14 Embodiment and disembodiment in networked music performance

Georg Hajdu

In this chapter the physical, sonic and spatial aspects of networked music performance (NMP) or networked multimedia performance (with an added autonomous visual layer, which for simplicity's sake we will also refer to as NMP) will be explored from a particular perspective – that of remoteness or spatial distance. In NMP the performers (agents) are typically separated by distances, which in extreme cases add up to tens of thousands of kilometres. This implies that the interactions between the agents as well as the sounds they perceive are offset by latencies that can exceed the tolerable threshold for on-the-beat performance and may require the establishment of new musical genres.

There are two concepts in NMP, one in which high-quality, multi-channel sound is streamed across the globe, and another in which the sounds are created locally, albeit shaped remotely via control messages. The first – the method of choice for musicians on acoustic instruments – requires high-bandwidth, low-latency network connections, while the other – the preferred method of digital performers – brings sound generation processes to the foreground. As the players are separated by (large) geographical spaces, the relationships between them need to be represented virtually. The performances can take place in virtual space alone or in hybrid spaces where local audiences intersect with virtual ones. The interplay of the physical, sonic and spatial dimensions, taken for granted in most traditional forms of music performance, becomes all the more crucial when spectators create mental maps of the interactions they observe, which greatly influence the extent of their appreciation. The working hypothesis is that this appreciation relies on the plausibility between physical action and sonic result. Because of the remoteness of the participants, these actions may not always be perceived directly or immediately. Hence, they have to be embodied by adequate representations to overcome the disconnect, whereby a mutual bond between performers and the spectators, perceiving this interaction, is being negotiated.

In order to fully realise the implications of NMP, I will present in this chapter a short historical overview before turning the attention to concepts underlying the nature of distributed performances. Next, with those

concepts in mind, I will zoom into and critique current practices by acknowledging that NMP is particularly susceptible to various types of disembodiment, namely spatial separation, microtemporal fluctuations and poverty of sound production and transmission. I will conclude this chapter by extrapolating the present state of affairs to predict what the future might hold in stock.

Brief and personal history of networked music performance

NMP has a long history that started in the late 19th century fuelled by advances in telegraphic, radiophonic and early recording technologies. The first system is this area was the Telharmonium (Weidenaar, 1995), which consisted of a gargantuan additive synthesis engine delivering live music performances on an organ-like instrument to households via telephone wires. As with the radio, which had its first public performance only three years after the Telharmonium's opening season (1907 vs. 1910), both media delivered content without the performers having 'to be there'. This physical *disconnect* between a musician creating a sound and a loudspeaker reproducing it must have been a bewildering experience for people at the time, whether they were playing music into or listening to music from the machine. Hungarian composer and ethnomusicologist Béla Bartók, who had made most of his phonograph recordings and transcriptions between 1907 and 1918, visited Turkey in 1936 for a further field study and reported this amusing anecdote about a folk singer:

> Bekir oğlu Mustafa was the first singer to sing into the phonograph in his tribal settlement, and he could only be persuaded with great difficulty to do so; he was afraid he would 'lose his voice' permanently if the machine should 'take it'.
>
> (Bartók, 1976, p. 58)

As curious as this account may sound, it touches on some fundamental issues relating to embodiment and disembodiment of recorded and electronically generated sounds, which are an essential part of NMPs.

Electronic music

After the short-lived attempts of the Telharmonium and the Pianola to bring music to households without musicians needing to be there, the radio (along with the turntable) established itself as a mass medium of the 1930s, transmitting music and speech to an ever-growing number of receivers. It comes as no surprise that John Cage, an early adopter of any technological progress, turned his attention to the radio as a musical instrument. His *Imaginary Landscape No. 4* from 1951 was conceived for 12 radios, whereas the first piece from this series was written for piano, cymbal and two

variable-speed turntables with frequency recordings. While these pieces actually required performers acting in front of an audience, electronic music since the late 1940s was mostly listened to via radio loudspeakers or in front of empty stages. See the section 'Cause and effect' (below) for a more in-depth discussion.

Network performance

The advent of affordable digital computers, customised controllers and MIDI (as a standard to connect music devices) introduced new opportunities for experimental developers. In 1975, Kraftwerk started using custom-built electronic controllers on stage and, in 1978, the League of Automatic Music Composers began performing with networked microcomputers such as the Commodore KIM-1 (Brown & Bischoff, 2002). The effect on the audience was entirely different: While in case of Kraftwerk the audience could appreciate the connection between physical actions and the sounds produced, this aspect was partially hidden in the performance of the League and The Hub. In a memorable performance at the 1992 International Computer Music Conference in San José, California, the six male members of The Hub sat around tables mounted with computers, a MIDI hub (hence the name of the band) as well as various controllers and sound-producing devices. Although the performers did not face the audience directly, the audience could observe intense concentration and physical involvement during the show. After the last piece, the band radiated a sense of relief and satisfaction upon the successful ending of the concert, which the spectators appreciated even though it was difficult to connect the gestures made by the band members to the sonic outcome.

The 1980s also saw further development of a different kind of network music in which recorded sounds were sent over large distances via satellites or landlines and remixed in radio studios in real-time. People such as Max Neuhaus (Public Supply) and Bill Fontana (Soundbridge Cologne/San Francisco) fostered this approach between the 1960s and the 1980s. These distinct practices of local network performances mediated via *control messages*, on the one hand, and long-distance *audio streams*, on the other, started to merge in the mid to late 1990s when technologies became available to musicians to send data regardless of their type around the globe; yet a conceptual distinction between those approaches (audiovisual (AV) streaming vs. exchange of control messages) is still observable today. Improvising musicians such as Pauline Oliveros and Chris Chafe motivated the development of *telematic* performances, which essentially consist of musicians playing together over low-latency audio and video links – a practice that existed even before the popularisation of the internet. However, the exchange of control messages (discreet and continuous) takes advantage of the peculiarities of digital media and therefore is to be regarded as the actual innovation in NMP.

