



Music in the digital age: commodity, community, communion

Ian Cross¹

Received: 10 October 2022 / Accepted: 11 April 2023 / Published online: 28 April 2023
© The Author(s) 2023

Abstract

Digital systems are reshaping how we engage with music as a sounding dimension of cultural life that is capable of being transformed into a commodity. At the same time, as we increasingly engage through digital media with each other and with virtual others, attributes of music that underpin our capacity to interact communicatively are disregarded or overlooked within those media. Even before the advent of technologies of music reproduction, music was susceptible to assimilation into economic acts of exchange. What is new in the digital world is the way in which modes of engagement with music are themselves being absorbed into an economy built on the datafication of virtual acts and the digital shadows of casual preferences. But music is more than just sounds that are culturally sanctioned as musical. Music is manifested *as* behaviours, and *in* interactive behaviour. Music is participatory as well as presentational, and in the participatory mode—involving collective, non-specialist, interactive real-time music-making—has significant individual and social consequences. Yet music as real-time participation is largely absent from the virtual world, with potential social costs that remain to be understood. Moreover, our everyday, face-to-face communicative—conversational—interactions are imbued with patterns between interlocutors that are musical, in that they share features with what we are happy to describe as “music”. These features are presently lacking in digital systems designed to subservise communicative functions, and this paper will consider the significant implications for our interactions with machines to which their successful incorporation into voice–user interfaces would give rise.

Keywords Music · Commodity candidacy · Presentational and participatory music · Back-channel · Communicative interaction · Phatic · Deontic commitment

1 Introduction

The rise of the microprocessor has transformed our engagement with music in ways that were unimaginable even 40 years ago. For musicians, whether formally trained or untrained, amateur or professional, the computer has revolutionized the production of music from creation through recording to performance (see Collins and D’Escrivan 2017). But digital technologies, and the economic and social affordances that have accompanied their emergence, are also impinging on music in contemporary societies in at least three further ways:

1. they are accelerating the consolidation of music’s status as a commodity, appropriating and altering the ways in

which we can value and engage with music, and leveraging music’s exchange value so as to commodify our acts of engagement with it

2. they are systemically inimical to the development and implementation of systems that would enable music to be used in real-time computer-mediated participatory interaction that has the capacity to enhance sociality
3. in the design of systems for computer-mediated communication, they are unmindful of attributes of music that are embodied in the interactive affordances which underpin our capacity to communicate.

Whilst each of these points requires a degree of elucidation, the first is probably uncontentious. A substantial literature has emerged around the ways in which the technologies and institutional structures of the internet have affected the ways in which—and indeed the roles through which—we engage with music (see, e.g. Cayari 2011; Erickson et al. 2013; Aguiar and Martens 2016). The effects of these technologies and institutional structures on how music can be

✉ Ian Cross
Ic108@cam.ac.uk

¹ Emeritus Professor, Faculty of Music, University of Cambridge, 11 West Road, Cambridge CB3 9DP, UK

interpreted have not radically restructured our relations to music in that they are aligned with existing Western historical trends, except in the sense that the virtual world has transformed the economic actuality of our engagement with most aspects of our daily lives. The literature surrounding these issues has tended to accept implicitly that music is a commodity. When it has addressed music in digital contexts as a focus for participatory interaction, it has done so in terms of virtual communities, or scenes clustered around engagement with particular musical genres as frames for identity formation and presentation (e.g. Bennett 2004), rather than in terms of systems for real-time creation of music (with a very few exceptions). Hence, it has neither touched significantly on digital technology's capacity for mediating music in real-time participatory forms nor has it addressed the absence from computer-mediated communication systems of the "musical" features that are increasingly being found to underpin everyday human communicative interactions. This paper will explore the idea that digital technologies encompass a severely constrained and unacknowledged culture-specific representation of "music", at best distorting and at worst suppressing the affordances that distinguish it as a flexible and intensely functional medium for social interaction (Cross 2022).

2 Music: commodity

Music is probably as old as our species, *Homo sapiens*, appearing some half-million years ago (Stringer 2016). This claim for music's antiquity as a domain of human experience may seem to be belied by the archaeological record, given that the oldest known unambiguously musical instruments (i.e. artefacts that cannot be interpreted as having any function other than to produce patterns of sounds) are comparatively recent. They take the form of bird-bone and mammoth-tusk ivory pipes found in Hohle Fels in southern Germany (Conard et al 2009), which have been dated to around 40,000 years before the present—about as soon as modern humans arrive in Europe. Nevertheless, the sophistication of these instruments indicates that they were not the beginning of a new tradition but an extension of a very ancient one. Music came out of Africa with modern humans and it is likely that aspects of something-like-music are intertwined with our emergence as a species (Cross 2016).

But whilst Germany is not the origin of human musicality, together with Britain it is one of the main sources of ways of thinking that have shaped our *understanding* of music from the eighteenth century to the present day. Over that period, thanks to Hume and Kant, concepts of aesthetic value have become key to how we have conceive of music (see Cross and Tolbert 2016). Notions of aesthetic character, aesthetic judgement and aesthetic experience have been

used to argue that music has value that is irreducible and that inheres in the unique quality of the experiences that it affords, experiences that are typically held to be "disinterested" in that they are not directly concerned with the furtherance of our own interests. We may enjoy music, but for that enjoyment to have aesthetic value and thereby cultural validity our experiences must transcend the merely sensual or hedonic pleasure that music affords us.

This idea—that music should be esteemed primarily for its aesthetic value—still receives lip service in current political discourse about "cultural products" and value: see, e.g. the *Culture White Paper* produced by the UK government's Department of Culture, Media and Sport (DCMS) in 2016. But that idea has lost significant traction over the last half-century, despite continuous attempts within music scholarship to use the concept of the aesthetic to segregate properly *musical* value (whether expressed in terms of potential for transcendence, originality, authenticity, etc.) from any form of exchange or *commodity* value (see, e.g. León, 2014). Nevertheless, the idea that music exists to be appreciated aesthetically has shaped the ways in which we conceive of engaging with music to the present day; aesthetic theories have framed music as something that exists to be heard, whether or not that hearing results in an aesthetic evaluation or merely a hedonic experience. And increasingly, music's value is equated with its hedonic value, that desirable attribute underpinning its economic value; music has unequivocally become a commodity in contemporary societies (see again the DCMS *Culture White Paper*, in which cultural value is rapidly reduced to economic value). Music may also be other things with other forms of value (see, e.g. Johnson 2002; Marett 2005; Turino 2009), but its pre-eminent contemporary form is that of commodity, with all that that entails for how music is and should be treated and conceptualised in the global neoliberal economy.

We can trace at least some of the ways in which music becomes assimilated into the capitalist system over the last few hundred years of Western history, using Appadurai's (1994) notion of *commodity candidacy*. For Appadurai (1994, p. 81), a thing can be thought of as a commodity when "...its exchangeability (past, present or future) for some other thing is its socially relevant feature". In order to be a commodity, the thing must possess the potential to become so; it must possess *commodity candidacy*. And in order for that potential to emerge, a context must exist that enables commodity candidacy to be expressed.

Music's transformation into commodity can be thought of as dependent on the routes and contexts whereby it becomes reifiable and reproducible, initially as text and latterly as sound. In pre-modern and early modern Europe (see, e.g. Dillon 2002; Brauner 2002), scribes were commissioned to create manuscripts that included musical notation. The cost or value of these scribal services, and the potential

mobility of the scribes and perhaps of the manuscripts, can be thought of as having allowed music, in its notated form, to move into a state of commodity candidacy; in effect, whilst not a full-blown commodity, musical notation, and the skills required to create it, allow music to begin to figure in an economy of exchange rather than of service or obligation. With the arrival of printing in the early modern period, music attained more than a vestigial commodity status by virtue of the reproducibility of printed musical notation, the production, reproduction and distribution of which was controlled by rights-owning individuals licenced by the (monarchical) state (see, e.g. Albinsson 2012). The rights were usually monopolistic (at least in theory), were generally inalienable (they were not transferrable), and allowed rights-holders to profit from the sale and distribution of music in notated form.

