MASs which have been encountered, and their counts. This represents the system's memory of patterns found from human interaction.

During the interaction, the results from this were used to inform the playing of the system as follows:

- Values were extracted from the human's playing, using the same automatic procedures described above.
- Any new values are combined with the current context to produce a MAS. For each facet under analysis, a sequence of MASs is maintained, containing both the human and the system's actions.
- When a new value from the human is encountered, the current sequence is passed to the sequence completion algorithm. This contains all of the sequences of MASs learnt in the previous step, and will
  - match the longest subsequence of the current MAS sequence which is present in the database,



(a) Numeric lattice for average dynamics values

(b) Numeric lattice for dynamics pattern values

Fig. 9. Example numeric lattices.

  – randomly pick a subsequent MAS based on the counts for each possible continuation of this sequence found in the corpus.[10]
- Once a suitable MAS was chosen, the system would search for a value which fulfilled it, and then enact that in its playing. In general terms, the search can be seen as partitioning off the graph according to the current values and the relations given by the chosen MAS, and then randomly choosing a value from the selected partition.

D.3. Example deliberation

In order to demonstrate the behaviour of the system when playing with a human pianist, we present a short constructed example. This example will focus on the mean dynamics facet, i.e. the general overall loudness with which one is playing. During operation of the system, an identical process is carried out for all of the 9 facets under analysis.

- During the course of playing, both the human and the system are playing *mf* when the human's average level shifts to *ff*. The action has the signature H: $(\texttt{SUBSUMED}, \texttt{SAME}, \texttt{SAME})$.
- This is added to the sequence which is being maintained for mean volume, and a subsequent MAS is suggested: C: $(\texttt{DISJOINT}, \texttt{DISJOINT}, \texttt{SUBSUMES})$, i.e. to play something which is significantly different from the playing of both the computer and the human.
- In this case, this is then limited to values from the other branch of the lattice (see Fig. 9), and anything from *mp* to *pp* could be chosen.
- Alternatively, if C: $(\texttt{SUBSUMED}, \texttt{SUBSUMED}, \texttt{SUBSUMED})$ had been chosen, this would limit actions to values which were further down the right hand side of the lattice, and only *ff* would be possible.
- Whichever value is chosen, the system would then start playing music at that volume.

D.4. Results

The experiment was run as a pilot study, using pianists studying at Napier University. 5 subjects were used, each carrying out 12 interactions, with the condition of each action randomly chosen and blinded. Responses consisted of numeric scores for 13 questions, grouped roughly into questions about interactivity, expressivity, and general competence.[11] The 13-dimensional scores were then reduced to 3 factors using PCA. By looking at the loadings of each factor onto the questions, the factors were loosely mapped onto:

1. interactivity,

[10] This is very loosely modelled on behaviour from the Continuator [27].
[11] These groupings were not available to the participants.