In the early years, though, internet music performances were hampered by severe issues as stated by Brown and Bischoff (2002):

> As such [musicians first and technologists second], we were the first (as far as we know) to make interactive, live electronic music in a computer network, and despite the primitive nature of that network (compared to those available at present) we were the first to experience its potentials and its problems. One of those problems has to do with distance. As instruments, and ensembles, get more complex, the direct interaction of people with sound becomes difficult to maintain. Computer music instruments are best when they take on a life of their own, surprising their creator/performers with a liveliness and character that could not be predicted; but there remains a need to guide them directly, to nudge their behaviour in this direction and the next by physical gestures, and to hear the results of those gestures begin to emerge immediately. When the network mediates those gestures further, a disconnect can take place that alienates the player from interaction with the music.

According to Brown and Bischoff this alienation could only be overcome by maintaining a balance between the behaviour and responsiveness in the design of electronic instruments and networks – an issue which has since been addressed by some developers and practitioners such as the people affiliated with SLork and PLork (the Stanford and Princeton laptop orchestras). Table 14.1 presents a comparison of the technical conditions surrounding early and late internet music performance in respect to local scenarios.

Quintet.net

Unaffected by (or unaware of) the sobering experiences reported by The Hub, the newly available technologies attracted a significant number of computer musicians and multimedia specialists such as Pauline Oliveros, Atau Tanaka and Chris Chafe at the turn of the millennium to embark on a new journey on the internet. I described my own approach to NMP in an article I presented at the 1999 SoundArt Festival in Osnabrück, Germany (Hajdu, 2003). It contained the layout of a system which has been in constant development ever since and was the basis for countless multimedia projects ranging from an internet opera project at the Munich Biennale to local network performances with the European Bridges Ensemble or students of the Hamburg University of Music and Theatre.

Quintet.net (Hajdu, 2005) is a multi-user, real-time, server–client-based digital multimedia performance environment consisting of different types of nodes (up to five players, an unlimited number of listeners and viewers, one conductor) assuming different roles and communicating via network protocols by exchanging discrete and continuous control messages of different types. As the early years of internet performance were an uphill battle

Table 14.1 Comparison of different network scenarios

	Early internet music (discrete control data and low-quality streaming)	*Late internet music (continuous control data and high-quality streaming)*	*Local performance (electronic/ acoustic)*
AV cues (musicians/ audience)	No	Via streaming	Immediate
Other cues	Avatars	Avatars	Immediate
3D visual perception	No	Technically possible	Yes
Spherical 360 degree visual perception	No	Technically possible	Yes
Sound localisation	Vector-based amplitude panning (VBAP)	Ambisonics, wave-field synthesis (WFS)	Immediate
Sound quality	mp3, aiff (16 bit, 44.1 kHz)	Any file type	Immediate
Network bandwidth	Dial-up modem, ISDN, low-bandwidth DSL	Internet2, GÉANT	Local
Network quality	Low throughput, Dropped packets, Errors, latency, jitter, out-of-order delivery	Quality of Service (QoS)	Local
Music styles	Collaborative soundscapes, control of generative music, loops, time brackets, free improvisation	Latency < 50 ms: beat-based/ metered music Latency > 50 ms: Collaborative soundscapes, control of generative music, loops, time brackets, free improvisation	Any style
Music notation	Simple	Complex	Any means

against technological shortcomings, the challenges that internet performance had been posing in the early days, and the solutions to be found to compensate for them, had to be addressed.

Music performance on the Internet also leads to socio-psychological questions as to the identity of the performer playing on the Net and how

> his or her presence is felt by the other musicians and their audiences, as well as where 'adequate' performances are supposed to take place and how they should be presented – questions that touch on the fundamental roles that musicians and their audiences have taken for granted for centuries.
>
> (Hajdu, 2005, p. 24)

In the performance of my piece *MindTrip* at the 2000 Mystik und Maschine Festival in Münster, Germany, on 28 October 2000 (the very first performance of a Quintet.net composition), the five performers were located in different cities across the globe. They were linked by the Quintet.net server running locally in Münster, connected via a 128 kbit dial-in ISDN connection to the internet. Münster performer Erhard Hirt, who played a Roland MIDI guitar, reported that it was difficult for him to localise and identify his partners, as their sonic identities were masked by sounds assigned to each performer via the Quintet.net sampler. As a seasoned improviser who had been used to read subtle visual and acoustic clues from fellow musicians, Hirt felt insecure about his hybrid role and double presence in two distinct spaces with a real audience and remote, virtual partners. In an email to me after the performance he commented on the deadpan quality of the exchanged sounds:

> Isn't the problem most likely that the complexity of a musical action by a [remote] partner is not being communicated via Internet, but instead just its reduction via pitch-to-midi conversion?

Hirt as well as the audience, the latter largely not used to computer-generated music, let alone to sounds produced by musicians who weren't there, had to rely on the spatial arrangement of the five loudspeakers on which name tags of the remote musicians had been placed – in addition to the display of 'real-time event notation' projected onto a large screen in back of the stage. Although participating Hub member Chris Brown pointed out that using shared notation was 'breaking new ground', many people in the audience felt that the approach held promise but was not yet living up to its potential. This was partially due to the lack of real physical interaction between the performers, but also because of the blandness of the sampled sounds. Because of technological limitations, we had to take special care to keep the memory footprint of the audio samples small enough and just use a limited number of voices in the polyphonic sampler. Using reverb or other digital audio effects to smooth and enliven the sounds was absolutely out of the question, as this would have exceeded the computational power of our machines. Consequently, the sonic results sounded artificial, machine-like and disembodied, but we also had to deal with the hitherto unresolved psychological issues of remote performance.

In 2002, Dutch-American composer and flutist Anne La Berge devised a piece called *Vamp.net* in which the performers received textual instructions

via the Quintet.net interface while improvising on their computer keyboards to a video. In the rehearsals the musicians, who all played from their private residencies, generally produced too many events, indicative of a communication problem. It was as if their solitary situation, the lack of secondary clues that musicians are sensitive to and the use of an on–off general-purpose controller had created a fear-of-void situation, for which they compensated by hyperactivity. This ultimately led to a breakdown of the type of interaction free improvisation ordinarily relies upon.