In Britain by the mid-eighteenth century we see the emergence of printed music as a mature notational and textual commodity (Hunter 1986), with property rights—now fully alienable—usually vested in the printer rather than the composer. It is the nature of the property rights that constituted the innovation rather than any particular technical development; music's existence in printed form within the capitalistic and exploitative context of eighteenth century Britain allowed its exchangeability to come to the fore. The printer was typically in physical possession of the engraved printing plates (the means of producing and reproducing the music as notation) and was thus in practical control of its sale and distribution, though this was increasingly contested by composers and by opportunistic entrepreneurs eager to exploit the limits of national and international means of policing the ownership of rights.

A little later, around the turn of the nineteenth century the emergence of the *work concept* (Goehr 1989)—the idea that music exists in the form of distinct and identifiable entities termed “works” that are created by specific individuals—helped further to consolidate music's commodity status by making it easier to think of music as taking the form of discrete and tradeable units. This status was thoroughly cemented into place by the emergence of sound recording and reproducing technology in the last quarter of the nineteenth century (Katz 2010). Music appears to escape from its own ephemerality as sound as “the work” can now be embodied in the infinitely reproducible sonic trace of its performance. Through the twentieth and into the twenty-first centuries, and transmuting in form from the physical to the virtual, music, and value in music, becomes assimilated to fit with the dynamics of a market economy as a (primarily) sonic commodity with hedonic value. It constitutes an output of the “creative and cultural industries”, a technology for auditory entertainment with which every society on earth is now familiar, in one way or another. Ownership of the means of production is largely vested (as in the eighteenth century)

not in the artists or performers but in those who own the means whereby it can be reproduced and distributed. As Lhermitte et al. (2015) document, music contributes very significantly to global economic activity in its own right, quite apart from its impact as a substantial component of TV programmes, computer games, films, advertising, etc.

Music's production, reproduction and dissemination was radically transformed over the first two decades of the twenty-first century by the advent of degraded or constrained consumer file formats, such as the (lossy) MP3 and streamable Digital Rights Management (DRM) protected files. We now engage with music in quite different ways from those prevalent even as recently as the 1990s, when the CD was just a new type of LP which we could “own”, freely exchange and autonomously replicate for our own purposes with no degradation of sound quality. The advent of the internet brought new modes of music delivery and new types of transactions involving music, allowing access to a much wider range of music than was available even to the most rabid audiophile. But it also brought new conceptions of music's “ownership” which, these days, is likely to appear much more like renting than owning (Anderson 2011; Sinclair and Tinson 2017), hedged with conditions and likely to be accessible either in a time-limited manner or with quite significant restrictions on autonomous use.

The advent of internet media consumption has been hailed by some (e.g. Shirky, quoted in Green and Jenkins (2011)) as a release from the “tyranny of one-way chains of communication”, as media consumers can now become media producers, either through creating their own music or by repurposing and gaining a degree of control over otherwise corporatised media content (Liikkanen and Salovaara 2015). However, as Green and Jenkins (2011, pp. 110–111) argue, that “release” is illusory; “every mouse click or video view is logged”, and even those who do not participate in creating content but are happy to lurk in the fringes of the web's music playpens and simply “consume” music are “... ultimately ... generating data to refine content delivery systems or recommendation engines, and ... drive up the popularity of online media businesses”.

The new players—in effect, the new publishers—are the internet platforms *Google (Alphabet)*, *Facebook*, *Apple*, etc. and their systemic associates such as *Spotify*, who have facilitated the commodification of *engagement* with music. For example, *YouTube* will harvest preferences and associations, present ostensibly targeted ads, and use the data acquired from users as they engage with music through its systems to optimise its algorithms and to link to other information that is unified by the user's presence as data hub (in, for instance, *Google*), allowing complex demographic and economic inferences to be developed, refined and employed in ways that bear no relation to music other than as the price paid for acquisition of user information (see, e.g. Drott 2018).

Music, together with other media, has become a contingent feature of commodified data which has itself become a form of capital that has value as it can be used to profile and target people, to optimise systems, to model probabilities, to grow the value of assets, and to manage, control and build things (Sadowski 2019); that commodified data can be distilled because of the online, accessible and quantifiable nature of the residues of our engagement with music.

As van Dijck (2014, p. 200) notes, “life mining”—“... extracting useful knowledge from the combined digital trails left behind by people who live a considerable part of their life online”—allows the corporates behind social media platforms to “measure, manipulate, and monetize online human behaviour” to their own advantage, in effect appropriating not only the labour (Fuchs 2010) but also the simple online existence of others to their own economic ends. Apparently private acts are appropriated as data, acquiring commodity status within markets that are largely free of independent ethical oversight (see, e.g. Shah 2018) and that may have social and political consequences in respect of which the individual originator of the data has no means of control or redress (Leurs and Shepherd 2016; de Kloet et al. 2019). Quite apart from a growing potential for social alienation and inequity, for music the result can be a flattening and homogenising of the landscape; pockets of artistic or cultural resistance remain, evidenced as bubbles of innovation, but the internet as mode of dissemination and engagement rarely allows a user’s actions to evade scrutiny and escape re-appropriation back into the corporate data economy.

Digital technology does not cause music’s aesthetic value to be reduced to mere entertainment; it simply accelerates processes already in place. Historical processes of social and technological change endow music with the status of a reifiable and reproducible commodity that may exist as text, sound or “song” (after iTunes), in forms that can be owned and exchanged. In turn, music, together with the delivery systems provided by the digital media conglomerates, becomes part of the context that facilitates the commodity candidacy of our acts of engagement with it. The factor that underpins its own continuing commodity status—its desirability consequent on the pleasure it affords—comes to be deployed as an incentive that allows the harvesting of the really valuable commodity, demographic data, for use in a market to which the music consumer has no real access and in which they have no significant role (despite initiatives such as MyData: see Lehtiniemi and Haapoja 2019) other than data hub. We are a sort of digital krill, browsing on diatomic music and in turn being grazed on by corporate leviathans.

So far, so straightforward... music’s status may have come to be increasingly alienated and potentially socially toxic in the digital world, but at least we seem to know what we are dealing with, and that understanding might eventually

enable some political restructuring of the technological and economic systems into which music has been assimilated.

3 Music: community

But music is more than either an aesthetic object or a hedonic commodity, and “engagement with music” is more than the ability to appreciate or consume music through listening. The music that we consume or appraise for its hedonic or aesthetic value is a trace or imagining of another, older, manifestation of music—music as multi-modal interaction. In other words, music as digital commodity or as bait for data capture is only one facet of music as it exists in the non-virtual world.

All known cultures have music, and all cultures expect their members to be able to make sense of their music, whether by making it, or moving with it, or listening to it. And the music that is listened to, moved to or made is more than just sound that is hedonically consumed or aesthetically appreciated. It is dynamic pattern in embodied minds, movement, and social interactions, shaped by biology and culture. It is actions and interactions that can have significant social functions that may be neither hedonic nor aesthetic and that may not rely on the activities of a specialist class, musicians, to make music but instead afford the status of music-maker to all members of a culture.

Turino (2008) makes an extremely helpful distinction between two principal “fields” of music, the *presentational* and the *participatory*. The field that tends to be privileged in the western conception is the *presentational*, which entails a clear distinction between those charged with music-making (the performers) and those whose task is music consumption (the audience); roles are relatively fixed and differentiated by expertise between performers (who will typically have to undergo formal training and /or commit significant time to acquiring musical skills) and audience members (who may themselves be distinguished by possession of different degrees of connoisseurship). A typical instance of music in the presentational mode would be a concert; whilst this may involve interaction between performers, and between performers and audience, roles and modes of interaction will tend to be relatively fixed. Within this type of “segregationist” framework there is significant potential for power and dominance to play a central role in interactions by virtue of performers’ and audiences’ possession of different degrees and types of expertise and cultural authority.

