AI & Soc
DOI 10.1007/s00146-013-0511-x

25TH ANNIVERSARY VOLUME

A FAUSTIAN EXCHANGE: WHAT IS IT TO BE HUMAN IN THE ERA OF UBIQUITOUS TECHNOLOGY?

"Does not compute"? Music as real-time communicative interaction

Ian Cross

Received: 29 September 2012 / Accepted: 3 September 2013

Abstract Mainstream operationalisations of music in contemporary digital culture tend to take forms that fit with Western folk-theoretic conceptions of music: as discrete sonic entities—songs, pieces, works—that fall within an autonomous domain of human experience, that have determinate structure and that have both affective and exchange value. This perspective is problematised in alternative digital manifestations of music as constituted in and through interaction, in which music is emergent from interactive processes that are computationally mediated. This alternative digital approach fits with broad conceptions of music that are grounded in ethnomusicological accounts and that have increasing weight in the cognitive sciences, in which music is understood and explored as a communicative medium. This paper will outline some of the possibilities, potentials and problems for digital approaches that are likely to arise in operationalising music as communicative interaction.

Keywords Music · Interaction · Speech · Digital representation · Phatic communion

1 Contemporary digital culture and Western folk-theoretic conceptions of music

In this paper, I shall be arguing that the ways in which music has been conceived of and addressed in the digital domain have limited the scope of computational applications to music. In general, computational approaches have dealt with music as an autonomous domain of human thought and behaviour based on complexly patterned sounds that are engaged with through listening for their emotional or hedonic value—music is manifested as sonic objects or entities that have affective, individual and (potentially) commercial value. Music thus appears as an aural commodity, representable in digital terms directly (as audio) or symbolically.

This approach to music fits with Western "folk theories" concerning music and its powers in Western culture: theories that are not intended to be definitive or to provide foundations for scholarly analysis, but rather that arise informally to guide action (see Walton 2007). In Western folk theories, music is complex, humanly produced, expressive sound (Feld and Fox 1994, p. 28), engaged with through listening because of its capacity to move our emotions (see McLucas 2010, Ch 4) rather than for any message it might convey; it is produced—composed and performed—by the few, and the predominant means through which the many engage with it is listening; and it exists as works, pieces or songs, entities that have exchange value as commodities.

Within the principal areas in which digital approaches to music are currently represented—audio representation and manipulation, music information retrieval, audio-to-symbolic translation and vice versa—music is conceived of and computationally represented in terms of objects (pieces, songs) that may be decomposed into smaller objects (sections, phrases, motifs, rhythms, pitches). The digital representation and manipulation of music as complexly patterned sound still pose interesting challenges after more than 50 years of research and development. Applications and developments of music information retrieval (MIR) as outlined in Downie et al. (2009) are only really viable in respect of discrete entities—works, pieces or songs,

I. Cross (✉)
Centre for Music and Science, Faculty of Music, University of Cambridge, West Road, Cambridge CB3 9DP, UK
e-mail: ic108@cam.ac.uk

Published online: 11 October 2013

grouped into distinct corpora—within an overarching taxonomic system (e.g. genre typology), entities that have complex internal structures that conform to generalisable principles (in terms of structure, elicited affect and perhaps consistent association of verbal correlates). And the problems of translating the complex audio signals that are taken to constitute music into symbolic form are still manifold, with only limited progress having been made on many fronts.

The contexts within which digital approaches have been elaborated have contributed to the operational value of handling music in the digital domain as discrete entities. Over the last 30 years, the transmutation of music into digital audio files that have, over the last dozen years or so, been released into the virtual wilderness of the varied ecosystems sustained by the Internet and its ancillary devices has strengthened the case for conceiving of music as made up of discrete sonic entities with determinate structure and specifiable value. And these developments are at the tip of an ideological and technological iceberg with a considerable time depth in Western cultural history, in which social currents have driven the reification or commodification of music as works with individual, cultural and exchange value (see e.g. Goehr 1994) while at the same time stimulating the development of technologies for music's storage, reproduction and production.

Of course, the digital approaches mentioned above represent only one sector; the areas of composition, performance and improvisation almost of necessity adopt a broader range of perspectives on music. In performance or in an explicitly interactive setting, the ontological primitives implicated in digital musical processes can be much more fluid. Programs such as George Lewis's *Voyager* (some 500,000 lines of Forth code, evolved over several years) provide complex, unpredictable and interacting improvisational contexts for live performers (see Lewis 2000), while systems such as Jonathan Impett's *Meta-Trumpet* (Impett 1996) extend the range of functionality of conventional instruments so as to afford the emergence of innovative sonic and gestural structures in the creation or performance of music. Here, the idea of "the work" is called into question, in line with ways in which musical and ideological developments beyond the digital domain throughout the twentieth century (including the rise of jazz and the spread of avant-garde ideas) have problematised performers', composers' and audiences' understanding of folk-theoretic notions of music. As a consequence, the nature of the "music" that is emergent from human–computer interaction (or even computer–computer interaction) appears open-ended or even undetermined, and perhaps not susceptible to being understood in terms of general principles—and if conceived within something like John Cage's aesthetic of chance, may not be intended to be subject to such principles. However, setting the notion of music as interaction in a broader cultural context and exploring it within a generalisable frame should provide a means to address computationally the issues raised by recent interactive approaches to computer music—and indeed, by the broader notion of music as an interactive, communicative medium.

2 Ethnomusicology and music as interaction

One development that has lent focus to the problematisation of Western folk-theoretic notions of music is the growth and consolidation of ethnomusicology as a discipline (as distinct from the assimilation of "world music" into pre-established Western taxonomies; see Stokes 2003). This process has led to an increasing awareness within ethnomusicology (and beyond) of musical ontologies that stand in sharp distinction to those of the West. Of primary importance to the current argument is the fact that in contrast to the *presentational* mode that predominates in Western musical thought and practice, much music in other cultures is *participatory*, involving ordinary culture-members—musical non-specialists—in active participation in performing and creating music. This distinction is perhaps most coherently elaborated by Thomas Turino (2003, 2008), who notes that presentational and participatory musics tend to fulfil different functions and to exhibit different sound features and performance practices, although both modes of music coexist within complex contemporary societies (Finnegan 1989).

Not only is music in many societies a social practice rather than a commodity, people make music together for reasons other than to give pleasure to others. Frequently, participatory music-making is an integral component of ritual; typically, it involves singing and dancing as well as playing instruments; and it is often not an end in itself but a means to an end—it has instrumental value for the participants. It is found in all world cultures, usually in the form of song, though the significance accorded to participatory music-making varies considerably from culture to culture. For some—quite disparate—cultures, it can be one of the most important elements of social life (see e.g. Nettl 1967; Lewis 2002); alternatively, it can be conceived of as primarily recreational and optional (as appears to be the case in many contemporary Western societies—see McLucas 2010).

In all cultures, participatory music-making appears to enhance social bonding (see Nettl 2005, p. 253) and exhibits musical features that support this function. As Turino notes (2008, p. 59), it tends to be based around short, open, redundantly repeated forms with "feathered" (rather than sharply defined) beginnings and endings and to

involve intensive variation in which individual virtuosity is downplayed; it is usually highly repetitive, with few dramatic contrasts, and is grounded in rhythmic constancy, often involving dense textures. In all these respects, its features stand in opposition to those of presentational musics, which usually adhere to closed, scripted forms with organised beginnings and endings and involve extensive variation and individual virtuosity, and a balance between repetition and contrast expressed over a varied rhythmic context. Presentational music's observance of closed forms is one of the primary attributes that allow it to be represented digitally as a collection of decomposable objects. However, in the participatory domain, a "piece of music" is not so much a closed form as (ibid.) "a collection of resources refashioned anew in each performance like the form, rules, and practiced moves of a game", with much in common with the processes and products emerging from interactive computer music systems.

Nevertheless, the ways in which interactive computer musics have been conceptualised and analysed have tended to align themselves with frameworks for conceptualising and analysing presentational musics (see e.g. Borchers 2001; Drummond 2009)—unsurprising, as the contexts in which most interactive computer music systems have been developed are those of the academy or the concert hall, primary habitats for presentational musics in the present day. Although interactive computer music systems have moved well beyond these habitats (often in the form of laptop systems, e.g. Prior 2008), here too they still tend to be conceptualised in terms that are probably more appropriate to those employed to frame the presentational and commodified musics of the last two centuries of Western tradition (see e.g. Grossmann's (2008) critique of the notion of "autonomous" laptop music).