Therefore, under these circumstances, free improvisation was not the way to go. Instead, the participants had to be involved in scenarios in which clear instructions as to how to fill time had to be given. John Cage's number pieces, of which I transcribed *Five* (written in 1988), held promise in that they feature scores in which musical events are supposed to be performed within *time brackets* allowing a certain amount of indeterminacy as to their onsets. This also allowed us to regard network latency as an integral part of the performance rather than something that needed to be avoided or compensated for.

In the early days of Quintet.net, musical events were generated either by typing on the computer keyboard, playing on a MIDI instrument or using a pitch tracker to track the actions of acoustic instruments, with the former being our favourite playing mode, as it represented an aesthetical purity, particularly in local performances where laptops would be mounted on stable music stands arranged in semicircles (reminiscent of Kraftwerk's minimalistic stage set-up). The downside was its on–off paradigm allowing little expressive control over the sounds as they developed, further contributing to the sense of disembodiment.

Curiously, though, this was a welcome ingredient in the Munich Biennale opera *Orpheus Kristall* – an opera whose creation was instigated by musicologist Golo Föllmer and cultural manager Jens Cording of the Siemens Arts Program, touted as the first internet opera. In this opera, the stage of the Gasteig Carl-Orff-Saal was connected to four other locations in Amsterdam and the USA (New York and California). The sounds were converted into numeric data via pitch tracker. While the remote players spent the entire duration of the opera improvising, the result of that was to be heard in the concert hall only during certain predefined moments ('windows') which composer Manfred Stahnke had indicated in his score. During one of these moments, the digital multiplication of the notes produced by the performers was filtered through the spectrum of the baritone's speaking voice, analysed in real-time by the pitch tracker. The result eerily merged the qualities of the voice with the ongoing network improvisation. It meant to metaphorically *embody* the chaos of the outside world infringing on the life of Orpheus: by giving the chaos a voice as well as shaping and absorbing it, Orpheus thus reveals his quality as a true divine creator.

Again, for the audience, the role of the network was difficult to decode, as the remoteness of the participants prevented the establishment of a *causal*

relationship between their actions and the sonic outcome audible through loudspeakers. Conceptually, the integration of the NMP posed serious challenges to the director, who in the end decided to hide the projection of the Quintet.net interface during the opera performance (only the Bayerischer Rundfunk production blended the imagery of the stage with the video of the computer screen). Although technically we succeeded to play through three performances and the dress rehearsal without crashes or other major accidents, a shift of paradigm for NMP was mandated. It can be summarised as follows:

- The current praxis of NMP needs to be put under scrutiny and its constituents extensively studied.
- NMP should not just mimic traditional forms of performance.
- A *plausible* relationship between the remote actions and their sonic and visual results ought to be established for performers and audiences alike.
- Expressiveness is to be achieved via continuous tracking of physical activity.
- Along the lines of Marshall McLuhan ('the medium is the message'), the *narrative* of network compositions needs to have a reference to its very technical nature.

With this aim in mind, I co-founded, in 2005, the European Bridges Ensemble (EBE) consisting of five electronic performers, a video artist and a 'conductor'. Two large grants allowed us to finance a number of activities, the first Music in the Global Village conference at the Budapest Kunsthalle among them. The first period from 2005 to 2007, in which we transitioned to using computers equipped with the powerful line of Intel Core processors, focused on the improvement of the audio system, allowing higher degrees of polyphony as well as other types of sound generation (besides simple sample playback), sound spatialisation and a more versatile video generation and processing component. A successful piece from this period was Sascha Lino Lemke's *Netze spinnen # Spinnennetze*, which premiered at the Akademie der Künste in October 2007 on the occasion of the György Ligeti composers' contest. Its two parts, inspired by the musical and nonmusical ideas of the great Hungarian composer, featured sophisticated user interfaces and sound generation modules as well as a rich textual score with instructions as to how and when to operate these interfaces via the computer keyboard. It also included a *social, game-like* component allowing the participants to interfere with each other, for instance by stopping the sounds of the other performers.

Performativity in musical networks

Cause and effect

The members of the EBE often found themselves in the position of having to explain the nature of their interactions to their audiences, which were

to a great extent attracted by the novelty of their project. But does novelty itself suffice? Does an artwork or a performance need a lengthy explanation before it can be appreciated by the audience or would there have been ways of establishing a more intimate sense of *causality* (or, more accurately, *plausibility*, as we will see later on) between actions and the sonic and visual results? Would the disconnect between performers and audience thus be overcome by appealing to the knowledge embodied in the spectators?

The relationship of cause and effect is perhaps the most fundamental principle in music. In the acoustic environment before the advent of loudspeakers, a sound could only be produced by a physical effort causing a resonating body to vibrate. This relationship was so intimate that even in the absence of a perceivable cause, a trigger for a sonic event could safely be inferred. It is obvious that from an evolutionary standpoint such a close association between cause and effect makes sense when it comes to quickly deciding whether to *fight* or *flight* in front of imminent threats (Wallin, Merker & Brown, 2000). Experiments of audiovisual phenomenal causality (Guski & Troje, 2003) have shown that for the perception of audiovisual synchrony (i.e., synchrony between visual and auditory triggers), 'a sound may be judged to be synchronous with a visual event even when the sound occurs after a short delay. In contrast, a preceding sound is less likely to be judged as synchronous' (p. 790). In a similar vein, Chafe, Cáceres and Gurevich (2010) conducted tapping experiments with artificial latencies between two visually and sonically separated musicians and have shown that our nervous system is best attuned to naturally occurring latencies such as those happening on concert stages.