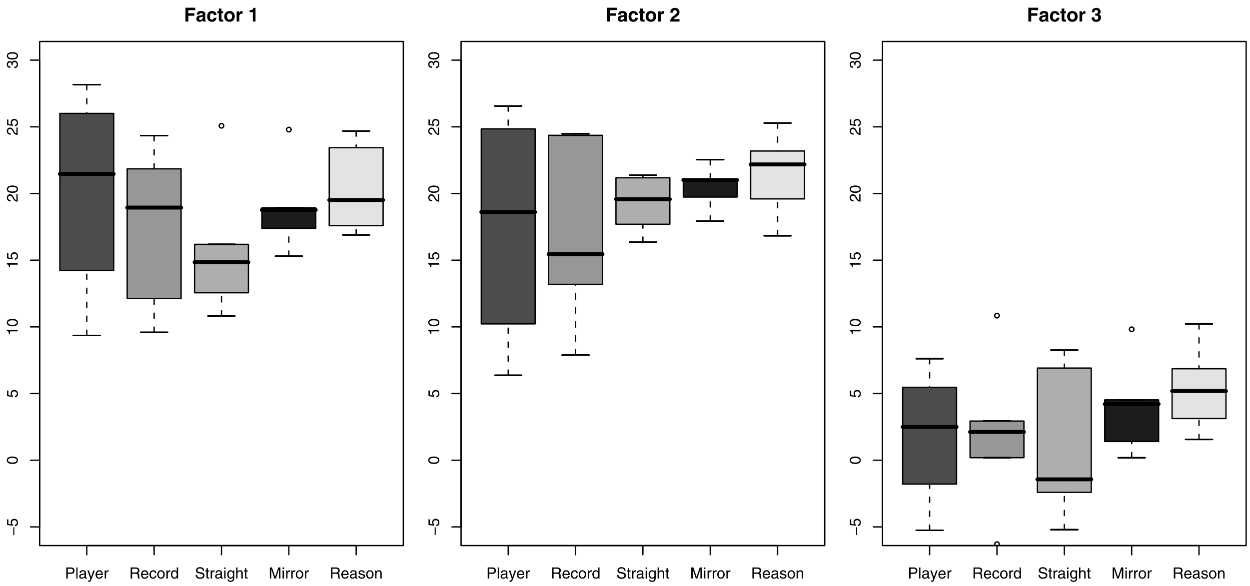


Fig. 10. Distribution of factor scores for all conditions (for 3 factor varimax factor analysis).

2. competence and expressivity,
3. interactivity and general performance.

The aggregate results are shown in Fig. 10. Looking at the mean scores of these conditions, and performing t-tests, the following hypotheses were tested (using $\alpha = 0.05$):

- A human is seen as more interactive than a recording of a human (Factor 1). $t(3) = 0.56$, $p = 0.3$, not significant.[12]
- The system in "mirroring" mode is more interactive than a straight recording. $t(4) = 1.98$, $p = 0.06$, not significant.
- The system in "reasoning" mode is more interactive than the system in "mirroring" mode. $t(4) = 1.26$, $p = 0.14$, not significant.

Despite the fact that all of the differences in means are in a direction which supports these hypotheses, no significant differences were found. However, this may be attributed to the small sample size ($n = 5$).

References

[1] F. Lerdahl, R. Jackendoff, A Generative Theory of Tonal Music, MIT Press, 1983.
[2] R. Jackendoff, Consciousness and the Computational Mind, MIT Press, 1987.
[3] J. Searle, Speech Acts, Cambridge University Press, 1969.
[4] J.L. Austin, How to Do Things with Words, Clarendon Press, 1962.
[5] FIPA, FIPA communicative act library specification, foundation for intelligent physical agents, www.fipa.org/specs/fipa00037, 2002.
[6] M. Wooldridge, Reasoning about Rational Agents, MIT Press, Cambridge, MA, 2002.
[7] A. Herzig, D. Longin, A logic of intention with cooperation principles and with assertive speech acts as communication primitives, in: Proc. 1st Int. Joint Conf. on Autonomous Agent and Multi-Agent System (AAMAS 2002), ACM Press, Bologna, 2002, pp. 920–927.
[8] K.R. Coventry, T. Blackwell, Pragmatics in language and music, in: M. Smith, A. Smaill, G.A. Wiggins (Eds.), Music Education: An Artificial Intelligence Approach, Workshops in Computing, Springer, 1994, pp. 123–142.
[9] W.F. Walker, A computer participant in musical improvisation, in: Proceedings of Conference on Human Factors in Computing Systems CHI97, 1997, pp. 123–130.
[10] M. Pelz-Sherman, A framework for performer interactions in western improvised contemporary art music, PhD thesis, Music, UC San Diego, 1998, http://pelz-sherman.net/mpsdiss.pdf.
[11] H. Grice, Logic and conversation, in: P. Cole, J. Morgan (Eds.), Syntax and Semantics, vol. 3: Speech Acts, Academic Press, New York, 1975.
[12] D.S. Murray-Rust, A. Smaill, Musical acts and musical agents, in: Proceedings of the 5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications, 2005, http://homepages.inf.ed.ac.uk/s0239182/Murray-Rust_MusicalActs.pdf.
[13] D. Murray-Rust, A. Smaill, M. Edwards, MAMA: An architecture for interactive musical agents, in: Proceedings of European Conference on Artificial Intelligence, IOS Press, 2006, pp. 36–40.
[14] D. Murray-Rust, Musical acts and musical agents: theory, implementation and practice, PhD thesis, Informatics, University of Edinburgh, http://hdl.handle.net/1842/2561, 2007.
[15] G. Wiggins, Music, syntax, and the meaning of 'meaning', in: Proceedings of the First Symposium on Music and Computers, 1998, pp. 18–23.
[16] J. Sloboda, Does music mean anything? Musicae Scientiae 2 (1998) 21–32.