In order for digital approaches to music to reflect the richness and complexity of music as it is manifested both across and within cultures, there is a need to develop frameworks that deal adequately with music in both presentational and participatory guises. This requires identification of general features of the types of interaction that may underpin music as a participatory medium, as well as identification of the generic features of music that may be functional in interactive contexts. It also requires consideration of the ways in which musical interactions share attributes and mechanisms with other modes of human communicative interaction, in particular, with speech.

3 Music and speech as communicative media

From an ethnomusicological perspective, interactive musics are generally simple in structure but have a potent role in mobilising a sense of social cohesion. While few ethnomusicologists have gone so far as Nettl in claiming a universal and primarily social functionality for music, other researchers exploring music within a range of scientific disciplines have suggested that music possesses features that make it particularly effective in scaffolding social interaction. Music typically exhibits temporal regularity that enables interacting individuals to anticipate the actions of, and to entrain their attention and their behaviours to, each other (Large and Jones 1999; Clayton et al. 2005); inter-participant entrainment in music is likely to lead to an enhanced sense of mutual affiliation. Music, across cultures, also *means* in ways that appear paradoxical but that may aid social cohesion. The idea that music embodies "natural", direct or unmediated meaning for participants is found in many societies (Feld 1981; Leman 1992; Turino 1999), albeit embedded in different ontologies in different cultures. However, at the same time, music's meanings appear to be manifold and unresolvable (see e.g. Qureshi 1987). As Swain (1996, p. 135) puts it, "…music seems full of meaning to ordinary and often extraordinary listeners, yet no community of listeners can agree among themselves with any precision that comes close to natural language about the nature of that meaning", an attribute that I have described elsewhere (Cross 1999) as "floating intentionality". This paradox—that music appears to embody unmediated, direct meaning, but what any particular instance of music may mean seems different in the experience of different individuals—can be dealt with by the realisation that the meanings elicited by music are *not* required to be made mutually explicit by individuals interacting in music. Each interacting individual can thus interpret musical meanings more or less entirely idiosyncratically without necessarily coming into conflict with the interpretations of others,[1] a situation that seems to stand in direct opposition to that manifested in language where most speech acts require a degree of consensual referentiality between participants in order to be interactively efficacious.

Hence, music provides a minimally conflictual framework for ostensibly communicative interaction; its seemingly direct expression of meaning, together with the affiliative qualities that derive from its temporal regularity,

[1] An exception can be found in those situations where music is co-opted as a means for the assertion of within-group identity that may be directed at out-groups; a common example in western European societies is the use of chanting by rival groups of football fans. Here, it could be argued that football chants are "musical" only insofar as they serve to coordinate collective *verbal* behaviour that is directed aggressively towards opposing supporters. However, some chants do co-opt existing songs which may become emblematic for particular groups of supporters and these may come to be directed towards bonding with fellow supporters as much as they are intended as a hostile group display directed towards fans of rival clubs (as in the use of the song "You'll never walk alone" by fans of Liverpool F. C.).

affords participants the sense that their experiences are in alignment even while the meanings that each is attributing to a joint musical activity may diverge widely. In a very generic sense, making music together can thus be conceptualised not so much as an aesthetic act but more as a process of establishing and sustaining a sense of interrelatedness between participants. As an interactive medium, music's proximal functions appear more directed towards managing the relationships between participants (see e.g. Turino 2008) than towards goals extrinsic to those relationships, again in apparent contrast with speech. In my own recent work (see e.g. Cross 2011), I have suggested that music may best be thought of as *a communicative medium that is optimal for the management of situations of social uncertainty*; music is, at root, an excellent means of coordinating social attitudes and behaviours and can be viewed as complementary to and coextensive in its forms, structures and primitives with speech as an interactive medium.

Of course, speech can also be employed to coordinate social attitudes and behaviours, by means of its *phatic* dimension, concerned with the mutual recognition of each other's presence by interlocutors in conversational contexts. The term "phatic", introduced by Malinowski (1923), is often conceived of as applying only to the function of formulaic phrases such as greetings at the outset of an interaction. To give one of Malinowski's own examples, in meeting someone and saying the phrase "Nice day today", one is producing a cue that invites a similarly formulaic response from an interlocutor (whether in the form of a nod, grunt, an interjection such as *uh-huh* or an explicit agreement). One is also articulating a proposition that purports to represent a state of affairs in the world that is susceptible to debate, and though this construal may be subsidiary, most instances of phatic talk are intrinsically multifunctional. Cue and response together constitute a mutual act of social recognition of each other as potentially communicative individuals. However, the phatic dimension permeates conversational interaction and is not limited to formulaic opening or closing phrases; the phatic dimension is better interpreted as applying to those elements of a conversation that are concerned with both establishing *and* maintaining its social context—with the *relational* dimension of the conversation—rather than with any referents or goals that are extrinsic to it (see Coupland et al. 1992). Coupland and Jaworski (2003) suggest that the phatic dimension of linguistic interaction is best characterised as employing ritualised sequences, being oriented towards the strengthening of relational ties, and involving a "low commitment to veracity"; phatic talk tends to the formulaic, manages the social context of a conversation and does so in part by disregarding the literal meanings of utterances.

The study of interaction in language has tended not to devote too much attention to the phatic dimension; while it is typically recognised as present, in the vast majority of the literature the focus has been on the ways in which speech—language in action—effects instrumental ends, particularly in coordinating joint action (see e.g. Bangerter and Clark 2003). This aspect of speech—termed *transactional* by Coupland et al. (1992)—has been analysed in respect of both the temporal patterning and semantic structure of conversational turns as well as the relationships between the content of a speaker's discourse and any "back-channel" responses (interjections or brief statements) that their interlocutor may feed back to them to indicate understanding or to stimulate continuation on the part of the speaker. While back-channel responses have typically been interpreted and explored as functional within speech's transactional dimension, some studies—e.g. that of Kita and Ide (2007)—have shown that the nature and function of verbal back-channel responses are susceptible to wide cultural variability and that, at least in Japanese culture, back-channel responses of the type termed *aizuchi* are best interpreted as signs of "emotional support for the turn-holder [speaker]" (ibid., p. 1244) and hence fulfil phatic functions. It seems highly likely that the functions fulfilled by *aizuchi* in Japanese culture are fulfilled by a range of means—verbal, gestural, postural, etc.—in other cultures' communicative interactions. Indeed, Stivers (2008) notes that listeners' nods as they interact with someone who is recounting a story are functionally affiliative (and hence, in the sense outlined above, relational or phatic), being best interpreted as endorsements of the teller's perspective on the story that is being told; by contrast, in the same situation, listeners' vocal interjections are best interpreted as acknowledgements of the information provided by the teller (and hence fundamentally transactional).

Of course, the phatic aspect of language is rarely clearly distinct from other, transactional, aspects. As Senft (2009, p. 230) notes, "The observation that there is generally more behind an utterance which is said to serve only a phatic function, also holds for all of the rather few studies that explicitly deal with the concept of 'phatic communion'". Typically, features of a verbal interaction that are oriented towards establishing and maintaining mutual social recognition simultaneously bear functions that are (Laver 1975, p. 236) "relevant to structuring the interactional consensus of the present and future encounters". In other words, relational aspects of verbal interactions are likely to fulfil dual functions: (i) setting up and consolidating mutual recognition amongst participants of their communicative engagement and (ii) providing a framework that can support each participant in ensuring that their individual and (assumed) joint conversational goals are achieved.

Moreover, as Sidnell (2009, p. 135) points out, there is "ample evidence that gesture, gaze, and body orientation" are all involved in (ibid., p. 125) the "little world of shared attention and involvement" that characterises talk-in-interaction, and such non-verbal aspects of communicative involvement are likely to be crucially important in all dimensions of communicative interaction. Hence, relational and transactional dimensions are intertwined in speech interaction and are manifested not only in speech but also in "functional gesture" (ibid.); to adapt Laver's words (1975, p. 217), "…the fundamental social function of the multidimensional communicative behavior that accompanies and includes phatic communion is the detailed management of interpersonal relationships during the psychologically crucial margins of interaction".