In music, the bond between cause and effect is so intimate that in musical semiotics the index ('a sign that is linked to its object by an actual connection or real relation') can be regarded as the most prevailing sign. In his book *Grundriss der musikalischen Semantik* [Fundamentals of musical semantics] Czech musicologist Vladimír Karbusický – basing his theory on the writing of Charles Sanders Peirce – elaborates on this intimacy:

> Its first ontologically significant factor is the immediate bond: something indicates … the state … of an object. … The psychosomatic relevance of indices makes its application to music – an art form that touches the soul and moves the body … The second factor is the energetic nature of this bond. Peirce spoke of the 'reactions of Secondness between the ego and the non-ego' and related this to 'Aristotle's use of energeia, action'.
>
> (Karbusický, 1986, p. 59, translated by the author)

The development of recorded media since the 1930s has fundamentally challenged this bond: By *decontextualising* (like in the early pieces of the *musique concrète* era), composers and sound artists were encouraging the listeners to experience sounds as such and not within the context of

their production; and by *processing*, which allowed sounds to be 'denatured' to borrow a term from biochemistry, that is, change their properties until they become unrecognisable. Tape recorders permitted composers and sound artists to play a sound backwards metaphorically, reversing cause and effect. Extreme transpositions and time stretching also made sounds nearly unrecognisable. The 1950s added electronically generated sounds to the equation, and composers and their audiences experienced, for the first time in history, sounds devoid of the physical act of creation. Acousmatic music was the genre that logically grew out of this scenario and turned the non-physicality of an empty stage filled with nothing but loudspeakers into a virtue. Paradoxically, the most approachable pieces of this genre, often still maintain causality on higher semantic and syntactic levels, which can be exemplified by comparing 'intentional' music by Smalley and Gobeil to the non-intentional and non-causal pieces by John Cage (such as *Williams Mix*).

Gestures, performativity and agency

In the same context Karbusický (1986) also indentified music as being permeated by gestures. He differentiated between three types: (a) *real-life gestures* (e.g., the motion of a conductor) represented as the *energetic substance*, constituting (b) the *mimetic basis* for music as well as (c) *codified gestures* in ballet. While Karbusický fails to make an explicit reference to the then budding genre of interactive music (his book was published in 1986), gestural control (sometimes referred to as *technical gestures*) has become an indispensable ingredient in the interactivity between man and machine. Kim and Seifert (2007) have extensively studied the nature of interactive music systems and discussed them in the context of recent theories of *agency*, *situatedness* and *embodiment*. Establishing the term performativity as the 'capacity to achieving efficiency which the moment of performance gives rise to', they posit that 'the framework of performativity offers an alternative perspective on music performance, which serves as the 'stage' integrating the processes of music production and perception' (p. 234):

> Francisco J. Varela [the Chilean biologist, philosopher and neuroscientist], who understands cognition as embodied action, places an emphasis on cognitive processes which depend on 'the kinds of experience that come from having a body with various sensorimotor capabilities' that are embedded in an environmental context, whether biological or cultural. In other words … cognitive processes are considered as emerging from the 'continuous reciprocal causation' between mind/brain, body, and environment. According to this embodied approach in cognitive science, not only physical instantiation, but also an agent's situatedness in its environment is of importance to cognition.
>
> (Kim & Seifert, 2007, p. 234)

Kim and Seifert also discuss the distinction between embodied and disembodied music systems. According to them, closed as well as knowledge-based interactive systems (e.g., those which rely on score following to elicit a musical response), are disembodied whereas coupled interaction of one or several agents with the environment, guided by action-perception loops is a necessary condition for embodied systems.

Embodiment

As the term *embodiment* (and its antonym *disembodiment*) has several meanings and connotations we should stop for a moment and carefully define them so as to not confound them later on (see also Leman et al., this volume). We encounter the term in the context of *embodied cognition*, referring to the unity of perception and action manifest in the mirror system (see below). *Embodied agents* are intelligent agents, which communicate with their environment via a physical or virtual body (Cassell, 2001). A branch of artificial intelligence focuses on empowering such agents to interact autonomously with humans. The most common use, though, refers to the materialisation of abstract concepts. In that sense, the semantic field of embodiment overlaps with those of the terms *metaphor* and *allegory*. Brandon LaBelle (2015) in his book *Background Noise – Perspectives on Sound Art* analyses Atau Tanaka's early network piece *Global String*, a multi-site network music installation, connected via the internet, consisting of a musical instrument in which the network forms the resonating body, by use of a real-time sound synthesis server. La Belle writes:

> Tanaka's own virtual creature, as embodiment of global bodies, inputs and connections, thrives on the hybridity of information/digital code/data files/etc., giving voice to the terrors and pleasures wearing a new skin presents. (p. 277)

In this context he also makes a reference to McLuhan's concept of the 'extended nervous system'. For McLuhan the physicality of electronic media (or lack thereof) is central to his media theory. In *The Global Village: Transformations in World Life and Media in the 21st Century* (McLuhan & Powers, 1989), McLuhan became increasingly concerned about the cultural and ethical effects of what he termed *discarnetism*, that is about the disembodiment implied and realised by much of our communication via electronic media. According to McLuhan, we assume a robotic state when we become unreflective extensions of our own technologies. But, when we are 'on the phone', we are disembodied or discarnate. Angelism, is what McLuhan called this state. According to McLuhan, angelism was actually fairly dangerous for humans, especially their social life:

> If he [the caller] is not identified or chooses not to identify himself, he loses touch with a geographic location and a social function. He

> becomes truly discarnate and, in that psychic sense, uncontrollable – a phone poltergeist … .
>
> (McLuhan & Powers, 1989, p. 124)

For the audience, the success of a NMP (or of a staged performance in general) depends to a great extent on whether they can understand and anticipate the actions of the performers. Two mechanisms, the *mirror system* and the *theory of mind* have been implicated in how the human brain might accomplish these tasks (Gazzaniga, Ivry & Mangun, 2014). The mirror system consists of mirror neurons in premotor cortex and other areas that respond to an action both when the action is produced by a subject or observed when produced by another subject. The mirror system has been assumed to be an essential part of comprehending the actions of other subjects. It provides an intimate link between perception and action with the consequence that the brain does not form abstract representations of visual patterns that conform to various actions, but rather our comprehension of such actions depends on our own ability to perform those deeds. The fact that our conceptual knowledge is grounded in our body knowledge has also been referred to as *embodied cognition*. Remarkably, the extent and intensity of activation patterns reflect a subject's repertoire of particular motor patterns. It was suggested (Aglioti, Cesari, Romani & Urgesi, 2008) that a well-practised motor system is anticipatory in nature, giving it the ability to predict others' actions in the arena of their expertise. Mirror systems are also thought to play an important role in imitation and learning new skills and for the understanding of other subjects' *intentions*. A similar capacity of humans to infer mental states of other people has been coined *theory of mind* by Premack and Woodruff (1978).