In contrast, the *participatory* field embraces music-making where roles can be open and fluid, expertise need be no more than minimal, and interactions tend towards the egalitarian, involving a high degree of mutual adaptiveness or reciprocity and the promotion of affiliation. Participatory music-making may manifest varying levels of expertise,

from the complexities of the interlocking collective vocal performance of Aka pygmy polyphony (see, Lewis 2002; Fűrnis 2006) to the simple and repetitive melodies of the *ayllu* ensembles of southern Peru (Turino 1989). It is always culturally particular (as in a group of people singing “Happy Birthday” at a party), and usually welcoming to any who are willing to accede to the (typically minimal) requirements for participation. More often than not it is multi-modal, with participation taking the form of sound and movement, as in a *ceilidh*—informal Scottish social dance (see Shoupe 2008)—or more-or-less any event involving music in West African societies (see, e.g. Stone 2010).

Participatory music-making frequently appears as a contingent element of other social activities: the singing of hymns as communal elements in the conduct of religious rites; the singing of lullabies and play-songs in the intimate and soothing interactions between mother and infant; the chants of spectators at football matches, ranging from the rehearsed and humorous offensiveness of the ultra-supporters (the *Çarşı*) of the Beşiktaş Football Club of Istanbul to the improvised and apparently casual (though still scabrous) communal singing encountered in the lower reaches of the English football leagues (Kytö 2011; Clarke 2006). In all these diverse situations, the quality of the collective music-making is not the principal focus. Its success typically lies in the degree to which it fulfils the function of enhancing social bonds, whether in the carnivalesque and hedonic atmosphere that surrounds the Aka and *ayllu* performances, the approach to the liminal represented in the conduct of religious ritual, the dyadic affect-modulation effected through the relationship between the mother's speech and song and the infant's responses, or the overt attempt in the football chant to form and project a unitary group identity capable of intimidating an opposing group.

Almost all music is, in reality, partly presentational and partly participatory. In presentational contexts, performers need audiences, whose responses may shape the mood, and sometimes govern the direction, of the performance itself—in effect coming to constitute part of the performance (see, e.g. Clayton 2007; Brand et al 2012; Moran 2013). In participatory contexts, music-making can involve fixed roles and diverse levels of expertise between contributors, exhibiting presentational features such as complexity of structure or hierarchical distinction between participants with the roles of some being more important, and more directive, than others (see, e.g. Fűrnis 2006).

Music in primarily interactive, participatory manifestations has been found to have profound effects, observable even under laboratory conditions. It involves sharing time—organisation of behaviour around a beat or periodic pulse that may or may not be physically expressed—so that participants' behaviours become entrained; they exhibit coordinated temporal structure (Clayton et al. 2005). It typically

involves a high degree of mutual adaptiveness, of awareness of and reciprocal sensitivity to each other's musical behaviours (see Cross 2008). Engagement in participatory music-making has been shown to result in enhanced sociability, empathy, prosocial behaviour, and has been linked to positive change in a range of biomarkers for wellbeing (see, e.g. Rabinowitch et al. 2013; Croom 2014; Fancourt 2016). Even listening to recordings of music—which can be regarded as constituting conserved traces or residues of virtual interaction that nevertheless afford a *sense* of interaction—may result in changes in feelings of empathic connectedness on the part of listeners (Clarke et al. 2015).

In its participatory guise, music appears to be resistant to being co-opted into the commodity role by virtue of its actualisation as transient lived experience. Participatory music-making exists as actions and interactions between people, making sense for them in the moment in ways that are not replicable, just as the music-making itself is not replicable. It may be repetitive, and a given type of participatory music-making (such as the Aka *mokondi massana*) may be repeated, but that repetition will be neither exact nor predictable in detail. Its significance, and indeed its identity, is in the sense of connectedness that it affords between participants as it is enacted. Moreover, participatory music almost always lacks the attributes of virtuosity, complexity, designed temporal structure, and sonic seductiveness that mark out most presentational music (see Turino 2008, p. 59), reducing its audience appeal and hence diminishing its commodifiability.

The imperviousness to commodity candidacy of music in its participatory form reduces the likelihood of its assimilation into the internet economy; it is largely excluded from the digital domain. Shirky may be right to claim that the internet has created something of a digital democracy in production and consumption; as Vernallis (2013) notes, those whose previous role was unalterably that of music consumer now have access to tools for music creation and dissemination and can transform into music producers. However, such activity does not breach the institutionalised boundaries of the presentational field. Whilst there has been an explosion of free or nearly free digital tools (such as *Audacity*, *Ableton Live*, *Reaper*, *Garageband*, etc.) that enable formally untrained consumers to repurpose or produce music in the presentational mode, outside the art-music world, participatory music's resistance to commodification offers little incentive to develop tools to enable digitally-mediated musical participation.

Recurrent attempts to create platforms for interactive musical engagement on the web, from Duckworth's *Cathedral* in the late 1990s (Duckworth 2003) through to present-day models based around live coding (see, e.g. de Campo 2013), have tended to be of limited impact or longevity, usually requiring specialist knowledge or privileged access

to resources. Moreover, from the outset, the conceptual framework of the conventional concert has tended to pervade the make-up of tools for real-time interactive musical creation (see, e.g. Duckworth 1999). The overview of recent systems provided in Rottondi et al. (2016) makes evident a tendency to conceive of online musical interaction as presentational performance rather than participatory event, and also delineates the enormous technical challenges posed by developing systems for remote online real-time musical interaction. These derive, at least in part, from the incompatibility between latencies in systems for online interaction (see, e.g. Chafe et al 2010, who suggest that such latencies must be less than 60 ms in order to enable successful musical interaction, ideally within the range 8–25 ms; though see also Cheston et al. 2023) and the requirements of participatory music-making for real-time reciprocity, mutual co-adjustment and co-adaptation, though these might be addressed in part by acknowledging them and incorporating that awareness into musical structures and prescriptions (see, e.g. Rofe et al. 2017).

It can be claimed, nonetheless, that there is an unambiguously participatory dimension of music in the digital domain, evident in the activities of the online communities that form around attachment to particular music tokens associated with an individual or genre. These online music communities (OMCs: Waldron 2018) take multiple forms and engage with each other and with the objects of their attachment, through diverse and novel practices. For example, Baym (2007) points to the ways in which the internet has enabled the genre of Swedish indie music to have a global reach through actors engaging with it and each other through networks of social networks. As she puts it, “Swedish indie fandom exemplifies a new form of online social organization in which members move amongst a complex ecosystem of sites, building connexions amongst themselves and their sites as they do”. This type of community blurs the boundaries between producer and consumer, with agency becoming distributed across a complex online space in ways that do not necessarily privilege the music creators (Baym 2013).

Underpinning such communities, according to Waldron (2018, p. 110) is the open-ended exchange of “social capital in the form of shared knowledge and information”, in part a legacy from the countercultures of the 1960s. This is evident even in respect of copyrighted material; whilst artists and corporates make material accessible via *YouTube*, that material is frequently appropriated, shared, repurposed and reposted in spite of threats of (and actual) legal action. Music producers have themselves begun to bypass the effects of piracy and reposting by making their materials available as means of engaging creatively (and commercially) with their audiences; in effect, the materials become social capital amongst the OMC to be shared and, critically, reworked, remixed, mashed and reposted (Michielse and Partti 2015).

Such remixing can itself, in the right social media savvy hands, result in commercial success; an extreme recent example is the rapid rise to global chart dominance of “Old Town Road”, by Lil Nas X (see Cevallos 2019). Nevertheless, for all the sense of community that music can engender in the context of OMCs, it is qualitatively different from that created through engagement in real-time, interactive, participatory contexts; some of the reasons for this should become evident in Sect. 3 below.

Overall, real-time, non-expert, multi-modal, and open engagement with music is largely absent from the digital economy other than as a niche or specialist activity because of lack of access to tools, and lack of incentive to develop tools, for online and *unpractised* musical interaction. Music in its real-time participatory forms is minimally represented in the digital world. The economics and affordances of networked digital technologies provide neither the means nor the incentive to implement systems for participatory music that would enable it to emerge in any form analogous to its existence in the analogue world, erasing the possibility of engaging collectively in music in ways that are inexpert and transgressive. Hence, music's capacity to engage and form connexions between non-expert interacting individuals—one of its principal powers in the non-virtual real-time face-to-face social world—is barely reflected in its digital manifestations.