Bearing these ideas in mind, we can propose that music can be interpreted as functional in the phatic dimension. Indeed, in contrast to the situation in speech, in music the phatic or relational dimension is foregrounded, the transactional dimension being more or less absent. Yet music—even in the form of interactive music-making—and speech seem so different that the suggestion that they may share functional characteristics requires further substantiation. In interactive music-making, participants produce sound simultaneously and are likely to be producing overlearned patterns, while in speech participants take turns and are generating utterances on the fly. Moreover, interlocutors can organise joint action by virtue of language's powers of referential specificity —speech can have an unambiguously *transactional* dimension—while people making music together are likely to experience a mutual *sense* of shared purpose, thanks to music's affiliative nature and its lack of requirement for referential consensus—the elements of its *relational* powers.

Nevertheless, as we have just seen, much of speech is concerned not with effecting instrumental ends extrinsic to the ongoing linguistic interaction, but with developing and sustaining the social relationships that constitute the framework for conversational participation—like interaction in music, the relational dimension makes up a significant aspect of speech interaction. Moreover, the ability to anticipate others' actions that is critically important to music is also central in speech, in managing the organisation of conversational turn-taking (see Levinson 2006). Speech, like music, makes use of overlearned formulae (as in greetings), while music, like speech, can exhibit on-the-fly generativity (as in joint improvisational performance). Music, like speech, may involve turn-taking (as in "call-and-response" structures, or "lining-out"), while speech can involve simultaneity of utterance—intriguingly, in Japanese *aizuchi* (see above), fulfilling a phatic function. And in its employment of the phatic, relational, dimension, speech, like music, can motivate a mutual sense of shared purpose, while interactive music, often embedded in broader ritual, can effect joint action just as does speech. Speech and music are not so distinct as interactive, communicative media as might at first appear; indeed, in many societies, the clean distinction drawn in contemporary Western societies between language and music is much more difficult to discern (see e.g. Seeger 1987).

Rather than constituting discrete domains, music and speech are perhaps better conceptualised as opposing poles on the continuum of human communicative resources in terms of function. While speech is optimal for mobilising joint action by virtue of language's powers of referential specificity, music is optimal for motivating a sense of shared intentionality (cf Tomasello et al. 2005) because of its provision of an explicit framework for "sharing time" and its inexplicitness in respect of meaning. In speech, we articulate and mutually demonstrate understanding of the propositional content of utterances (the transactional dimension); we also manage relationships with interlocutors in speech (the relational dimension), but the inexplicitness, or lack of veracity, that Coupland and Jaworski (2003) suggest is requisite for this function can always be undermined by the potential for our utterances to be interpreted not as tokens of recognition of each other's communicative presence, but as definite statements about the world that are capable of being contested. In music, we cannot formulate or convey semantically decomposable propositions. But music has the advantage over language in the relational domain in that music sets up and maintains its affiliative, relational, frame, without its affiliative qualities having to be continually re-negotiated, and the individual significances that participants may attribute to the ongoing musical interaction are not required to be made mutually manifest in order for the interaction to be sustained and to succeed.

4 Formal (or formalisable) frameworks for analysing or specifying computer music interaction

Returning to the initial problem—that treatment of music in the digital domain has tended to be in terms of discrete objects and has not reflected its manifestations as a process of interaction—the question arises of how one might represent music in digital contexts in ways that capture its interactive attributes. How can one produce any *formalisable* account of music in interactive computer systems that might be applied computationally or even analytically across systems? The ethnomusicological literature that demonstrates, and indeed insists on, music's status as a mode of interaction offers few helpful paths. Its characterisation of music as participatory tends to focus on specific ethnographic examples rather than seeking to develop

generalisable frameworks that may be applied across cultures for understanding music as participation. Those few cases from that literature that do adopt a more universalising stance in describing attributes of music as a participatory medium (e.g. Lomax 1968; Turino 2008) tend to depict features that are better interpreted as more diagnostic than constitutive. The characteristics that they pinpoint as distinguishing musical activities that are participatory from those that are presentational (Turino 2008), or those that are "groupy and integrated" from those that are "individualized and little integrated" (Lomax 1968, p. 22) are sufficient to enable musical activities to be so categorised, but are not articulated with enough specificity to enable the principles of categorisation to be formalised.

The literatures that describe, analyse and critique music in interactive digital contexts (see e.g. Kim et al. 2011) also tend to focus on specific instances and do not generally aim to provide generalisable frameworks in terms of which their subjects can be understood—unsurprising, as these literatures generally deal with interactive music systems that are conceived of and explained as aesthetic constructs of which the effectiveness can only be judged in ways that are largely *sui generis*. Two recent papers, those by Wilkie et al. (2011) and by Murray-Rust and Smaill (2011), constitute exceptions to this general trend in aiming to provide generic frameworks within which human–computer musical interaction can be conceptualised and explored. Wilkie et al. (2011) suggest that Lakoff and Johnson's (2003) notion of image schema or conceptual metaphor can usefully be employed to understand the bases of human–computer musical interaction; however, they provide only a preliminary sketch of how these concepts might be applied in practice to enhance the development of software for producing music, and are not concerned with real-time issues. Murray-Rust and Smaill (2011), on the other hand, go beyond these general considerations in presenting what is intended as a comprehensive and generic model for analysing and implementing multiagent interactive musical systems, and I shall consider this in detail.

The model presented by Murray-Rust and Smaill (2011) treats music not in terms of finite pieces but, in Turino's (2008, p. 59) words, "a collection of resources refashioned anew in each performance". It constitutes an intriguing attempt to produce a generic model of musical interaction which has many strengths, though it stands somewhat at odds with the general account of music in interaction developed above. It explores (p. 1697) "the way in which the playing of one musical agent is related to that of another", aiming "to focus on the communicative aspects of the musical experience rather than those which can be directly derived from the musical surface". They start by outlining what is intended by an action in the context of playing music and then develop a framework that relates actions to each other. Their approach is guided by ideas of communicative action grounded in speech act theory, which they characterise (p. 1698) as "… 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". Speech act theory, or the theory of illocutionary acts, presents an account of the ways in which linguistic utterances may be used in context to effect particular ends. Illocutionary acts can be categorised after Searle (1976) as *representatives* (which commit a speaker to something being the case); *directives* (attempts by a speaker to get a hearer to do something); *commissives* (which commit a speaker to some future course of action); *expressives* (the expression of a psychological state relative to the content of an utterance); and *declarations* (which (ibid., p. 14) "bring about some alternation in the status or condition of the referred-to object or objects solely in virtue of the fact that the declaration has been successfully performed").

Murray-Rust and Smaill develop an analogous theory of musical acts, which characterises the relationships between the actions of musical agents and assigns these relationships to particular classes of musical performatives. They state that musical acts must have the qualities of *embodiment* (p. 1699, "musical acts must have a manifestation in music"), *intention* (ibid., "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") and *intelligibility* (ibid., " if it is not understood, then it will fail to change the world"). They explicitly ignore "extramusical actions" such as nods, glances and gestures, focusing (p. 1700) on a "musical surface which has been segmented into discrete events such as notes, percussive strikes, glissandi, trills, etc." This is justifiable on the grounds of principle (imposing the quality of embodiment on the musical act means that the acts that produce music must be recoverable from the musical surface or signal), as well as heuristically, in that computational means of treating music in terms of entities such as notes, etc., are well established. However, as we shall see, it does create some problems for developing a computational understanding of music in interaction.

They present an account of the derivation of descriptions, which they define as "essentially perceptual objects", from the musical surface, expressing descriptions in terms of values. In the examples given, these are largely music-theoretic: e.g. a particular configuration of notes may be given the value "Cm" (the minor triad on the note C). They categorise the resulting value-representations in terms of the relationships that may exist between pairs of values which may be simultaneous or successive, although they generally treat the latter type of relation (e.g. in the

context of accounting for the relationships that are evident when musical agent A's act is followed by that of agent B). These relationships (*Rel*) are of five types: *same* (two successive values are identical), *subsumed* (the first value constitutes a specialisation of the second), *subsumes* (the second value is a specialisation of the first), *alter* (the two values have some, but not all, elements in common) and *disjoint* (the two values are different).

Underpinning their model are concepts of *musical context* and *musical common ground*; musical context provides the basis for producing an account of musical interactions as a series of states that can be interrogated as conjunctions of both sequential and simultaneous events, while musical common ground is defined as (p. 1706) " the set of values which an agent reasonably believes to have been extracted by every other agent". They develop the idea of an "action signature", which they describe (p. 1707) as "a compact, transferable representation of the components of the action which are useful from the point of view of analysing interactions". Each action signature is a triple of relations between the acts of two musical agents that describes the relationships between the prior acts of each and the new act of one within the five-term *Rel* framework sketched above. Hence, the possible values of *Rel* appear as:

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

and the expression of an action signature as:

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

where R_{self} is the relation between self's new and old values, R_{other} is the relation between self's new playing and other's current playing and R_{prev} is the relation that held between self's old value and other's current value.