Causality, plausibility and illusion

While Kim and Seifert (2007) present their ideas from the perspective of agents involved in a particular performative situation, there is no mention of the role of the audience in observing the outcome of such performances. Taking the role of the neural mirror system and theory of mind into account it is safe to assume that the audience is either empathic or non-empathic to the outcome of a performance, depending on its effectiveness. As the audience in most cases does not participate in the performance, it cannot verify whether an action is causal or only pretends to be – nor does it really matter dramaturgically. Mark Applebaum has demonstrated this paradox convincingly in his piece *Aphasia* (Figure 14.1). The performer seems to trigger sounds with gestures typical for interactive music performances, yet the sounds are played back by a sound system. In his score, Applebaum (2010) writes that 'the hand gestures must be precisely synchronised with the sound, the illusion being that the gestures cause the sound or vice versa' (p. 2). He also points out that 'the flow will be most persuasive when memorised

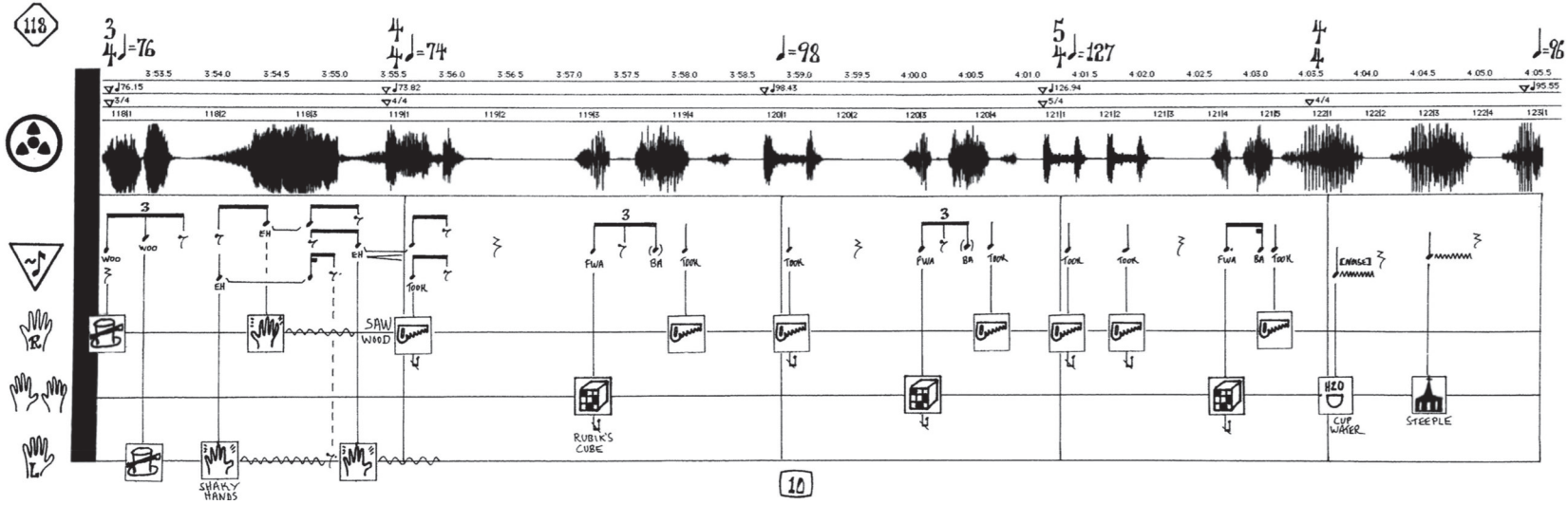

Figure 14.1 Excerpt from *Aphasia* by Mark Applebaum. The performer synchronises his/her gestures to the playback and thus creates the illusion of causality.

and also allow the performer's eyes to remain fixed on the audience; furthermore the very presence of a score creates a barrier – both psychological and one of sight lines – between the performer and the audience' (2010, p. 2).

We are therefore encountering a contradiction between the definition of performative music systems as embodied – and closed, non-interactive systems (such as the one used in *Aphasia*) as disembodied. The performance may be disembodied from the strict definition given by Kim and Seifert, but it certainly does not *appear* that way. We have to therefore abandon *causality* as necessary condition for perceived reality and replace it with the concept of *plausibility,* which also encompasses the theatrical phenomenon of *illusion*.

Another issue with Kim and Seifert's definition of interactive performativity lies in their rejection of the *intending* artist as an agent wilfully acting in a closed artwork in favour of embodied aesthetic experience that emerges during an interactive performance. This may hold true for the types of interactive installations that audiences encounter in museums or other public spaces but falls short in describing the experience of musicians that practice and rehearse for any extended time to achieve expertise and fine control in their interaction with human–computer interfaces. It seems, therefore, that their definition of performativity calls for an extension in which hybrid, that is score-, memory- or knowledge-based scenarios can be integrated without contradiction.