4 Music: communication

Music as commodity pervades and is pervaded by the digital world whilst, in contrast, music as real-time creative participation has a minimal representation. A third aspect of music—its manifestation in the suite of interactive affordances underpinning our everyday communications and conversations—is even less evident. Whilst speech production and conversational interaction on computer have come a long way from Dennis Klatt's *DECtalk* (to become famous as Stephen Hawking's robotic voice: Klatt 1988) and Weizenbaum's (1966) keyword-based transformational sleight-of-hand embodied as *ELIZA*, coordinative features that shape most everyday conversational interactions and that appear to be built on the same foundations as music remain, for the present, beyond the capacities of computational systems.

In some ways, this is unsurprising whilst intuitively music and language have some sort of relationship (after all, the preponderance of music in the world is actually song), we tend to think of them, and investigate them, as two quite distinct domains of human experience. But, increasingly, research is indicating that music and language—or, more properly, music and speech, language in action—may overlap substantially in what they are and what they do. These

findings are in line with what we know from the ethnomusicological literature, where we find that aspects of many other cultures' participatory practices that appear to us to be “musical” (in that they are grounded in melodic, rhythmic, metrical, and perhaps even harmonic patterning) do not appear to be regarded emically as independent from other culturally situated modes of thought and behaviour that to us would seem to be communicative practices (see, e.g. Wachsmann 1971; Seeger 1987; Roseman 1998). In fact, a degree of ambiguity about what constitutes speech or music is the case even in Western societies, where categories of cultural practice such as poetry may seem neither speech nor music, but something in between. Poetry, however, generally exists in Turino's presentational mode, consumed from the page in private or publicly performed for an audience. It is when they are regarded from a *participatory* perspective that speech and music clearly manifest common foundations.

Research is showing that aspects of human interaction that we tend to think of as musical—shared pulse, alignment of pitch patterning between participants, coordinated movement—permeate conversational interactions, reinforcing a view of music as a component of the human communicative toolkit. It can reasonably be claimed that features of music as an interactive medium underpin our ability to communicate and are manifested in all our communicative endeavours. Viewed from this perspective, a capacity for music is as universal as a capacity for language—or, more accurately, speech. Music is as embedded in our genome as is speech. Indeed, music and speech can be interpreted as overlapping as interactive communicative media to the extent that they are best both thought of as components of a more general human communicative toolkit that can be flexibly deployed and that is typically configured in different ways in different cultures.

This claim appears to contradict the idea that conversation is a straightforward exchange of messages, constrained by the informational exigencies mapped out by Shannon and Weaver (1949). However, conversation has long been known and shown to involve more than plain message transmission; Watzlawick and Beavin (1967, p. 5) noted that “...two orders of information are present in all communication... the content and the relationship aspects... communications composed of only the one or the other are impossible”. The relational aspect reflects (at least) the operational, cultural, social, personal and affective dimensions of face-to-face (FtF) communicative interaction (Burgoon et al 1984; Boston Change Process Study Group 2005), bearing on participant identities, relationships and interactive capacities:

- Can participants hear and see each other clearly enough, and share enough awareness of the underlying premises of the interaction, that a conversation can occur?

- Do the participants share an implicit understanding of the cultural conventions of conversational interaction?
- What social obligations do the participants bring to the conversation—is their relationship socially symmetrical?
- How, and how well, do the participants know each other—intimately, casually, formally?
- Are the affective states of participants positive or negative, shared or distinct, and are participants equally aware of each other's affective state?

All these factors are likely to bear on a conversational interaction, shaping its progress and determining its effectiveness.

Feedback from the person who is not speaking or “holding the floor” is the most obvious means of managing the relational dimension of communication, together with acknowledgement of that feedback; that feedback has been referred to (rather musically!) as “accompaniment signal” (Kendon 1967) but is now more commonly known as *backchannel* (Yngve 1970; Duncan 1972). Backchannel fulfils multiple functions in real-time FtF communication and can be vocal, or gestural, or both (Starkey and Fiske 1979). Conversational interaction's multimodality is stressed by Kendon (1970, 2004), who finds that, across cultures, gesture and the vocal speech signal tend to be highly coordinated, if not temporally coupled (it is notable that such vocalization-gesture coupling appears to emerge prior to the production of first words in infancy: see Esteve-Gibert and Prieto 2014). Co-speech gestures may contribute to the import of the vocal speech signal or they may shape the context in which that signal is intended to be experienced, potentially contributing to the relational dimension (see Kendon 2004).

Backchannel and its acknowledgement establish that the communicative channel is working (“message received”) and signal participants' levels of engagement in the communicative act to each other. They can imply that assumptions about the context of the conversation—its purpose and focus—are shared between participants (“message understood”). They may allow participants to infer whether their attitudes towards the conversational focus are aligned or not, by revealing shared or opposing stances. They can enable a conversation to flow without explicit negotiation about how it should proceed by cueing participants to continue their turns or to cede the floor. In effect, backchannel and its reception are so multifarious and pervasive in their contributions to the ongoing conversational flow that they have to be regarded as integral to the joint achievement of a communicative interaction (Schegloff 1981; Coupland et al. 1992).

Over recent decades, a small but growing literature has identified aspects of communicative interaction that make use of features that can be thought of as “musical”. Ward (1996) and Ward and Tsukahara (2000) analysed English and Japanese conversational corpora and found that pitch,

particularly low pitch, was an effective predictor of the occurrence of backchannel in both English and Japanese, though the effect was stronger in Japanese. Heldner et al. (2010) found that backchannel tended to match in pitch to its preceding turn, based on analysis of a corpus of English conversations of pairs of speakers playing a cooperative card game (the Columbia Games Corpus). From more fine-grained analysis of the same corpus, Levitan and Hirschberg (2011) found alignment or convergence between speakers across turns in multiple speech parameters including intensity, maximum pitch, jitter, shimmer, Noise-to-Harmonic-Ratio and speaking rate. A different analysis of the Columbia corpus (Levitan et al. 2015), this time exploring speakers' tendencies to use different types of turns, found that speakers tend (i) to become more similar to each other as they conversed in terms of frequency of types of conversational interaction (backchannel, interruptions, etc.), and (ii) to converge on similar inter-turn latencies (the time between the end of a turn and the beginning of the next).

These analyses have established that parameters that are significant in organising the relational dimension of speech include those central to musical interaction, in particular, coordination of pitch use. Other researchers have shown that features more directly associated with music, such as pitch patterning and rhythmic organisation around a steady pulse or beat, can play a significant role in managing conversational interactions. Gorisch et al. (2012) explored a different corpus, the AMI meeting corpus (see <http://www.amiproject.org/ami-scientific-portal/meeting-corpus.html>), analysing the relationships between speakers in terms of pitch pattern. Using a complex metric for F0—pitch—contour similarity originally developed to compare amplitude modulation contours, they found that non-interrupting insertions (interpretable as type of extended backchannel) were more similar to their preceding turns in pitch contour than were interrupting insertions. A different aspect of music is revealed as potentially central to conversation in Erickson's (1981) analysis of a corpus of the natural conversations of an Italian American family. He found multiple instances of within-turn rhythmic structure and organisation of cross-turn coordination around rhythmic patterning involving temporally-coordinated alternating contributions (“hocketing”, in music). A later analysis of natural conversational interactions between a student and a counsellor (Erickson 2012) found instances of the emergence of a rhythmic structure around turn transitions, organised around a periodic pulse or beat. Both these cases require (i) the organisation of a speaker's utterances so as to highlight a regular beat, (ii) recognition of this beat and accommodation to it by the interlocutor, though it is highly unlikely that these processes are consciously evident to the participants.

Starting from the premise that spontaneous interaction in both music and speech might be grounded in common

coordinative processes, Hawkins et al. (2013; see also Cross 2013) established an audio–visual corpus of eight pairs of same-sex friends talking, improvising music together using simple instruments and playing games involving physical manipulation of objects (this last to provide a comparison condition for the musical interaction). Half the pairs were classified as “musicians” and half as “non-musicians”, though there was little if any difference in the “success” of the music-making of the two groups. We found that both music-making and conversation in the vicinity of music-making tended to be organised around a consistent pulse, which in a few instances was manifested without evident intention in cross-speaker utterances immediately prior to music-making.