They then propose (p. 1710) "a semantics for the conditions of expressing a musical act", expressed in terms of a set of performatives and their formulation in terms of musical action signatures. Their taxonomy of musical performatives (analogous to Searle's classification of illocutionary acts) comprises *propose* (initiating a musical act); *confirm* (accepting an idea proposed by another; hence, a relationship *Rel* will be "same" or "subsumed"); *reject* (not accepting a proposal; hence, *Rel* is "disjoint" or "alter"); *extend* (extending currently accepted material, with *Rel* being "subsumed"); *alter*; *argue* (independently producing musical acts without any agent accepting any other's; hence, *Rel* is alter or disjoint); and *request* (which they describe as "a conventionalised action that cannot be modelled with action signatures"). They note that some musical acts could be realisable as different performatives and that in order for musical acts to be computable, their approach would require extension in the form of a model of intentional musical behaviour.

Murray-Rust and Smaill's model provides a complex and nuanced account of the exigencies of musical interaction in formal terms. It has substantive value as a means of analysing musical interactions in formal terms (as they demonstrate in an appendix), as well as in having created a framework within which the functional characteristics of interactive computer music systems may be understood in comparative terms. Nevertheless, as should be evident, their model of musical interaction is somewhat at odds with the account developed in the preceding section of this paper. This derives, in part, from their use of speech act theory to underpin their approach. While speech act theory does indeed, as they note (p. 1698), expand the exploration of language in action well beyond an exclusive focus on the truth values of propositions, it does not provide a safe haven from the problems associable with understanding communicative systems principally in terms of the expression of truth-conditional propositions.

Searle (1976) implies that speech acts always have propositional content, whether that content is a matter for negotiation or can be presupposed (as is the case for his category of *expressives*). Wharton (2003), in considering the boundaries of language by exploring the status of interjections (which he characterises (p. 201) as "paralinguistic" and (p. 211) as communicating "attitudinal information, relating to the emotional or mental state of the speaker"), suggests that for any communicative act to be considered part of language, at least some of its elements must be expressible in terms of (ibid., p. 196) a "…grammar, a code pairing phonological and semantic representations of sentences", and hence foundationally propositional. This is, of course, not a significant issue in the analysis of the vast majority of linguistic interactions, which, after Searle, are likely to manifest some propositional content. However, it does pose problems when applied to music, as it brings with it presumptions as to a need for something analogous to propositional content. In Murray-Rust and Smaill's theory, this seems to exist at the level of the "performatives" that they build from their action signatures. Their terms *propose*, *confirm*, *reject*, *extend*, *argue* and *request* have indeed all been widely applied to music by musicologists and music critics, though generally in metaphorical terms; the functions of these terms as descriptors of goal-oriented linguistic discourse can be interpreted as bringing with their application in Murray-Rust and Smaill's model the undesirable baggage of propositionality.

In some senses, Murray-Rust and Smaill (2011, p. 1711) have only gone part of the way in their approach, as they themselves acknowledge when they note the need to develop "…a model of intentional musical behaviour" in order for their model to yield computable interpretations of

human musical interaction. In their paper, they present a framework for representing objectively quantifiable correlates of the musical interactions based on what they describe as "the musical surface" comprised of (ibid, p. 1700) "discrete events such as notes, percussive strikes, glissandi, trills, etc.", which appear, in their approach, to fulfil the function of propositional content in speech act theory, without aiming to represent the intentions or attitudes in which those correlates are necessarily embedded. This interpretation of their work is supported by their theory of musical acts being concerned with implementing a means of (p. 1699) "reasoning about the musical beliefs of other agents", which seems to imply too heavy a reliance on the propositionally grounded architecture of speech act theory to provide an adequate reflection of the relational dimension of musical interaction outlined above.

These problems can be, to some extent, be resolved by redefining the idea of the musical surface so as to be grounded in music perception and cognition rather than, as in Murray-Rust and Smaill's account, *music-theoretic*. Hence, the musicality of the acts that may be recoverable from the "musical surface" would be linked to engagement, attention, affect, intention and attitude. Conceiving of the musical surface as experiential, as constituted by factors that have been shown to be focal in communicative interaction, and as susceptible to empirical analysis would fulfil their "embodiment" criterion and provide a means of motivating the interpretation of musical interactions. A cognitive-perceptual reconfiguring of the musical surface would take into account the types of phenomena that Lomax (1968) described as "paramusical"—vibrato, mode of sound production, micro-timing—about which much more is now known than in Lomax's day, in terms of the extent to which features such as jitter (see Patel et al. 2011) are indicative of affect, and micro-timing may be informative in respect of intentions as to the articulation of structure (see e.g. Gabrielsson 2009). An approach intended to provide a generalisable account of musical interaction must also, unlike that of Murray-Rust and Smaill, take into account the "nods, glances and gestures" that have been shown to underpin music as interaction (see e.g. Moran 2007), otherwise it would be unable to provide an account of gestural or multimodal interactive computer music systems (see e.g. Lindström et al. 2005; Varni et al. 2009).

5 An integrated framework for formalising the representation of interaction

An alternative approach to that of Murray-Rust and Smaill—or perhaps a complementary approach—can be proposed that would start from the characterisation of real-world communicative interactions in generic terms (to which computer-based interactive systems could be seen as formal approximations). Such an approach might be derived from the work of Jens Allwood and his collaborators (in particular, Allwood et al. 1992, and Allwood 2007), although that work requires some revision in order to be generalisable to the case of interactions that can be characterised as musical.

In the more recent paper, Allwood outlines an approach to understanding communicative interaction that is explicitly intended as a critical response to speech act theory (ibid., p. 13) and as a means of taking into account the context and background of such interactions that goes beyond those that are the typical focus of conventional conversation analysis (CA). He puts forward a method of investigating communicative interaction that is based on the idea of individual *contributions* to the interaction (that may be successive or simultaneous) and the functions that these might fulfil in the unfolding communicative context. The term "contribution" is preferred by Allwood to the term "turn" that is typically used in CA; as Allwood suggests, the "turn" can simply be viewed as a special case of the interaction class "contribution", when one contributor holds the floor. Contributions may fulfil message or communication management functions; in speech, "The main message (MM) is related to the main communicative acts and their associated cognitive attitudes and referential content of the contribution" (Allwood 2007, p. 8), while communication management functions are either other- or self-directed (for Allwood, interactive versus own communication management, ICM vs OCM). Self-directed functions of contributions (OCMs) include holding a turn while managing the structure, timing and/or role of the contribution (to which one might add the functions of affective and/or motoric self-regulation). Other-directed functions of contributions (ICMs) may be articulated in terms of Allwood's responsive, evocative and expressive functions, all of which may incorporate elements aimed at making evident to the other participant the contributor's willingness to engage communicatively and their construal of the joint focus of the interaction.

The scope of the "main message" (MM) function in speech interaction is likely to differ from that in musical interaction. In speech, whether or not it is evident in any given instance, the MM function is likely to be dependent on the referential content of the contribution. In music, this is unlikely to be the case. But contributions in both domains are likely to be comprehensible in terms of what Allwood describes as the *responsive* and *evocative* functions, the *responsive* function of a contribution being its "relation to preceding or turn-holding discourse", while the *evocative* function of the contribution is the response it is designed to elicit, often assessable through its "relation

to succeeding discourse". In addition, Allwood's *expressive* function is likely to be applicable in both domains, being concerned with the articulation of [cognitive or affective] attitudes and emotions.

Having characterised the contributions to a communicative interaction as individual acts, it remains to provide an explicit account of their inter-relationships. Casting back to Allwood's earlier work (Allwood et al. 1992), we find a framework that explicitly takes as its primitives pairs of contributions. As Allwood et al. note (ibid., p. 2), they are concerned with specifying and exploring "linguistic mechanisms which enable the participants of a conversation to exchange information about four basic communicative functions, which are essential in human direct face-to-face communication". This framework, developed to account for the role of linguistic inter-individual feedback, can be used as a means of characterising in semi-formal terms the elements of the unfolding interaction rather in the manner of Murray-Rust and Smaill's (2011) "action signatures", but here taking into account the communicative intentions underlying the actions of the participants involved in the interaction. Assignment of the functions subserved by individual contributions in a communicative musical interaction after Allwood (2007) can be thought of as setting the parameters for the analysis of the interaction in terms of a version of the framework outlined in Allwood's 1992 paper, in which the basic unit of analysis is not an individual contribution but a successive pair of contributions. As Allwood et al. (ibid, p. 18) put it, each "…[contribution] pair … can be represented in attribute-value terms as a matrix of (semi-nested) binary values…" in respect of four functions: contact, perception, understanding and attitudinal reactions.