In contrast, Gurevich and Fyans (2011) also studied digital musical interactions (DMI) as relationships between performers and digital music interfaces as well as their perception by spectators. Referring to an analysis of interactions with 'virtual musical instruments' (Johnston, Candy & Edmonds, 2008, pp. 208–209), the authors differentiate between three different modes of interaction – instrumental, ornamental and conversational – of which only the first corresponds to the traditional mode of continuous expressive control of an acoustic musical instrument – citing Cadoz (2009, pp. 207) that 'in the case where the gesture is performed on a physical device like a key or a joystick' the 'transduced signals have no physical (ergotic) meaning'. Non-instrumental modes, common in DMIs, are typically differentiated from the instrumental mode in that they rely on cognitive rather than perceptual-motor skills, as they control out-of-time or temporarily remote events (as opposed to continuous in-time events). It appears that instrumental modes facilitate the construction of mental model of the interaction, which is a key ingredient of the spectators' appreciation of the DMI–performer interaction. Without it 'some participants' perception seemed primarily informed by bodily cues – facial expressions or physical "confidence"'. This fact leads Gurevich and Fyans to stipulate that DMIs engage spectators in 'different modes of interaction, sometimes simultaneously'. Therefore, 'successful DMIs are more likely to be those that … capitalise on this flexibility that digital devices afford' (p. 174).

One of these modes of engagement can also be found in the game-like nature of many network music projects,[1] which was studied in detail by Golo Föllmer (2005). He based his investigations on the social theories of Dutch

philosopher Johan Huizinga, whose seminal work *Homo Ludens* provides an analysis of the *play element of culture*. This also explains why participatory concepts, in which the boundaries between performers and spectators disappear (Freeman, 2008), have become a popular ingredient in network music performances.

Virtual reality and telepresence

Illusion is a key component of both magic and games, particularly on the internet, where people often mask their true identities and take on virtual ones. It has been shown that subjects are willing to accept low-dimensional virtual representations of themselves called avatars as long as the movements of such avatars are plausible. This is further supported by Wöllner (2012) who reports that conductors are capable of recognising their own gestures just from simple point-light displays.

In an article in the *Scientific American* Jaron Lanier wrote about avatars:

> Although the computer power of the day limited our early avatars to extremely simple, cartoonish computer graphics that only roughly approximated the faces of users, they nonetheless transmitted the motions of their hosts faithfully and thereby conveyed a sense of presence, emotion and locus of interest.
>
> (Lanier, 2001, p. 69)

In the same article, he also described a tele-immersion system that may eventually bring the world to one's desk. His system, designed for users to share virtual spaces involves streaming video from a battery of cameras creating a disparity map reflecting the degree of variation among the images from all aspects to create a single viewpoint-independent sculptural model of the scene. Experimenting with the system, Lanier noted that 'some critical aspects of a virtual world's responsiveness should not be subject to more than 30 to 50 milliseconds of delay. Longer delays result in user fatigue and disorientation, a degradation of the illusion and, in the worst case, nausea' (2001, p. 74). This is on a par with the observations of Chafe, Cáceres and Gurevich (2010), who reported that for remote musicians locking into each other's performances 35 milliseconds represent an optimal value within the acceptable range of latency.

There has been some hope that machines capable of *predicting* users' *intent* will be able to compensate for latency (Chafe, 2009; Alexandraki & Bader, 2014) and allow performances synchronised to a beat in which the combined latencies of round trip times (RTT) and audio buffering noticeably exceed the optimal range. (It is said that Internet2-type networks allow transmission speeds of up to half the speed of light. Under these conditions, tight beat-based performances may not work for distances larger than 5000 km, roughly the distance between the two North-American coastlines.)

Yet, is doubtful whether such *prediction* mechanisms hold the key to overcoming lag, as they also have to take discrepancies between intent and prediction into account, which may be perceived as mistakes. We would therefore trade a lack of temporal accuracy for a more pronounced likelihood of syntactic errors. The performance would also suffer from a decrease of expressiveness, as it would start to increasingly resemble a stream of algorithmically generated events controlled by what Johnston et al. (2008) call non-instrumental modes of interaction.

Dramaturgy of networked music performance

The beginning of the millenium was the time when the first telematic performances took place – their organisers seeking to employ telepresence technologies for remote interactions. The frequency of network music performances rose considerably over the following years and gave researchers an incentive to study the network topologies and social nature of such interactions, providing opportunities for *distributed creativity*. Rebelo, Schroeder and Renard (2008) came up with the notion of network dramaturgies encompassing aspects of authorship, collaboration, structure, content and as general term for a number of aspects that characterise performance practice. They posit that 'authorship and collaboration in particular are problematised in the context of network performance through questions of involvement of multiple sites/nodes and the relation between multiple kinds of artistic input' (p. 1). Therefore, notions of presence and environment become crucial to performance which is characterised by its multi-nodal nature.

Analysing the degree of independence the individual nodes may assume in a network, they distinguish between the following three approaches:

- projected dramaturgy (one node acts as author and the others as contributors);
- directed dramaturgy (authorship remains with an individual or group who take on the role of director);
- distributed dramaturgy (each node retains authorship).

These categories also apply to the nearly 30 compositions that have been written for Quintet.net and the EBE since 2000 (see https://en.wikipedia.org/wiki/European_Bridges_Ensemble for a selection thereof).

Praxis of networked music performance

In 2007, two EBE members became partners in the European Culture 2007 project Cooperation and Mediation in Digital Arts (CO-ME-DI-A) with the aim of exploring the potential of NMP (http://e-b-e.eu). The opening three-way concert between Belfast, Graz and Hamburg was

called *Disparate Bodies 2.0*, named after the eponymous piece by Pedro Rebelo of Queens University's Sonic Arts Research Centre in Belfast (Hajdu, 2008).

Cage's *Five* was also included in the programme, except this time Quintet.net was solely used to display the score in real-time, while the five performers played the music on acoustic instruments. This performance thus added a significant new component to our praxis of NMP: multi-channel, high-fidelity, low-latency audio streaming via JackTrip – with all participating partners being connected to the GÉANT backbone (the European equivalent of Internet2). This software relying on the Jack Audio Kit (http://www.jackaudio.org) is based on streaming technologies in development since 2000 and is maintained by Stanford University's SoundWire Group (Cáceres & Chafe, 2010).