Overall, we found that features that are typically musical—adherence to a periodic beat, coordination of pitch—were operational in conversations, not so much at the level of the individual but operating across turns between participants to coordinate and facilitate the flow of the interaction. For example, when speaker B wished to mark alignment of stance or attitude in their response to speaker A's question, the timing of the first accented or stressed element of their turn fell at a temporal location predictable from the timing of the last two or three accented elements of speaker A's turn (Hawkins 2014; Ogden and Hawkins 2015). Similarly, attitudinal alignment may be signalled by the formation of a musical pitch interval between the modal fundamental frequency of speaker B's turn and that of speaker A's prior turn (Robledo et al. 2016). In subsequent motion-capture experiments using same-sex pairs of strangers, movement coordination between conversational partners *after* an improvised musical interaction was found to show a significant increase in spatio-temporal alignment compared to alignment during conversation *before* the musical interaction (Robledo et al. 2021).

These explorations of spontaneous interaction in music and speech have allowed the development of clear hypotheses about what differentiates and integrates the two domains. In such circumstances, music as participation has a function that is intrinsic to it—simple continuation of the joint activity—with no extrinsic goal in view. Proximally, it has the relational effect of aligning the affective and attitudinal states of participants, affording an enhanced sense of sociality and of mutual affiliation. Distally, we can think of musical interaction as constituting an optimal means of managing situations of social uncertainty, from the dyad to the group (Cross 2006). That aspect of speech referred to as “musical” above is doing much the same relational job in conversational interactions, except that conversational interactions can express a function that can be thought of as proper to speech and that is extrinsic to the interaction: the organisation of joint action for a mutually explicit purpose. The musicality of speech underpins the grounds required

for communication that can lead to joint action towards a specific goal.

Communicative interactions, whether in the form of music or speech, are collaboratively co-constructed in real time by interlocutors; indeed, humans seem to be remarkably motivated to engage in collaborative, cooperative interactions (Levinson 2006). Whilst at any given moment one speaker may be holding the floor, the carefully timed, shaped and targeted backchannel contributions of the other participant(s) have been shown to be effectively co-creating conversations (Bavelas et al. 2000, 2002; Tolins and Fox Tree 2014). Viewed from this perspective, conversational interactions, particularly those that can be characterised as affiliative or as forms of *phatic communion* (after Malinowski 1923), seem remarkably like participatory music, with contributions from all participants being carefully coordinated in time and differentiable in function so as to shape the overall patterning of events and facilitate their continuation. And it would seem that the processes that enable this coordination are common to both speech and music, making it likely that real-time human communicative interaction cannot be modelled as a process explicable solely in terms of individual generative and representational capacities (Hari et al 2015). Any attempt to model such interactions must take account of the underlying relational processes, processes that are foundationally musical and that are dependent on the *mutual interdependence* of the interactants.

The extent to which such relational features can be or have been incorporated into systems for computer-mediated communication (CMC) is extremely limited, in part because of the multimodality of the types of cues that underpin the relational dimension of FtF communication, in part because perhaps the majority of systems for CMC have tended to be text-based, and in part because how the relational dimension is operationalised in FtF conversation is still not clearly understood, particularly cross-linguistically (for example, it is still unclear how computational systems can make the types of inferences about speaker intent that are characteristic of FtF communication—see Schuller et al. 2016). Research into CMC has indeed explored ways in which relational or interactional aspects of communication can be incorporated into computer–human communication systems, explicitly noting that there is a need to represent the functions that deal with managing the interaction itself (Vilhjálmsson 2005). Methods for integrating these types of function into CMC have been explored in text-based systems (e.g. Shankar et al 2000; Cech and Condon 2004; Liebman and Gergle 2016).

In addition, a few researchers are beginning to address the crucial issue of whether the ways in which real-life FtF communications, and non-FtF CMC interactions can be managed are similar (Degand and van Bergen 2018). There have also been some explorations in CMC systems of the

communicative functionality of simulated social signals using animated virtual conversational agents (e.g. Prepin et al. 2012; Cafaro et al. 2016), as well as a very few studies of real-world “conversational” engagement with Voice User Interfaces (VUIs). In their innovative explorations of user interactions with VUIs, Porcheron et al. (2018, p. 9) state that they “...reject the notion that such devices and interfaces are *conversational* in nature and that interaction with the interface is a *conversation* ...it is hard to make a case based on our data that responses from the device have a similar *status* to the conversation into which they are embedded”. In general, as Luger and Sellen (2016) put it in the title of their paper, interacting with a conversational agent is all too often “Like having a really bad PA”. Despite these and a surprisingly small number of other studies, overall, the ways in which the real-time interactivity of the relational dimension of communication may be integrated into the capacities of CMCs and VUIs has been severely under-explored, and it seems that this pattern of neglect is being continued in the development of LLM-type systems such as ChatGPT and its competitors (see, e.g. Bang et al. 2023; Lynch et al. 2022).

It has been claimed (Firth et al 2019, p. 124) that “...it is highly conceivable that the social connexions formed in the online world are processed in similar ways to those of the off-line world, and thus have much potential to carry over from the Internet to shape ‘real-world’ sociality, including our social interactions and our perceptions of social hierarchies, in ways that are not restricted to the context of the Internet”. However, real-time FtF communicative interaction has features that are simply not incorporated into CMC or VUI systems; real-time communicative interaction with quasi-autonomous computational interlocutors present profoundly impoverished interactive capacities, generally lacking what can be called musical attributes of reciprocity and mutual co-adaptation. One could thus equally argue—*contra* Firth—that extensive and intensive virtual interaction may desensitise an individual to features crucial to engagement in productive and flexible real-world communicative interaction, or may decrease motivation to engage in real-time FtF interaction, leading to an impoverished capacity for real-time FtF social interaction—and in the intransigence of much online interaction we may already be seeing warning signs of just such a deficit.

5 Conclusions

Music has been transformed by the digital age. Some of that change has resulted from an acceleration of existing processes; music's commodification—its absorption into an economy of exchange—was well under way before the advent of the idea of computational theory, though its more recent transformation into a contingent generator of

demographic data has added a new twist. But those accelerated changes have abstracted music from the world of real-time sociality, channelling it into a presentational mode bound to its commodity status and constraining our access to—and control over—it as a participatory medium. Computer-mediated engagement with music has become distorted by a corporate hunger for data, shaping how we can listen to it or own it. The all-encompassing consecration of music as a commodity has virtually excised music as collective real-time interactive participation from the digital domain, in which the musicality of everyday communicative interaction is likewise almost entirely unrepresented.

Is any of this truly problematic? Surely we have access to more music, and types of music, than we have ever dreamt of; we can access it instantly and much of the time at no cost; it is in the nature of things to change, and whilst we may mourn apparent losses, there are more gains than losses. Perhaps—but perhaps not. Should music be more than clickbait, more than the price paid by internet leviathans to possess more than a little of our non-corporate souls? Historically, yes: in virtually all known world cultures, music is and has been believed to manifest values that are not simply reducible to the terms of economics (see, e.g. Nettle 2015). If we wish to salvage music from its commodity status there are routes available: some already well-charted, as in the rediscovery of patronage in the form of *Patreon*; the development of artist-driven remix cultures sketched in Michielse & Partti (2015) and the online communities described by Waldron (2018); with a probable proliferation of others yet to be discovered. We can also monitor our engagement with music through corporate sites such as *YouTube*, tracking the trackers or using anti-tracking software: taking back a degree of control and disrupting the datafication of our access to music online. Beyond music as commodity, participation in music in the digital domain is largely offline, presentational and non-real-time. Those systems that do exist for real-time computer-mediated musical interaction usually require high levels of expertise and access to specialist resources. At present, there is no financial incentive to develop systems for inexpert computer-mediated music-making, though acknowledgement of a need to address “social justice issues through music-making in community” (after Waldron 2018, p. 20) could serve to stimulate the emergence of such systems.