This framework is proposed by Allwood et al. in the context of the analysis of talk-in-interaction. In the formulation that they present, it would appear as in Table 1, with the first three functions constituting binary variables and the fourth, attitudinal reactions, taking a number of nominal values depending on the specific functional relationship embodied by the contribution pair.

The framework of Allwood et al. requires some modification to be adapted so as to be appropriate for characterising the ongoing flow of a musical interaction. It is concerned with feedback in linguistic interchange—largely, sequential and responsive—but is adapted here to fit with the simultaneity and bidirectionality of musical interaction. It necessarily has to reflect both transactional and relational aspects of linguistic interchanges, but as noted earlier, the primary dimension motivating musical interactions is the relational. Hence, that aspect of the framework that most directly reflects the transactional dimension—(iii) UNDERSTANDING—would need to be construed as being bound to particular stylistic norms and expectations, as well as being of less moment or weight than those aspects reflecting the relational (here, interpreted as (i) CONTACT, (ii) PERCEPTION and (iv) ATTITUDINAL), largely because its conditions can be assumed to be met in a musical interchange almost by the mere fact of sustained interaction. (The opposite can be expected to hold for the speech domain, in that the transactional dimension would be likely to be of most moment and would be generalisable—at least via translation—as well as being constrained by the specific facts of the context and motivations of the interlocutors in respect of those facts, for at least most utterance contexts.) Moreover, while the background assumption underlying the framework of Allwood et al. is turn-taking (i.e. sequential contributions to the interaction), the background assumption in characterising musical interactions would need to be simultaneity (i.e. participants typically performing together in time rather than one after another and potentially playing equal roles within the interaction).

The framework would be applied to the actions of the individuals engaged in the interaction, actions here being defined as "musical gestures" rather than as discrete music-theoretically conformant events (as appears to be the case in Murray-Rust and Smaill's approach), so as to provide a more generalisable system that may be applied across a range of musical styles or cultures. A musical gesture can be defined, for present purposes, as an objectively specifiable phrasal or sub-phrasal, or metrical or sub-metrical, unit, typically occurring within a temporal

Table 1 After Allwood et al. (1992): "four basic communicative functions, which are essential in human direct face-to-face communication"

SPEECH	
FUNCTION	VALUE
(i) CONTACT—willingness and ability to continue interaction	+/−
(ii) PERCEPTION—willingness and ability to perceive expression and message	+/−
(iii) UNDERSTANDING—willingness and ability to understand expression and message	+/−
(iv) (other) ATTITUDINAL reactions—willingness and ability to give (other) attitudinal reactions to expression, message or interlocutor	Accept, reject, belief, agreement, surprise

Each function is assigned one of the available values so as to yield an attribute-value structure which describes (ibid., p. 16) "how the occurrence content of a particular feedback utterance is constructed by combining a type content with features of the context". As Allwood et al. note, "Category (iv) has the word *other* in brackets, since contact, perception, and understanding also involve attitudes, albeit of a very fundamental cognitive and volitional sort"

window of between 1 and 3 s. The concept of "musical gesture" here conforms with Godøy's (2011, p. 18) notion of a *sound-action chunk*, the level of musical organisation at which we find "significant style-determining features of musical sound such as rhythmical and textural patterns… as well as the associated sense of body motion and even mood and emotion". The idea that events at this timescale should constitute the basic units for the analysis of interaction—at least in the first instance—is supported by Pöppel's (2009) proposal that underlying perception—and action (see e.g. Gerstner and Goldberg 1994; Schleidt and Kien 1997; Lemke and Schleidt 1999)—are processes of pre-semantic temporal integration that have an amodal time constant of around 2–3 s, within which component events may be integrated so as to allow for (Pöppel 2009, p. 1893) "maintenance of perceptual… identity". A "musical gesture" may, of course, still take the form of a discrete note or chord, but is more likely to articulate a pattern of notes or chords that can be partitioned from preceding or succeeding patterns by the identification of discontinuities (analogous to the procedures underlying Lerdahl and Jackendoff's (1983) grouping principles).

In adapting the framework so as to enhance the focus on representing the relational dimension, each of the first three components should preferably be scalar rather than binary, in order to reflect the potential for a range of levels of engagement, while the fourth component, ATTITUDINAL, would need to be composite, with a range of scalar subcomponents. In addition, the first three functions would need to refer not only to perception/reaction, but also to production. This adaptation affords the structure shown in Table 2.

The use of scalar rather than binary values in the version adapted for music is motivated by a basic assumption of a focus on the relational dimension, in the context of which fine-grained interpersonal contributions and responses are required to be representable (the scale used here can be thought of as probabilistic, from 0 to 1). In its original use, the framework is intended as a means of characterising the pragmatic and semantic functions of linguistic feedback in communicative interactions, with the semantic dimension largely articulated in terms of functions (iii), UNDERSTANDING, and (iv), ATTITUDINAL, contributions. Adapted as a means of characterising music in interaction, the presence of an explicitly semantic dimension that can be linked to a "message" becomes moot; however, function (iii), UNDERSTANDING, is still required to deal with those features of musical interaction that arise from shared cultural intuitions and values and that underpin musical syntax. UNDERSTANDING in the context of speech interactions is mediated by common ground, the "mutual knowledge, mutual beliefs, and mutual assumptions" that allow participants in a communicative interaction "to coordinate on content" (Clark and Brennan 1991, p. 128). Common ground is established by interaction and is built on assumed commonalities of knowledge and belief (Lee 2001), being at least partly oriented towards the interchange of transactional meanings involving propositionality and truth-conditionality that must be underpinned by a consensus between the parties about the informational content of the interaction. This consensus is one of the features that enables linguistic interaction to unfold coherently in time in order to effect joint action.

Table 2 Framework for the analysis of musical interaction, adapted from Allwood et al. (1992)

MUSIC	
FUNCTION	VALUE
(i) ENGAGEMENT—willingness and ability to continue interaction	(0–1) SELF-DIRECTED, OCM/OTHER-DIRECTED, ICM
(ii) ATTENTION AND MOTIVATION—willingness and ability to perceive and produce expression and signal	(0–1)
(iii) UNDERSTANDING—willingness and ability to understand and act on expression and style-specific features of signal	(0–1)
(iv) ATTITUDINAL contributions—willingness and ability to give attitudinal contributions to expression, signal or co-participant	ALIGN [Initiate$_{(0–1)}$, match$_{(0–1)}$, complement$_{(0–1)}$, support$_{(0–1)}$, close$_{(0–1)}$] REALIGN [Remodel$_{(0–1)}$, contest$_{(0–1)}$, disregard$_{(0–1)}$]

In music, propositionality is not a key property, but music in interaction can unfold coherently over time, exhibiting principles that can at least be thought of as syntactical in organisation. Musical "syntax" can be thought of as a form of scaffolding that shapes musical interaction, making clear what are legitimate as participant contributions to the interaction while delegitimising others. There is thus a need for the adapted framework to represent those properties that enable this coherence of ongoing musical interaction, and it is proposed that the most appropriate locus for such features is in the expressive and style-specific features of the musical event. Hence, while musical interaction, like linguistic interaction, rests on a sort of common ground, it does not have the overt commitment to propositional content that can characterise language; music's common ground is established by interaction built on assumed commonalities of stylistic competence or cultural knowledge that are likely to be experienced reflexively—intuitively—rather than being the

object of any conscious reflection[2] (except in cases of breakdown).