[12] Since there were an odd number of participants, one of them did not play under the "Human" condition, so $n = 4$ here.

[17] I. Cross, Music as biocultural phenomenon, Annals of the New York Academy of Sciences 999 (2003) 106–111, http://www-ext.mus.cam.ac.uk/~ic108/PDF/IRMCNYAS2003.PDF.
[18] K. Atkinson, T. Bench-Capon, P. McBurney, A dialogue game protocol for multi-agent argument over proposals for action, Autonomous Agents and Multi-Agent Systems 11 (2) (2005) 153–171.
[19] C.D. Walton, Multi-agent dialogue protocols, in: AI&M 2004, Eighth International Symposium on Artificial Intelligence and Mathematics, 2004.
[20] B.E. Benson, The Improvisation of Musical Dialogue, Cambridge University Press, ISBN 0-521-00932-4, 2003.
[21] S.W. Smoliar, Parsing, structure, memory and affect, Journal of New Music Research 24 (1) (1995) 21–33.
[22] J. Pressing, Cognitive processes in improvisation, in: W.R. Crozier, A.J. Chapman (Eds.), Cognitive Processes in the Perception of Art, Elsevier, Amsterdam, 1984, pp. 345–363.
[23] M. Minsky, The Society of Mind, Simon & Schuster, Inc., 1986.
[24] M. Minsky, Music, mind, and meaning, Computer Music Journal 5 (3) (1981) 28–44, http://web.media.mit.edu/~minsky/papers/MusicMindMeaning.html.
[25] L.K. Ueda, F. Kon, Andante: A mobile musical agents infrastructure, in: Proceedings of the IX Brazilian Symposium on Computer Music, Campinas, Brazil, 2003, pp. 87–94.
[26] F. Pachet, Enhancing individual creativity with an interactive reflective musical system, in: I. Deliège, G.H. Wiggins (Eds.), Musical Creativity: Current Research in Theory and Practice, Psychology Press, Hove, 2004.
[27] F. Pachet, Beyond the cybernetic jam fantasy: The continuator, IEEE Computers Graphics and Applications 24 (1) (2005) 31–35.
[28] F. Pachet, The MusES system: An environment for experimenting with knowledge representation techniques in tonal harmony, in: Proceedings of the 1st Brazilian Symposium on Computer Music, Caxambu, Minas Gerais, Brazil, 1994, pp. 195–201, http://citeseer.nj.nec.com/148999.html.
[29] F. Pachet, G. Ramalho, J. Carrive, Representing temporal musical objects and reasoning in the MusES system, Journal of New Music Research 5 (3) (1996) 252–275, http://citeseer.nj.nec.com/pachet95representing.html.
[30] J.R. Fonseka, Musical agents, Honours thesis, Information Technology, Monash University, 2000.
[31] R.D. Wulfhorst, L. Nakayama, R.M. Vicari, A multiagent approach for musical interactive systems agents and multiagent systems, in: Proceedings of the Second International Joint Conference on Autonomous, ACM Press, 2003, pp. 584–591.
[32] M. Davis, Little Blue Frog, Big Fun (reissue), Columbia/Legacy, 2000.
[33] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 1997.
[34] R. Stalnaker, Common ground, Linguistics and Philosophy 25 (5) (2002) 701–721.
[35] FIPA Specification, http://www.fipa.org/repository/bysubject.html, 2009.
[36] KQML Spec, KQML Specification, http://www.cs.umbc.edu/kqml/kqmlspec/spec.html, 2009.
[37] J. Pressing, Improvisation: Methods and models, in: J.A. Sloboda (Ed.), Generative Processes in Music, Oxford University Press, 1988, pp. 178–192.