The incorporation of interactive capacities into CMC and VUI systems and LLMs, based on relational musical qualities of mutual adaptation, raises two different types of problem. One is technical; adaptive conversational systems would need to be inferentially flexible and largely accurate at multiple levels, ranging from the satisfactory interpretation of acoustic signals to the correct interpretation of interactions in the contexts of other ongoing interactions, as Porcheron et al. (2018) point out. Assuming that these

technical challenges can be surmounted, a further problem remains: the ethical issues that such systems would raise. If an interactive system behaves in such a way as to reproduce human behaviour to the extent that it is experienced as indistinguishable from human behaviour, our behaviour in respect of it is likely to change.

In a summative report commissioned by the Royal Society and the British Academy, a team of experts produced a set of five ethical principles as a basis for “responsible robotics”. Principle number 4 (see Boden et al. 2017, p. 127) states that “Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent”. The introduction of human-like types of responsiveness into CMCs and VUIs, based on the musicality of our everyday interactions, would straightforwardly breach this rule; it would introduce an asymmetry of deontic commitment (see Kissine 2008) into the interaction. In effect, we would be likely to attribute to an adaptive VUI properties such as autonomy and, perhaps, sociality, that could condition the ways in which we interacted with it (for a preliminary exploration of the effects of embedding rhythm in a human–computer interactive task see Yu et al. 2021). This would be likely to involve ceding a degree of our own autonomy to the system and its creators. We would interact with the system as though it was qualitatively of the same type as ourselves; in feeling that we have social obligations to it, we would implicitly infer that it has social obligations—that it has made or can make deontic commitments—to us. We would be likely to engage with the system—affectively, cognitively and deontically—on a mutual basis, rather than having an understanding of the system as artificial, a product of craft of which the ultimate moral authority is vested in non-present human creators (see Bostrom and Yudkowsky 2014) whose attitudes and intentions towards us might be quite other than those that we infer from the behaviour of the system itself.

Music is irrevocably a commodity within the digital domain, a situation reinforced by the monetisation of our acts of online engagement with it; we may be unable to reclaim music from commodity status, but we should be able to regain a degree of control over how we choose to access and experience it. Non-expert engagement with others through music is ubiquitous in the world of co-presence, yet inaccessible in the digital domain; to represent the participatory actuality of music in the world of computation we need to develop and disseminate tools and systems that allow us to exercise our incompetent, non-goal-directed yet socially crucial capacity for making music together in the digital world. Finally, we need urgently to reconsider how we, as a society, deal with ostensibly autonomous digital interactive systems that are premised on the simulation of human communicative capacities. We have statements of principle concerning the design and implementation of such

systems; we need to ensure that we are fully aware of where, to whom and to what those principles should be applied, otherwise we risk ceding control of our social commitments in the digital domain to corporate or state agents who impersonate affiliation whilst aiming for exploitation.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguiar L, Martens B (2016) Digital music consumption on the Internet: evidence from clickstream data. *Inf Econ Policy* 34:27–43. <https://doi.org/10.1016/j.infoecopol.2016.01.003>
- Albinsson S (2012) The advent of performing rights in Europe. *Music Politics* 6:1–22. <https://doi.org/10.3998/mp.9460447.0006.204>
- Anderson J (2011) Stream capture: returning control of digital music to the users. *Harvard J Law Technol* 25:159–178
- Appadurai A (1994) Commodities and the politics of value. In: Pearce MS (ed) *Interpreting objects and collections*. Leicester readers in museum studies. Routledge, London, pp 76–91
- Bang Y, Cahyawijaya S, Lee N, Dai W, Su D, Wilie B et al. (2023) A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *arXiv* 2302.04023
- Bavelas JB, Coates L, Johnson T (2000) Listeners as co-narrators. *J Pers Soc Psychol* 79:941–952. <https://doi.org/10.1037/0022-3514.79.6.941>
- Bavelas JB, Coates L, Johnson T (2002) Listener responses as a collaborative process: the role of Gaze. *J Commun* 52:566–580
- Baym NK (2007) The new shape of online community: the example of Swedish independent music fandom. *First Monday*. <https://doi.org/10.5210/fm.v12i8.1978>
- Baym NK (2013) Fans or friends?: seeing social media audiences as musicians do. *Particip J Audience Reception Stud* 9:286–316. <https://doi.org/10.11606/issn.1982-8160.v7i1p13-46>
- Bennett A (2004) Consolidating the music scenes perspective. *Poetics* 32:223–234. <https://doi.org/10.1016/j.poetic.2004.05.004>
- Boden M et al (2017) Principles of robotics: regulating robots in the real world. *Connect Sci* 29:124–129. <https://doi.org/10.1080/09540091.2016.1271400>
- Boston Change Process Study Group (2005) The “something more” than interpretation revisited: sloppiness and co-creativity in the psychoanalytic encounter. *J Am Psychoanal Assoc* 53:693–729. <https://doi.org/10.1177/00030651050530030401>
- Bostrom N, Yudkowsky E (2014) The ethics of artificial intelligence. In: Frankish K, Ramsey WM (eds) *The Cambridge handbook of artificial intelligence*. Cambridge University Press, Cambridge, pp 316–334. <https://doi.org/10.1017/CBO9781139046855.020>
- Brand G, Sloboda J, Saul B, Hathaway M (2012) The reciprocal relationship between jazz musicians and audiences in live performances: a pilot qualitative study. *Psychol Music* 40:634–651. <https://doi.org/10.1177/0305735612448509>
- Brauner MP (2002) Antoine Bousnoys: method, meaning and context in late medieval music. *J Am Musicol Soc* 55:337–346
- Burgoon JK, Buller DB, Hale JL, de Turck MA (1984) Relational messages associated with nonverbal behaviors. *Hum Commun Res* 10:351–378. <https://doi.org/10.1111/j.1468-2958.1984.tb00023.x>
- Cafaro A, Vilhjálmsón HH, Bickmore T (2016) First impressions in human-agent virtual encounters. *ACM Trans Comput Hum Interact (TOCHI)* 23:1–40. <https://doi.org/10.1145/2940325>
- Cayari C (2011) The YouTube effect: How YouTube has provided new ways to consume, create, and share music. *Int J Educ Arts* 12:n6
- Cech CG, Condon SL (2004) Temporal properties of turn-taking and turn-packaging in synchronous computer-mediated communication. In: Paper presented at the 37th annual Hawaii international conference on system sciences, 2004
- Cevallos C (2019) Lil Nas X and Billy Ray Cyrus, Old Town Road (Official Movie). Columbia Records. Directed by Calmatic. released on YouTube on May 17, 2019. *J Soc Am Music* 13:396–400. <https://doi.org/10.1017/S1752196319000300>
- Chafe C, Cáceres J-P, Gurevich M (2010) Effect of temporal separation on synchronization in rhythmic performance. *Perception* 39:982–992. <https://doi.org/10.1068/p6465>
- Cheston H, Cross I, Harrison P (2023) Leader-follower relationships optimize coordination in networked Jazz performances (in preparation)
- Clark T (2006) ‘I’m Scunthorpe ’til I die’: constructing and (Re) negotiating Identity through the Terrace Chant. *Soccer Soc* 7:494–507. <https://doi.org/10.1080/14660970600905786>
- Clarke E, DeNora T, Vuoskoski J (2015) Music, empathy and cultural understanding. *Phys Life Rev* 15:61–88. <https://doi.org/10.1016/j.pprev.2015.09.001>
- Clayton M (2007) Time, gesture and attention in a “Khyāl” performance. *Asian Music* 38:71–96
- 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
- Collins N, d’Escrivan J (eds) (2017) 2nd edn. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781316459874>
- Conard NJ, Malina M, Münzel SC (2009) New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460:737–740. <https://doi.org/10.1038/nature08169>
- Coupland J, Coupland N, Robinson JD (1992) “How are you?”: negotiating phatic communion. *Lang Soc* 21:207–230
- Croom AM (2014) Music practice and participation for psychological well-being: a review of how music influences positive emotion, engagement, relationships, meaning, and accomplishment. *Music Sci* 19:44–64. <https://doi.org/10.1177/1029864914561709>
- Cross I (2006) Four issues in the study of music in evolution. *World Music* 48:55–63
- Cross I (2013) “Does not compute”? Music as real-time communicative interaction. *AI Soc* 28:415–430
- Cross I (2016) The nature of music and its evolution. In: Hallam S, Cross I, Thaut M (eds) *Oxford handbook of music psychology*, 2nd edn. Oxford University Press, Oxford, pp 3–17
- Cross I, Tolbert E (2016) Music and meaning. In: Hallam S, Cross I, Thaut M (eds) *Oxford handbook of music psychology*, 2nd edn. Oxford University Press, Oxford, pp 33–46
- Cross I (2008) Musicality and the human capacity for culture. *Musicae Scientiae Special Issue*, pp 127–143
- Cross I (2019) Musics, selves, societies: the values of musics. *Music Sci* (in press)