Communicative function (i), ENGAGEMENT—willingness and ability to continue, initiate, or re-initiate interaction—would be assessable in terms of the attentional focus of each participant (in gesture-based or real-world systems, derived from, for example, posture or gaze direction), which should yield inferences concerning the extent to which a contribution is self-directed (OCM) or other-directed (ICM). Function (ii), ATTENTION AND MOTIVATION—willingness and ability to perceive expression and signal—would be measurable in terms of participants' responsiveness to the musical signal (the acoustic, behavioural and visual correlates of a musical gesture that give it an identity that is independent of the context in which it is realised) and its expression (the manner in which it is deployed), the level of responsiveness providing an index of level of arousal. Function (iii), UNDERSTANDING—willingness and ability to understand expression and style-specific features of signal—would be quantifiable in terms of the momentary and longer-term style-specific appropriateness of responses and initiatory "gestures" to their stylistic contexts. Communicative function (iv), ATTITUDINAL contributions—willingness and ability to give other attitudinal contributions and reactions to expression, signal, or co-participant—would be assessable by evaluating the function of a contribution in the context of the ongoing interaction. Two main attitudinal components are proposed, ALIGN and REALIGN, selected as the most appropriate categories for characterising attitudinal aspects of ongoing musical interaction.[3] The term ALIGN is preferred here to the term COOPERATE, as a base-level assumption that can be held to underlie musical interaction is cooperativity; the extent to which OCM or ICM are evidenced the context of function (i), ENGAGEMENT, may be taken as an index of cooperativity. Each attitudinal component can be split into subcomponents. In the case of ALIGN, a contribution may *initiate* an interaction (providing an invitation to interact), *match* with an existing contribution (by coordinating with it), *complement* another's contribution (by combining with it), *support* the other's contribution (by adopting a subordinate yet sustaining role) or in appropriate circumstances bring the interaction (or phase of the interaction) to a *close*. In the case of REALIGN, a contribution may derive from another's contribution yet *remodel* it so as to change its style-specific import; it may actively *contest* an existing contribution (e.g. by overtly failing to conform to fundamental and mutually manifest attributes such as tempo); or it may simply ignore or *disregard* existing contributions.

This framework is intended to be applicable to the analysis of any form of interaction that can be understood as musical. Its applicability is explored here in the context of an unscripted, non-expert musical interaction recorded as one of the pilot sessions undertaken in the course of developing an ongoing study of naturalistic unscripted interactions in speech and music (for interim results, see Hawkins et al. 2013). In these pilot sessions, conducted in the University of Cambridge's Cambridge Centre for Music and Science, participant dyads (friends, typically students at the university, with both either musically expert or non-expert) were seated in a recording studio at a round table on which was placed several collaborative games (such as a pack of cards) and a number of non-standard musical instruments (such as a Western version of a kalimba, or thumb piano, Australian aboriginal clap-sticks, three Sri Lankan drums, etc.). Participants were asked to start by discussing how they had arrived at the studio for about 10 min, then either to play with the collaborative games or jointly explore the musical instruments for 10 min, to engage in whichever of these two options they had not yet undertaken for 10 min and for the final 10 min, to make up a story together. Participants were recorded on two video cameras with integrated high-quality microphones throughout the experimental sessions.

While the main purpose of the pilot sessions was to identify an appropriate methodology for studying unscripted musical and speech interaction (and the final method used in Hawkins, Cross and Ogden differs in detail from that outlined above), pilot data revealed some interesting and novel facets of spontaneous musical interaction. The musical bouts produced by the participants varied in quality and in degree of success; however, all musician and non-musician dyads produced some bouts of fluent musical interaction, though that fluency and coherence was sometimes evident only for short stretches of an attempted musical interaction. A brief sequence of musical interaction from one of the pilot sessions is used here to explore the viability of aspects of the framework outlined above as a means of coding and understanding attitudinal aspects of musical interaction.

In this sequence, two musically non-expert female graduate students are participating. At the beginning of the sequence, one participant (identified as L) decides that she

[2] This notion of musical "common ground" differs somewhat from that proposed by Murray-Rust and Smaill (2011, p. 1706) who define "…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 definition of common ground seems to require a more exhaustive and explicit awareness of the "cultural context" than appears likely to be operational in any real-world musical interaction, though that comprehensiveness and particularity may be required by the specification of a particular computational system for human–computer musical interaction.

[3] Given the intrinsically cooperative nature of musical interaction, what would appear to be the antonym of ALIGN—disalign—is rejected here in favour of REALIGN; disalignment would be indicative of a refusal to engage or to continue to participate in a musical interaction.

Table 3 Analysis of unscripted musical interaction

Start	End	P	Event	Behaviour	Interpretations
21.920	*22.370*	*L*	*Action*	*Both hand lift prior to lap drum strike*	*L preparation to explore drums*
21.938	*23.239*	*L*	*Attentional focus*	*Gaze at table-top drum (rh)*	
21.958	22.333	R	Gaze at eyes	Gaze at L eyes	R assessment of L musical intentions
22.336	23.241	R	Attentional focus	Focus on L hands	
22.507	*23.214*	*L*	*Action*	*Bimanual pattern, lapRH-lapLH-lapRH_tableRH [pattern should have started with lapHL, adjusted at 23.214]*	*L sets up and repeats simple bimanual rhythmic pattern*
23.214	*32.928*	*L*	*Action*	*Fluent bimanual pattern repeat starting with three-note upbeat (lapLH-lapRH-lapLH)-tableRH and continuing through to end of bout*	
23.226	*24.706*	*L*	*Gaze at eyes*	*Gaze at R eyes*	*L assessment of R musical intentions*
23.220	25.360	R	Action	Move hands to kalimba	R joins mutual gaze, checking appropriateness of own behaviour
23.228	23.583	R	Gaze at eyes	Gaze at L eyes	
23.583	25.994	R	Attentional focus	Shift of focus to own hands on kalimba	
24.705	*30.130*	*L*	*Attentional focus*	*Gaze at table-top drum (RH)*	*L satisfied musical roles agreed*
25.360	28.460	R	Action	Bimanual exploration of kalimba, starting in time with drum but getting out as focus becomes stronger on self-management of kalimba	R explores kalimba—OCM
25.983	26.646	R	Gaze at eyes	Gaze at L eyes	R checking appropriateness of own musical behaviour—ICM[¶]
26.642	29.423	R	Attentional focus	Focus on own hands on kalimba	R focusing on OCM
29.421	30.013	R	Gaze at eyes	Gaze at L eyes	R checking appropriateness of own musical behaviour—ICM
28.460	29.890	R	Action	Comes back in time with drum	R oscillating between OCM and ICM
29.895	32.545	R	Action	Bimanual exploration of kalimba getting out of time with L	
30.018	39.004	R	Attentional focus	FOCUS on own hands on kalimba	
30.130	*30.555*	*L*	*Gaze at eyes*	*Gaze at R eyes*	*L assessment of R musical intentions*
30.561	*35.698*	*L*	*Attentional focus*	*Gaze at table-top drum (RH)*	
32.545	**38.935**	**R**	**Action**	**Exploration continues, eventually strikes note in time with drum downbeat, and proceeds to develop a coherent kalimba pattern at times subdividing drumbeat**	**At 34.021, a moment of synchrony when R plays kalimba simultaneously with L's drumbeats and develops short melodic pattern, several strokes synchronous with drumbeats**
32.928	***39.254***	***L***	***Action***	***Adjusts timing of tableRH stroke to align with kalimba stroke, continuation of bimanual (lapLH-lapRH-lapLH)-tableRH pattern repeat***	
35.684	***36.931***	***L***	***Gaze at eyes***	***Gaze at R eyes***	***L noting appropriateness of musical interaction—ICM***
36.937	***39.663***	***L***	***Attentional focus***	***Gaze at table-top drum (RH)***	
39.001	**45.348**	**R**	**Gaze at eyes**	**Gaze at L eyes**	**R acknowledging end of contribution**

Table 3 continued

Start	End	P	Event	Behaviour	Interpretations
39.254	***41.755***	***L***	***Action***	***Bimanual pattern on lap starting with upbeat LH-LH-RH-LH-RH, then final downbeat RH on table drum***	***L acknowledging end of interaction***
39.657	***40.465***	***L***	***Gaze at eyes***	***Gaze at R eyes***	
40.472	***42.376***	***L***	***Attentional focus***	***Gaze at lap drum (both hands)***	
41.755	*44.943*	*L*	*Action*	*Bimanual paradiddle (RH-LH-RH–LH–RH–) on lap drum*	*R laughing and verbalising, commenting self-deprecatingly on success of the musical interaction, L acknowledging success*
42.920	*44.943*	*L*	*Action*	*Bimanual paradiddle (LH-RH-LH–RH–LH–, RH-LH-RH–) on lap drum*	
43.532	*44.943*	*L*	*Gaze at eyes*	*Gaze at R eyes*	*L acknowledging end of bout*

The fluent musical bout is bold-faced

will play the drums; she places a conjoined pair of small drums on her lap and another, smaller, drum on the table. The other participant (identified as R) directs her attention towards the kalimba (an instrument evidently unfamiliar to her) and starts to explore how she can make sounds with it. The account of the sequence given in Table 3 derives from an analysis of the video carried out using ELAN, coding events in terms of *action*, *attentional focus* and *gazing at eyes*; events are coded for both participants, sometimes sequentially and sometimes overlapping in time. Table 4 gives an account of sequence to the end of the musical bout in terms of the coding framework presented in Table 2.

