Interactivity for Mobile Music-Making

GEORG ESSL* and MICHAEL ROHS†

Deutsche Telekom Laboratories & TU-Berlin, Ernst Reuter Platz 7, D-10587 Berlin, Germany
E-mail: *georg.essl@telekom.de, †michael.rohs@telekom.de

Mobile phones offer an attractive platform for interactive music performance. We provide a theoretical analysis of the sensor capabilities via a design space and show concrete examples of how different sensors can facilitate interactive performance on these devices. These sensors include cameras, microphones, accelerometers, magnetometers and multitouch screens. The interactivity through sensors in turn informs aspects of live performance as well as composition though persistence, scoring, and mapping to musical notes or abstract sounds.

1. INTRODUCTION

Mobile devices are both ubiquitous and computationally potent. Additionally, many are naturally designed to be sound-generating devices, whether for conveying speech signals for telephony or prerecorded music for mobile entertainment. And they have an increasingly versatile array of sensor capabilities which can be used to design interactivity for both live and pre-meditated music performance. These sensor capabilities range from push-buttons, through microphones and built-in cameras to accelerometers and multitouch screens. Hence it is natural to ask if these devices make good generic platforms for interactive music performance devices (Essl, Wang and Rohs 2008). And in answering this question affirmatively, one is led to a further concern: how do these technological choices impact and inform emerging musical practice?

In this paper we follow the traces of our own work in enabling interactive music performances on mobile phones, including new developments in engaging with the sensory and computational capabilities of multitouch mobile devices such as iPhones.

A primary component of our approach is the systematic study of the sensor capabilities through a design space. A design space is a taxonomy that suggests early on what mappings are possible given the characteristics of sensors. We then look which kinds of music performances are well served by the available technical capabilities.

The development in this area has very much been driven by what is technologically possible. First we study camera-based interactions. While it may appear counterintuitive that computer-vision-based 2D marker detection developed before gesture-based interaction, this has to do with the broad availability of built-in cameras in smart phones before it became more widespread to also embed accelerometers. Indeed, technological development has been rapid, and since the start of our work in the area in 2005 we already have gone through multiple steps of new sensor capabilities becoming incorporated into mainstream commodity devices. It is not clear if this trend will continue or if we are reaching saturation in input sensor modalities on typical mobile devices. Nevertheless, we feel this is the right moment to describe the development up to 2009. Also, at the same time we have already seen an array of possible musical uses for, and different types of interactivity in, these technologies.

The paper is organised primarily in chronological order. After discussing related work, we will describe four steps in the development of interactivity for mobile phones from the camera-based CaMus of 2005 to the multitouch based Fendrix of 2009 and discuss the musical uses of these different technologies.

1.1. Related work

Turning mobile devices into musical instruments has already been explored by a number of researchers. Tanaka presented an accelerometer-based custom-made augmented PDA that could control streaming audio (Tanaka 2004). Geiger designed a touch-screen-based interaction paradigm with integrated synthesis on the mobile device using a port of Pure Data (PD) for Linux-enabled portable devices like iPaqs (Geiger 2003, 2006). Various GPS-based interactions have also been proposed (Strachan, Eslambolchilar, Murray-Smith, Hughes and O'Modhrain 2005; Tanaka, Valadon and Berger 2007). Many of these systems used an external computer for sound generation.

Using a mobile phone as physical musical instrument has been pioneered by Greg Schiemer (Schiemer and Havryliv 2006) in his PocketGamelan instrument. At the same time there has been an effort to build up ways to allow interactive performance on commodity mobile phones.

The attempt to turn portable gaming platforms into generic sounding devices is in fact rather old. The original GameBoy inspired a fairly generic music performance platform called nanoloop, developed by Wittchow (Behrendt 2005), though the commercialisation of mobile music-making programs is fairly recent. For example JamSessions, a simple guitar-strumming

Organised Sound **14**(2): 197–207 © 2009 Cambridge University Press. Printed in the United Kingdom. doi:10.1017/S1355771809000302

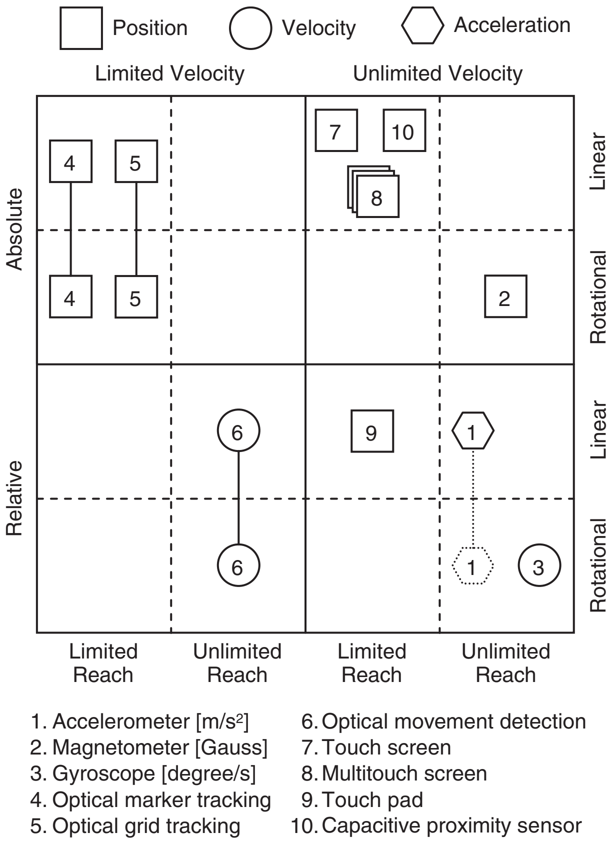

Figure 1. Design space of sensors for mobile music performance.