- Cross I (2022) Music, speech and affiliative communicative interaction: pitch and rhythm as interactive affordances Psyarxiv. <https://psyarxiv.com/tr9n6>
- De Campo A (2013) Republic: collaborative live coding 2003–2013. In: Blackwell A, McLean A, Noble J, Rohrhuber J (eds) Collaboration and learning through live coding (Dagstuhl Report 13382). Schloss Dagstuhl, Germany, pp 152–153. <https://doi.org/10.4230/DagRep.3.9.130>
- de Kloet J, Poell T, Guohua Z, Yiu Fai C (2019) The platformization of Chinese Society: infrastructure, governance, and practice. *Chin J Commun* 12(3):249–256
- Degand L, van Bergen G (2018) Discourse markers as turn-transition devices: evidence from speech and instant messaging. *Discourse Process* 55:47–71. <https://doi.org/10.1080/0163853X.2016.1198136>
- Department of Culture, Media and Sport (2016) The culture white paper. Crown copyright, London
- Dillon E (2002) Medieval Music-Making and the *Roman de Fauvel*. Oxford University Press, Oxford
- Drott EA (2018) Music as a technology of surveillance. *J Soc Am Music* 12(3):233–267
- Duckworth W (1999) Making music on the web. *Leonardo Music J* 9:13–17
- Duckworth W (2003) Perceptual and structural implications of “virtual” music on the web. *Ann N Y Acad Sci* 999:254–262
- Duncan S (1972) Some signals and rules for taking speaking turns in conversations. *J Pers Soc Psychol* 23:283–292. <https://doi.org/10.1037/h0033031>
- Duncan SP (2010) To infinity and beyond: a reflection on notation, 1980s Darmstadt, and interpretational approaches to the music of new complexity. *J New Music Cult* 7:1–20
- Erickson F (1981) *Money tree, lasagna bush, salt and pepper*: social construction of topical cohesion in a conversation among Italian-Americans. In: Tannen D (ed) *Analyzing discourse: text and talk*, vol 71. Georgetown University Press, Washington D.C., pp 43–70
- Erickson F (2012) Rhythm in discourse. In: Chapelle CA (ed) *The encyclopedia of applied linguistics*. Wiley, Hoboken. <https://doi.org/10.1002/9781405198431.wbeal1019>
- Erickson K, Kretschmer M, Mendis D (2013) Copyright and the economic effects of parody: an empirical study of music videos on the youtube platform and an assessment of the regulatory options. Intellectual Property Office Research Paper, 2013/24
- Esteve-Gibert N, Prieto P (2014) Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun* 57:301–316. <https://doi.org/10.1016/j.specom.2013.06.006>
- Fancourt D (2016) An introduction to the psychoneuroimmunology of music: history, future collaboration and a research agenda. *Psychol Music* 44:168–182. <https://doi.org/10.1177/0305735614558901>
- Firth J et al (2019) The “online brain”: how the Internet may be changing our cognition. *World Psychiatry* 18:119–129. <https://doi.org/10.1002/wps.20617>
- Fuchs C (2010) Labor in informational capitalism and on the internet. *Inf Soc* 26:179–196. <https://doi.org/10.1080/01972241003712215>
- Fürniss S (2006) Aka polyphony: music, theory, back and forth. In: Tenzer M (ed) *Analytical studies in world music*. Oxford University Press, Oxford, pp 163–204
- Goehr L (1989) Being true to the work. *J Aesthet Art Critic* 47:55–67
- Gorisch J, Wells B, Brown GJ (2012) Pitch contour matching and interactional alignment across turns: an acoustic investigation. *Lang Speech* 55:57–76. <https://doi.org/10.1177/0023830911428874>
- Green J, Jenkins H (2011) Spreadable media: how audiences create value and meaning in a networked economy. In: Nightingale V (ed) *The handbook of media audiences*. Wiley-Blackwell, Chichester, pp 109–127
- Hari R, Henriksson L, Malinen S, Parkkonen L (2015) Centrality of social interaction in human brain function. *Neuron* 88:181–193. <https://doi.org/10.1016/j.neuron.2015.09.022>
- Hawkins S (2014) Situational influences on rhythmicity in speech, music, and their interaction. *Philos Trans R Soc Lond B Biol Sci* 369:20130398
- Hawkins S, Cross I, Ogden R (2013) Communicative interaction in spontaneous music and speech. In: Orwin M, Howes C, Kempson R (eds) *Music, language and interaction*. Communication, mind and language. College Publications, London, pp 285–329
- Heldner M, Edlund J, Hirschberg J (2010) Pitch similarity in the vicinity of backchannels. In: Paper presented at the INTER-SPEECH-2010 [11th Annual Conference of the International Speech Communication Association], Makuhari, Chiba, Japan, September 26–30
- Hunter D (1986) Music Copyright in Britain to 1800. *Music Lett* 67:269–282
- Johnson J (2002) *Who needs classical music?: cultural choice and musical value*. Oxford University Press, Oxford
- Katz M (2010) *Capturing sound: how technology has changed music*, revised. University of California Press, Berkeley
- Kendon A (1967) Some functions of gaze-direction in social interaction. *Acta Physiol (oxf)* 26:22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendon A (1970) Movement coordination in social interaction: some examples described. *Acta Physiol (oxf)* 32:101–125. [https://doi.org/10.1016/0001-6918\(70\)90094-6](https://doi.org/10.1016/0001-6918(70)90094-6)
- Kendon A (2004) *Gesture: visible action as utterance*. Cambridge University Press, Cambridge
- Kissine M (2008) From predictions to promises: how to derive deontic commitment. *Pragmat Cogn* 16:471–491. <https://doi.org/10.1075/pc.16.3.03kis>
- Klatt DH (1988) Digital processor for use in a text to speech system. United States of America Patent 4754485
- Kytö M (2011) ‘We are the rebellious voice of the terraces, we are Çarşı’: constructing a football supporter group through sound. *Soccer Soc* 12:77–93. <https://doi.org/10.1080/14660970.2011.530474>
- Lehtiniemi T, Haapoja J (2019) Data agency at stake: MyData activism and alternative frames of equal participation. *New Media Soc*. <https://doi.org/10.1177/1461444819861955>
- León JF (2014) Introduction: music, music making and neoliberalism. *Culture Theory Critique* 55:129–137. <https://doi.org/10.1080/14735784.2014.913847>
- Leurs K, Shepherd T (2016) Datafication & discrimination. In: Schäfer MT, van Es K (eds) *The datafied society: studying culture through data*. Amsterdam University Press, Amsterdam, pp 211–231. <https://doi.org/10.5117/9789462981362>
- Levinson SC (2006) On the human “interaction engine.” In: Enfield NJ, Levinson SC (eds) *Roots of human sociality: culture, cognition and interaction*. Wenner-Gren International Symposium Series. Berg, Oxford, pp 39–69
- Levitan R, Hirschberg JB (2011) Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: INTER-SPEECH, Florence, Italy, August 27–31. pp 3081–3084
- Levitan R, Benus S, Gravano A, Hirschberg J (2015) Entrainment and turn-taking in human-human dialogue. In: *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*, Stanford, CA. pp 44–51
- 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*. Ph.D, London School of Economics