Adding low-latency, high-fidelity audio and video streaming to the equation increased the sense of presence for all participants. However, the EBE was actually pursuing a different route – a route that was to embrace a *virtual presence* of the performers. Miller Puckette (2008) in his article 'Not Being There' (!) also argues in favour of systems that, instead of (or in addition to) increasing the sense of presence by ever-refined streaming technologies, are based on the transmission of control messages, stressing the potential of game-like interactions by the participating parties:

> We quickly decided that instead of doing telepresence, … we would try to do something that better responded to the idea of the physical distance between the sites. Instead of sending sounds and images across the network, we opted to send real-time audio analyses of the instruments. … it was possible to associate any of the four instruments with a 'ghost' in the remote location, possibly very similar, but possibly quite different.
>
> (Puckette, 2008, p. 409)

In *Radio Music* (yet another adaptation of a piece by Cage), I devised for the five performers of Quintet.net an interface representing the physical actions of turning a USB rotary controller and acting as a dial for the virtual radio (Figure 14.2), for which I coined the term *radio avatar*. The audience experienced these actions in form of the combined movements of five cursors running along the display of virtual radio stations. At a performance at SIGGRAPH Asia with performers playing from Yokohama, Budapest and Belgrade the animation generated much interest among the audience. Despite the fact that three of the performers were in remote locations, there was no doubt that these actions were authentic.

EBE gave their last concert in November 2014 featuring Ádám Siska's piece *2_45* in which Wii Remote Controllers were used to simulate the actions of a ping-pong game. The piece was originally composed for a networked performance between the neighbouring capitals of Vienna and

Figure 14.2 Georg Hajdu's adaptation of *Radio Music* uses a *radio avatar* to represent the actions of the (remote) musicians.

Bratislava, featuring a variety of sounds that were supposed to be 'tossed' to and fro via Wii Remote Controller strikes.

Thanks to wireless technology employed to send continuous control data to the individual computers, the players, exaggerating their movements according to Siska's choreography, were able to move about freely while observing a computer screen. On the screen, they were following a slider indicating the arrival of an object to strike. A successful strike was translated into a sonic representation, either iconic sound samples of crashes or spoken words. This piece had all the key ingredients of a successful (network) performance: Unbound by cables or idiosyncratic controllers the players were free to explore their perceptual-motor skills while feeling situated in a sonic environment that paid attention to small details of their interaction with the machine (such as simulating the sound of the approaching object). As a result of the game/sports metaphor, the spectators were able to quickly construct a mental image of the digital musical interaction and were able to identify the meaning of the sonic events generated by the remote participants. This piece was both engaging and entertaining, and thus did not require AV streams or an accompanying video playback to maintaining interest. After nearly ten years of exploratory work, the ensemble had set standards for NMPs in terms of aesthetic experience and performativity.

The dialectical strive towards hyper-realism

The CO-ME-DI-A project was concluded in 2010 with a series of showcase events. A three-way, three-day network festival between Belfast, Graz and Hamburg epitomised the technical and aesthetical progress achieved

throughout the first decade of the millennium: 45 uncompressed audio streams and five video streams ran with surprisingly few dropouts, delivering realistic multi-channel sound and video projections of the partner stages and audiences in addition to the video artwork on the screen in the centre. One of the highlights on the second day was the performance of Luigi Nono's composition *A Pierre. Dell'Azzurro Silenzio, Inquietum* for contrabass flute, contrabass clarinet and live electronics by Carin Levine (playing from Belfast) and Carola Schaal (playing from Hamburg). The quality of the audio and video streams as well as the unnoticeable latency created a deep sense of presence, which, for the performers and spectators alike, overrode the perception of medium's technicality and allowed an immersive experience, that is the audience was allowed to focus on the aesthetical experience rather than the technical means facilitating this experience.

In December 2010, another showcase event took place in Pécs, Hungary, featuring the third Music in the Global Village conference/symposium. The symposium invited select participants of the CO-ME-DI-A project and attempted to draw conclusions on NMP in light of the experiences gained throughout the project. There was a consensus that low-latency video streaming had not reached the same potential as audio streaming: in her demonstration during the lecture session 'Strategies for Finding Artistically Meaningful Applications of Internet Technology', Andrea Szigetvári of the Hungarian Computer Music Foundation harshly criticised *tunnel* vision and lack of true immersion caused by cameras and video streams at the time, the small resolutions either offering too small a viewing area or too little detail, and thus seriously limiting the amount of visual information available to performers and audience alike.

Another issue was the lack of standardisation. While the CO-ME-DI-A project had originally set out to develop a *CO-ME-DI-A bus*, into which different media streams were to conveniently be plugged in and received at each receiving end, this was abandoned, as the technical requirements were just too different for each scenario. Instead, each performance had to be configured by hand, which required lengthy preparation and rehearsals.

Nonetheless, the CO-ME-DI-A project hinted convincingly that technologies might eventually mature enough to gradually overcome most if not all of these hurdles. Currently, Ambisonics and wave-field synthesis spatialisation technologies offer a high degree of immersive realism in the audio domain, while 360-degree spherical video streams viewable with head-mounted displays have just become a reality. Therefore, in combination, holographic and 'holophonic' projection techniques (with the possible inclusion of haptic, temperature, olfactory and balance cues) will eventually offer unprecedented realism, which will come as close to *teleporting* people as one can imagine. Using 3D CGI-technologies, it is already possible to create realistic animated models of people (absent or deceased) and to place them in the midst of real-musicians on stage such as with the holographic projection of John Coltrane in the 2002 movie *Vanilla Sky*.

However, I called this section 'dialectical strive towards virtual hyper-realism', as art and design do not always follow the same paths as technological advances and might even occasionally oppose them. This opposition is often fuelled by power and economic structures in which elite technology is only available to a select few. But sometimes it is just about aesthetics: an example is Apple's move away from skeuomorphs, that is realistic user interface (UI) elements in their iOS operating system. Ableton has always favoured simple and abstract UI designs in the software Live while Reason, another digital audio workstation, which realistically mimics a studio rack. In the same vein, we will see continued interest in low-tech network music performances such as with Live Coding, both in local and remote networked scenarios. A memorable project in this area is the Powerbooks Unplugged laptop ensemble (http://pbup.net/s/), who promote themselves mockingly as the 'first acoustic computer music folk band: the laptop is their only instrument'.