The application of the framework allows tracking of the fluctuating engagement between participants as one or the other directs attention towards managing their own instrument or musical pattern (notably R, as she tries to understand how to produce melodies on the kalimba); it highlights the consistent high levels of attention and motivation of both participants towards the interaction throughout the attempted musical bout; it tracks the fluctuating understanding as each achieves or fails to achieve fluency in what they are attempting to produce; and it tracks the roles and functions of the attitudinal contributions as evidenced in the musical behaviours as the sequence unfolds. Note that when fluency is achieved (albeit for only 7 s or so, between 32.545 and 39.001 s), both participants are coded as having the highest levels of engagement, attention and motivation, and understanding, and both are coded as $\text{ALIGN}_{[\text{complement}]}$ as they mutually adapt to each other's musical behaviours which combine to produce a coherent (though brief) musical sequence. In this short but musically fluent bout, these musically non-expert participants mutually adapted the timing of their playing to maintain fluency, producing sequences of musical events in almost exact synchrony (kalimba notes mostly occurring within 20 ms of a drum strike—or vice versa).

In the approach of Allwood et al. (1992), the production of an attribute-value coding of a communicative interaction constitutes a precursor step to the derivation of a *situation semantics*, which expresses the relationship between a feedback utterance (or back-channel contribution) in conversation and a foregrounded discourse theme. Within the adapted framework outlined above, which characterises the relationships between two interacting individuals whose contributions to the interaction may have equal status within the interaction, the contributions of both participants require to be coded. Moreover, the notion of a situation semantics cannot and should not be transposed directly to the musical case for an obvious reason: the absence of a propositional, consensually semantic system in the musical case. Murray-Rust and Smaill's (2011) *action signature* representation might provide a viable alternative, here operating principally on the attribute-values represented under function (iv), ATTITUDINAL, contributions weighted by a linear combination of the values accorded to functions (i)–(iii).

This would result in a sequential representation of the temporal relationships between the attitudes of the participants in the musical interaction, charting the dynamic flux of the intentions and roles that could be attributed to the participants. As the basic terms that feed into the action signatures reflect attitudinal contributions, the resultant structures would not require to be reinterpreted in terms of performativity (as is the case in Murray-Rust and Smaill's approach), as that performativity—at least, those aspects of it relevant to the relational dimension of musical interaction—would already be encoded in those resultant structures, constituting, in Allwood's (2007) terms, the "main

Table 4 Application of the coding framework to the unscripted musical interaction

Start	End	P	Engagement	Attention and motivation	Understanding	Attitudinal	Interpretations
21.920	*22.370*	*L*	*1*	*0*	*0*	–	*L preparation to explore drums*
21.938	*23.239*	*L*	*1*	*1*	*0*	–	
21.958	22.333	R	1	1	0	–	R assessment of L musical intentions
22.336	23.241	R	1	1	0	–	
22.507	*23.214*	*L*	*1*	*1*	*0.5*	–	*L sets up and repeats simple bimanual rhythmic pattern*
23.214	*32.928*	*L*	*1*	*1*	*1*	*ALIGN, initiate*	
23.226	*24.706*	*L*	*1*	*1*	*1*	–	*L assessment of R musical intentions*
23.220	25.360	R	1	1	0.5	–	R joins mutual gaze, checking appropriateness of own behaviour
23.228	23.583	R	1	1	0.5	–	
23.583	25.994	R	1	1	1	–	
24.705	*30.130*	*L*	*1*	*1*	*1*	*ALIGN, initiate*	*L satisfied musical roles agreed*
25.360	28.460	R	0.5	1	0.75	ALIGN, match	R explores kalimba—OCM
25.983	26.646	R	1	1	0.75	ALIGN, match	R checking appropriateness of own musical behaviour—ICM
26.642	29.423	R	0.5	1	0.5	ALIGN, match	R focusing on OCM
29.421	30.013	R	1	1	0.75	ALIGN, match	R checking appropriateness of own musical behaviour—ICM
28.460	29.890	R	1	1	1	ALIGN, match	R oscillating between OCM and ICM
29.895	32.545	R	1	1	0.5	ALIGN, match	
30.018	39.004	R	0.5	1	0.75	ALIGN, match	
30.130	30.555	L	1	1	1	ALIGN, initiate	L assessment of R musical intentions
30.561	*35.698*	*L*	*1*	*1*	*1*	*ALIGN, match*	
32.545	**38.935**	**R**	**1**	**1**	**1**	**ALIGN, complement**	**At 34.021, R plays kalimba simultaneously with L's drumbeats and develops short melodic pattern, several strokes synchronous with drumbeats**
32.928	***39.254***	***L***	***1***	***1***	***1***	***ALIGN, complement***	
35.684	***36.931***	***L***	***1***	***1***	***1***	***ALIGN, complement***	***L noting appropriateness of musical interaction—ICM***
36.937	***39.663***	***L***	***1***	***1***	***1***	***ALIGN, complement***	
39.001	**45.348**	**R**	**1**	**1**	**1**	**ALIGN, close**	**R acknowledging end of contribution**
39.254	***41.755***	***L***	***1***	***1***	***1***	***ALIGN, close***	***L acknowledging end of interaction***
39.657	***40.465***	***L***	**–**	**–**	**–**	**–**	
40.472	***42.376***	***L***	**–**	**–**	**–**	**–**	

See Table 3 footnote

message" function of the musical interaction. In effect, representation in terms of action signatures would provide a "map" of an ongoing musical interaction that embodies aspects of its intentional bases, constituting a tool for the analysis of the workings of a virtual or real-world interactive musical system, and a means of specifying the operational parameters and aspects of the dynamics of a virtual system. It also affords a descriptive framework that could be used for representing and comparing interactive computer music system, as well as for describing and analysing music across cultures (although in each particular instance of its application, it would be necessary to

model the relevant style-specific set of conventionalised relationships that underpins communicative function (iii), UNDERSTANDING, in the framework).

6 Digital approaches and music as communicative interaction

Music is more than just commodifiable sound, and any representation of music in the digital domain needs to be capable of recognising this fact. In this paper, I outline an alternative view of music as an interactive, communicative, participatory medium that has social and individual consequences, and present a preliminary sketch of an approach to characterising music as interaction that derives from existing systematisations of interactive computer music systems and talk-in-interaction. This approach seems capable of portraying musical interaction in ways that capture that feature which best distinguishes music from speech as a communicative medium: its foregrounding of the relational dimension of communicative interaction. At present, however, this is only a sketch; it is far less capable of being operationalised in computational terms than is the model described by Murray-Rust and Smaill (2011), but it does seem to possess an attribute that they acknowledge their model as lacking, that of going some way towards representing (ibid., p. 1711) "intentional musical behaviour".

Of course, there are aspects of musical interaction that remain outside the scope of this sketch, in particular, the ways in which music may *mean* independently for each participant, irrespective of the fit that may be objectively identified between the ways in which each deals with the style-specific aspects of the interaction. In fact, there is no way that one can analyse such "PERSONAL" meanings as may attach during the unfolding of the musical interaction. Perhaps the only way to deal with them within the present approach is to postulate their putatively independent and potentially conflicting existence as a context for the interaction that may condition the extent to which the UNDERSTANDING component—here, roughly the analogue of language's transactional component, concerned with style-specific elements and in respect of which at least elementary understanding should be objectively identifiable—is in mutual alignment across participants. Moreover, it would be desirable that a means of characterising music as interaction could deal with music in its presentational mode; while this may be achievable, it is beyond the scope of the present proposal.

Musical interaction and interaction in speech are, as suggested above, linked together as two ends of a communicative continuum. Any approach to understanding the one almost necessarily implicates the other, and there are likely to be instances of linguistic interaction to which the theory sketched out here is as applicable as it is to interaction in music. Digital approaches to music as interaction can learn from digital approaches to the understanding of language as interaction, but the converse is equally true; interactive computer-based systems for managing language in action might have much to learn from considering the problems, and potential solutions, posed by the domain of interactive computer music.