- Lhermitte M, Perrin B, Blanc S (2015) Cultural times: The first global map of cultural and creative industries. EY Advisory (Ernst & Young), France
- Liebman N, Gergle D (2016) It's (Not) Simply a matter of time: the relationship between CMC Cues and interpersonal affinity. In: Paper presented at the proceedings of the 19th ACM conference on computer-supported cooperative work & social computing, San Francisco, California, USA, February 27–March 02
- Liikkanen LA, Salovaara A (2015) Music on YouTube: user engagement with traditional, user-appropriated and derivative videos. *Comput Hum Behav* 50:108–124. <https://doi.org/10.1016/j.chb.2015.01.067>
- Luger E, Sellen A (2016) "Like Having a Really Bad PA": the Gulf between user expectation and experience of conversational agents. In: Paper presented at the Proceedings of the 2016 CHI conference on human factors in computing systems, San Jose, California, USA
- Lynch C, Wahid A, Tompson J, Ding T, Betker J, Baruch R, et al. (2022) Interactive language: talking to robots in real time. arXiv 2210.06407v1
- 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* (supplement 1). Routledge, London
- Marett A (2005) *Songs, dreamings, and ghosts: the Wangga of North Australia*. Wesleyan University Press, Hanover
- Michielse M, Partti H (2015) Producing a meaningful difference: the significance of small creative acts in composing within online participatory remix practices. *Int J Community Music* 8:27–40. https://doi.org/10.1386/ijcm.8.1.27_1
- Moran N (2013) Music, bodies and relationships: an ethnographic contribution to embodied cognition studies. *Psychol Music* 41:5–17. <https://doi.org/10.1177/03057356114400174>
- Nettl B (2015) *The study of ethnomusicology: thirty-three discussions*, 3rd edn. University of Illinois Press, Urbana
- Ogden R, Hawkins S (2015) Entrainment as a basis for co-ordinated actions in speech. In: Paper presented at the 18th international congress of phonetic sciences, Glasgow, 10–14 August
- Porcheron M, Fischer JE, Reeves S, Sharples S (2018) Voice interfaces in everyday life. In: Paper presented at the proceedings of the 2018 CHI conference on human factors in computing systems, Montreal QC, Canada
- Prepin K, Ochs M, Pelachaud C (2012) Mutual stance building in dyad of virtual agents: smile alignment and synchronisation. In: Paper presented at the 2012 international conference on privacy, security, risk and trust and 2012 international conference on social computing, Liverpool
- Rabinowitch T-C, Cross I, Burnard P (2013) Long-term musical group interaction has a positive influence on empathy in children. *Psychol Music* 41:484–498. <https://doi.org/10.1177/0305735612440609>
- Robledo JP, Hawkins S, Cornejo C, Cross I, Party D, Hurtado E (2021) Musical improvisation enhances interpersonal coordination in subsequent conversation: motor and speech evidence. *PLoS ONE* 16(4):e0250166
- Robledo JP, Hawkins S, Cross I, Ogden R (2016) Pitch-interval analysis of periodic and aperiodic Question+Answer pairs. In: Barnes J, Brugos A, Shattuck-Hufnagel S, Veilleux N (eds) *Speech Prosody 2016*. Boston, pp 1071–1075. <https://doi.org/10.21437/SpeechProsody.2016>
- Rofe M, Murray S, Parker W (2017) Online Orchestra: connecting remote communities through music. *J Music Technol Educ* 10:147–165. https://doi.org/10.1386/jmte.10.2-3.147_1
- Roseman M (1998) Singers of the landscape: song, history, and property rights in the Malaysian rain forest. *Am Anthropol* 100:106–121. <https://doi.org/10.2307/682812>
- Rottondi C, Chafe C, Allocchio C, Sarti A (2016) An overview on networked music performance technologies. *IEEE Access* 4:8823–8843. <https://doi.org/10.1109/ACCESS.2016.2628440>
- Sadowski J (2019) When data is capital: datafication, accumulation, and extraction. *Big Data Soc* 6:2053951718820549. <https://doi.org/10.1177/2053951718820549>
- Schegloff EA (1981) Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In: Tannen D (ed) *Analyzing discourse: text and talk*, vol 71. Georgetown University Press, Washington D.C., pp 71–93
- Schuller B et al. (2016) The INTERSPEECH 2016 computational paralinguistics challenge: deception, sincerity & native language. In: Proceedings of the annual conference of the international speech communication association, INTERSPEECH, 12 September. pp 2001–2005
- Seeger A (1987) *Why Suyá sing: a musical anthropology of an Amazonian people*. Cambridge University Press, Cambridge
- Shah H (2018) Use our personal data for the common good. *Nature* 559:7–8
- Shankar TR, VanKleek M, Vicente A, Smith BK (2000) Fugue: a computer mediated conversational system that supports turn negotiation. In: Proceedings of the 33rd annual Hawaii international conference on system sciences. IEEE.
- Shannon C, Weaver W (1949) *The mathematical theory of communication*. University of Illinois Press, Urbana
- Shoupe CA (2008) The 'problem' with Scottish dance music: two paradigms. In: Russell I, Alburger MA (eds) *Driving the bow; fiddle and dance studies from around the North Atlantic 2*. Elphinstone Institute, Aberdeen, pp 105–120
- Singer N et al (2016) Common modulation of limbic network activation underlies musical emotions as they unfold. *Neuroimage* 141:517–529. <https://doi.org/10.1016/j.neuroimage.2016.07.002>
- Stone RM (2010) African music in a constellation of arts. In: Stone RM (ed) *The Garland handbook of African music*, 2nd edn. Routledge, New York, pp 27–32. <https://doi.org/10.4324/9780203927878>
- Stringer C (2016) The origin and evolution of Homo sapiens. *Philos Trans R Soc Lond B Biol Sci* 371:20150237
- Tolins J, Fox Tree JE (2014) Addressee backchannels steer narrative development. *J Pragmat* 70:152–164. <https://doi.org/10.1016/j.pragma.2014.06.006>
- Turino T (1989) The coherence of social style and musical creation among the Aymara in Southern Peru. *Ethnomusicology* 33:1–30
- Turino T (2008) *Music as social life: the politics of participation*. Chicago Studies in Ethnomusicology. University of Chicago Press, London
- Turino T (2009) Four fields of music making and sustainable living. *World Music* 51(1):95–117
- van Dijck J (2014) Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveill Soc* 12:197–208
- Vernallis C (2013) *Unruly media*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780199766994.003.0007>
- Vilhjálmsson HH (2005) Augmenting online conversation through automated discourse tagging. In: Paper presented at the 38th Annual Hawaii international conference on system sciences, Hawaii
- Wachsmann KP (1971) Universal perspectives in music. *Ethnomusicology* 15:381–384

- Waldron J (2018) Online music communities and social media. In: Bartlett B-L, Higgins L (eds) Oxford handbook of community music. Oxford University Press, Oxford, pp 109–130
- Ward N, Tsukahara W (2000) Prosodic features which cue back-channel responses in English and Japanese. *J Pragmat* 32:1177–1207. [https://doi.org/10.1016/S0378-2166\(99\)00109-5](https://doi.org/10.1016/S0378-2166(99)00109-5)
- Ward N (1996) Using prosodic clues to decide when to produce back-channel utterances. In: Paper presented at the proceeding of fourth international conference on spoken language processing. ICSLP'96
- Watzlawick P, Beavin J (1967) Some formal aspects of communication. *Am Behav Sci* 10:4–8
- Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM* 9:36–45. <https://doi.org/10.1145/365153.365168>
- Yngve VH (1970) On getting a word in edgewise. In: Papers from the sixth regional meeting, Chicago Linguistic Society. Chicago Linguistic Society, Chicago, pp 567–577
- Yu CG, Blackwell AF, Cross I (2021) Perception of rhythmic agency for conversational labeling. *Hum Comput Interact*. <https://doi.org/10.1080/07370024.2021.1877541>
- Sinclair G, Tinson J (2017) Psychological ownership and music streaming consumption. *J Business Res* 71:1–9
- Starkey Jr D, Fiske DW (1979) Dynamic patterning in conversation: language, paralinguistic sounds, intonation, facial expressions, and gestures combine to form the detailed structure and strategy of face-to-face interactions. *Am Sci* 67(1):90–98

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.