Conclusions

In this chapter I have shown that NMP represent a special case of digital musical interaction in which participating musicians may be hidden from sight or their presence only be experienced through the use of streaming technologies. Cognitive science provides the notions (such as embodiment/disembodiment) to identify the nature of man–machine interaction in networked performances and thus contributes to the improvement of the physical, sonic and spatial aspects – themes that are key to this book. The first decade of the first millennium saw a rapid co-evolution of technological and artistic solutions specific to this new genre, with the objective to create an ever more palpable musical experience for audiences and performers alike – the software Quintet.net and the pieces created and/or performed by the EBE being exemplary for this co-evolution.

Many properties of NMP have been aestheticised by the artists, for instance by using network delays as structural elements, motivating the metaphorical use of the term embodiment such as in the case of Tanaka's Global String project. While multimedia technologies continue to be developed, eventually offering unprecedented realism and immersion, the aim of NMP remains to create an audience involvement fundamentally different from classical music performances. These performances take machine–performer–spectator interactions into consideration, which, to a great deal, rely on embodied cognition and the sense of causality. From the audience's perspective, digital music interaction, particularly in NMP, poses a challenge as it builds its effect on illusion. Hence, classical cause-and-effect relationships (which also permeate the 'genuine' musical sign of the index) are replaced by plausibility, that is the amount to which performers and spectators are capable of 'buying' the outcome of a performance by building mental maps of the interaction. In NMP, this can be facilitated by the use of avatars, projected visually, and carefully orchestrated dramaturgies, involving participants in game-like scenarios.

Note

1 In this context one should also mention the participatory piece *isms* by my former student Jacob Sello written for the European Bridges Ensemble (Lee & Freeman, 2013).

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109–1116.

Alexandraki, C., & Bader, R. (2014). Using computer accompaniment to assist networked music performance. In *Proceedings of the AES 53rd International Conference* (pp. 27–29), London, UK.

Applebaum, M. (2010). *Aphasia* [score]. Retrieved from http://londontreatise.blogs.wesleyan.edu/files/2011/06/APPLEBAUM-Aphasia1.pdf.

Bartók, B. (1976). *Turkish folk music from asia minor.* Princeton: Princeton University Press.

Brown, C., & Bischoff J. (2002). Indigenous to the net: Early network music bands in the San Francisco bay area. Retrieved from http://crossfade.walkerart.org/brownbischoff/hub_texts/points_of_presence_f.html.

Cáceres, J.-P., & Chafe, C. (2010). JackTrip: Under the hood of an engine for network audio. *Journal of New Music Research, 39*(3), 183–187.

Cadoz, C. (2009). Supra-instrumental interactions and gestures. *Journal of New Music Research, 38*(3), 215–230.

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine, 22*(4), 67–84.

Chafe, C. (2009). Tapping into the internet as a musical/acoustical medium. *Contemporary Music Review, 28*(4/5), 413–420.

Chafe, C., Cáceres, J.-P., & Gurevich, M. (2010). Effect of temporal separation on synchronization in rhythmic performance. *Perception, 39*(7), 982–992.

Föllmer, G. (2005). *Netzmusik. Elektronische, ästhetische und soziale Strukturen einer partizipativen Musik.* Hofheim: Wolke-Verlag.

Freeman, J. (2008). Extreme sight-reading, mediated expression, and audience participation: Real-time music notation in live performance. *Computer Music Journal, 32*(3), 25–41.

Gazzaniga, M. S., Ivry, R., & Mangun, G. R. (2014). *Cognitive neuroscience: The biology of the mind*. New York: W.W. Norton.

Gurevich, M., & Fyans, A.C. (2011). Digital musical interactions: Performer–system relationships and their perception by spectators. *Organised Sound, 16*(2), 166–175.

Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics, 65*(5), 789–800.

Hajdu, G. (2003). Quintet.net – Präliminarien zur Entwicklung einer vernetzten, interaktiven Echtzeitkompositionsumgebung. In B. Enders & J. Stange-Elbe (Eds.), *Global village – global brain – global music* (pp. 298–303), Osnabrück: Rasch.

Hajdu, G. (2005). Quintet.net: An environment for composing and performing music on the internet. *LEONARDO, 38*(1), 23–30.

Hajdu, G. (2008). Getrennte Körper – Musiker im Netz. *Positionen 74 – Dispositive*, 29–31.

Johnston, A., Candy, L., & Edmonds, E. A. (2009). Designing for conversational interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 207–212), 4–6 June, Pittsburgh, USA.

Karbusický, V. (1986). *Grundriss der musikalischen Semantik*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Kim, J. H., & Seifert, U. (2007). Embodiment and agency: Towards an aesthetics of interactive performativity. In *Proceedings of 4th Sound and Music Computing Conference* (pp. 230–237), 11–13 July, Lefkada, Greece.

LaBelle, B. (2015). *Background noise – perspectives on sound art*. London: Bloomsbury.

Lanier, J. (2001). Virtually there: Three-dimensional tele-immersion may eventually bring the world to your desk. *Scientific American*, *284*, 66–75.

Lee, S. W., & Freeman, J. (2013). Echobo: A mobile music instrument designed for audience to play. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 450–455), 27–30 May, Daejeon, South Korea.

McLuhan, M., & Powers, B. R. (1989). *The global village: Transformations in world life and media in the 21st century*. New York: Oxford University Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

Puckette, M. (2008). Not being there. *Contemporary Music Review, 28*(4/5), 409–412.

Rebelo, P., Schroeder F., & Renard, A. B. (2008). Network dramaturgy: Being on the node. In *Proceedings of the International Computer Music Conference* (pp. 551–552), Belfast, UK.

Wallin, N. L., Merker, B., & Brown, S. (Eds.) (2000). *The origins of music*. Cambridge, MA: MIT Press.

Weidenaar, R. (1995). *Magic music from the telharmonium*. Metuchen, NJ: The Scarecrow Press.

Wöllner, C. (2012). Self-recognition of highly skilled actions: A study of orchestral conductors. *Consciousness and Cognition, 21*, 1311–1321.