References

Allwood J (2007) Activity based studies of linguistic interaction. Gothenbg Pap Theor Linguist 93:1–19

Allwood J, Nivre J, Ahlsén E (1992) On the semantics and pragmatics of linguistic feedback. J Semant 9(1):1–26

Bangerter A, Clark HH (2003) Navigating joint projects with dialogue. Cogn Sci 27:195–225

Borchers JO (2001) A pattern approach to interaction design. AI & Soc 15(4):359–376

Clark HH, Brennan SE (1991) Grounding in communication. In: Resnick LB, Levine JM, Teasley SD (eds) Perspectives on socially shared cognition. American Psychological Association, Washington, DC, pp 127–149

Clayton M, Sager R, Will U (2005) In time with the music: the concept of entrainment and its significance for ethnomusicology. ESEM Counterpoint 1:1–45

Coupland J, Jaworski A (2003) Transgression and Intimacy in recreational talk narratives. Res Lang Soc Interact 36(1):85–106

Coupland J, Coupland N, Robinson JD (1992) "How are you?"Negotiating phatic communion. Lang Soc 21(2):207–230

Cross I (1999) Is music the most important thing we ever did? Music, development and evolution. In: Yi SW (ed) Music, mind and science. Seoul National University Press, Seoul, pp 10–39

Cross I (2011) Music as a social and cognitive process. In: Rebuschat P, Rorhrmeier M, Hawkins JA, Cross I (eds) Language and music as cognitive systems. Oxford University Press, Oxford, pp 313–328

Downie JS, Byrd D, Crawford T (2009) Ten years of ISMIR: reflections on challenges and opportunities. In: Proceedings of the 10th international society for music information retrieval conference (ISMIR 2009), pp 13–18

Drummond J (2009) Understanding interactive systems. Organised Sound 14(02):124–133

Feld S (1981) 'Flow like a waterfall': the metaphors of Kaluli musical theory. Yearb Tradit Music 13:22–47

Feld S, Fox AA (1994) Music and language. Annu Rev Anthropol 23:25–53

Finnegan R (1989) The hidden musicians: music-making in an English town. C.U.P, Cambridge

Gabrielsson A (2009) The relationship between musical structure and perceived expression. In: Hallam S, Cross I, Thaut M (eds) Oxford handbook of music psychology. Oxford University Press, Oxford, pp 141–150

Gerstner GE, Goldberg LJ (1994) Evidence of a time constant associated with movement patterns in six mammalian species. Ethol Sociobiol 15(4):181–205

Godøy R (2011) Sound-action chunks in music. In: Solis J, Ng K (eds) Musical robots and interactive multimodal systems, vol 74. Springer, Berlin, pp 13–26

Goehr L (1994) The imaginary museum of musical works: an essay in the philosophy of music. Oxford University Press, Oxford

Grossmann R (2008) The tip of the iceberg: laptop music and the information-technological transformation of music. Organised Sound 13(1):5–11

Hawkins S, Cross I, Ogden R (2013) Communicative interaction in spontaneous music and speech. In: Kempson R, Orwin M, Cann R, Howes C (eds) Music, language and interaction. College Publications, London

Impett J (1996) Projection and interactivity of musical structures in Mirror-Rite. Organised Sound 1(03):203–211

Kim YE, Grunberg DK, Batula AM, Lofaro DM, JunHo O, Oh PY (2011) Enabling humanoid musical interaction and performance. Proceedings 2011 international conference on collaboration technologies and systems (CTS 2011)

Kita S, Ide S (2007) Nodding, *aizuchi*, and final particles in Japanese conversation: how conversation reflects the ideology of communication and social relationships. J Pragmat 39(7):1242–1254

Lakoff G, Johnson M (2003) Metaphors we live by. University of Chicago Press, London

Large EW, Jones MR (1999) The dynamics of attending: how people track time-varying events. Psychol Rev 106(1):119–159

Laver J (1975) Communicative functions of phatic communion. In: Kendon A, Harris RH, Key MR (eds) Organization of Behavior in face-to-face interaction. Mouton & Co, The Hague, pp 215-238

Lee BPH (2001) Mutual knowledge, background knowledge and shared beliefs: their roles in establishing common ground. J Pragmat 33(1):21–44

Leman M (1992) The theory of tone semantics: concept, foundation, and application. Mind Mach 2(4):345–363

Lemke MR, Schleidt M (1999) Temporal segmentation of human short-term behavior in everyday activities and interview sessions. Naturwissenschaften 86(6):289–292

Lerdahl F, Jackendoff R (1983) A generative theory of tonal music. MIT Press, Cambridge, MA

Levinson SC (2006) On the human "interaction engine". In: Enfield NJ, Levinson SC (eds) Roots of human sociality: culture, cognition and interaction. Berg, Oxford, pp 39–69

Lewis GE (2000) Too many notes: computers, complexity and culture in "Voyager". Leonardo Music J 10:33–39

Lewis J (2002) Forest hunter-gatherers and their world: a study of the Mbendjele Yaka pygmies of Congo-Brazzaville and their secular and religious activities and representations, Unpublished Ph.D. London School of Economics, London

Lindström E, Camurri A, Friberg A, Volpe G, Rinman M-L (2005) Affect, attitude and evaluation of multisensory performances. J New Music Res 34(1):69–86

Lomax A (1968) Folk song style and culture. American Association for the Advancement of Science, Washington, DC

Malinowski B (1923) The problem of meaning in primitive languages. In: Ogden CK, Richards IA (eds) The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism. Routledge, London

McLucas AD (2010) The musical ear: oral tradition in the USA. Ashgate, Farnham, Surrey

Moran NS (2007) Measuring musical interaction, Unpublished Ph.D. Open University, Milton Keynes

Murray-Rust D, Smaill A (2011) Towards a model of musical interaction and communication. Artif Intell 175(9–10):1697–1721

Nettl B (1967) Studies in Blackfoot Indian musical culture, part i: traditional uses and functions. Ethnomusicology 11(2):141–160

Nettl B (2005) The study of ethnomusicology: thirty-one issues and concepts, 2nd edn. University of Illinois Press, Chicago

Patel S, Scherer KR, Björkner E, Sundberg J (2011) Mapping emotions into acoustic space: the role of voice production. Biol Psychol 87(1):93–98

Pöppel E (2009) Pre-semantically defined temporal windows for cognitive processing. Philos Trans R Soc B Biol Sci 364(1525):1887–1896

Prior N (2008) Ok computer: mobility, software and the laptop musician. Inf Commun Soc 11(7):912–932

Qureshi RB (1987) Musical sound and contextual input: a performance model for musical analysis. Ethnomusicology 31(1):56–86

Schleidt M, Kien J (1997) Segmentation in behavior and what it can tell us about brain function. Hum Nat 8(1):77–111

Searle JR (1976) A classification of illocutionary acts. Lang Soc 5(1):1–23

Seeger A (1987) Why Suyá sing: a musical anthropology of an Amazonian people. Cambridge University Press, Cambridge

Senft G (2009) Phatic communion. In: Senft G, Östman J-O, Verschueren J (eds) Culture and language use. John Benjamins, Amsterdam, pp 226–233

Sidnell J (2009) Participation. In: D'hondt S, Östman J-O, Verschueren J (eds) The pragmatics of interaction. John Benjamins, Amsterdam, pp 124–156

Stivers T (2008) Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. Res Lang Soc Interact 41(1):31–57

Stokes M (2003) Globalization and the politics of world music. In: Clayton M, Herbert T, Middleton R (eds) The cultural study of music: a critical introduction. Routledge, London, pp 297–308

Swain JP (1996) The range of musical semantics. J Aesthet Art Crit 54(2):135–152

Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci 28(5):675–691

Turino T (1999) Signs of imagination, identity, and experience: a Peircian semiotic theory for music. Ethnomusicology 43(2):221–255

Turino T (2003) Are we global yet? Globalist discourse, cultural formations and the study of Zimbabwean popular music. British J Ethnomusicol 12(2):51–79

Turino T (2008) Music as social life: the politics of participation. University of Chicago Press, London

Varni G, Camurri A, Coletta P, Volpe G (2009) Toward a real-time automated measure of empathy and dominance. Int Conf Comput Sci Eng (CSE) 2009:843–848

Walton K (2007) Aesthetics—what? Why? and wherefore? J Aesthet Art Crit 65(2):147–161

Wharton T (2003) Interjections, language, and the 'showing/saying' continuum. Pragmat Cogn 11(1):39–91

Wilkie K, Holland S, Mulholland P (2011) What can the language of musicians tell us about music interaction design? Comput Music J 34(4):34–48