game for the Nintendo DS, was released in the summer of 2007.[1]

Mobile music performance is only rarely discussed conceptually or analytically. One of the primary sources is still Behrendt's book (Behrendt 2005), though it considers mobile music in a much wider scope than we do here. Mobile music is seen as any technology that allows new music to be performed while mobile. A more recent review of work and theory along these lines can be found in Gaye, Holmquist, Behrendt and Tanaka (2006).

1.2. Designing interactions from sensor capabilities

There are excellent references available reviewing sensor technologies for the design of new interfaces for musical expression in detail (Miranda and Wanderley 2006). Here we give only a brief summary of the characteristics that are relevant for design decisions in the context of making mobile handheld devices into interactive performance instruments. Sensor capabilities form the technological foundation of the interactivity of a musical instrument. Hence, despite the undue technicality of these descriptions we consider them important from a conceptual perspective as well, in the same sense that a good composer of orchestral music has to know the capabilities and limitations of orchestral instruments. In order to help conceptualise the properties of sensors, we structure these characteristics into a design space of sensor technologies (figure 1).

The layout of our design space of sensor technologies is inspired by Ballagas, Rohs, Sheridan and Borchers (2006) and adapted to the requirements of mobile music creation. Whereas the design space presented in Ballagas et al. (2006) is more geared towards the result of an interaction (position, orientation, selection), the most relevant factors for our design space are the kinds of movements that can be detected (linear or rotational), the maximum velocity that can be sensed, and the maximum physical interaction range.

This physical range is relevant because it in turn forms what Delalande called the effective gesture component of the musical gesture of the performer (see Miranda and Wanderley 2006). According to Delalande, the further two aspects of a musical gesture are accompanist gestures (actual gestures which have a non-effective quality, such as related head and shoulder movements) and figurative gestures (which are assumed by the audience but not physically existing gestures). The technology indirectly also

[1] Details about JamSessions can be found at http://www.ubi.com/US/Games/Info.aspx?pId=5560.

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

influences these other aspects of musical gesture, by providing the possible gesture space for accompanist gestures and by suggesting context for figurative gestures.

The sensors we consider measure static position and orientation, velocity or acceleration. They operate either with a fixed outer frame of reference or relative to previous sensor states and with no outer frame of reference.

This classification with respect to outer frames of reference is important conceptually because having access to a fixed reference offers access to the absolute. This has a multitude of musical consequences. Absolute measures can be stored, can also be more easily repeated and reproduced, and relate well to musical qualities which themselves are absolute such as absolute pitch or loudness. An absolute reference is, however, not a necessity. Expressive gestures can be derived from relative measures as well.

Movements can be linear or rotational, in one or more dimensions. Some sensors can measure linear as well as rotational movements (indicated in figure 1 by a connecting line). In the context of musical performance, maximum velocity and reach are important. The body of the performer prescribes a meaningful space for interaction, and sensor technology for body-centred performance is only relevant when it works well within the space that is reachable by the performer. For some sensors the maximum velocity is constrained by the technology: for example, with optical sensing at low frame rates. Other sensors can detect much faster movements than human beings are able to produce. Hence, we call this category *unlimited velocity*.

Musically the given maximum velocity directly dictates what kinds of musical phrases are technologically possible. If the speed required to perform a phrase is faster than the technology it becomes inaccessible.

The *reach* dimension denotes the maximum extent of the physical interaction space that the sensor is able to cover. For optical marker and grid tracking, reach is limited by the extent of the grid and the maximum recognition distance. For touch screens, reach is limited by the size of the screen, and for proximity sensors the limit is the maximum sensing distance. Musically, reach has multiple meanings. For the performer it defines the confinement of the space of performance. Large reach may in fact not be a desirable quality because it means extra body motion to perform. From the point of view of the audience, large reach means that gestures can be chosen that form a larger region of space and that can be more easily observed and associated with sound.

We do not use the *direct/indirect* dimension (Ballagas et al. 2006), because there is no strong notion of spatial coincidence or separation in auditory output. Audio output could be described as *environmental* or *ambient*. As discussed in Ballagas et al. (2006), interactions might be continuous or discrete. *Continuous* sensing with auditory and visual feedback is, for example, suitable for compositional interfaces. *Discrete* sensing is best suited for generating one-time sound events.

1.2.1. Optical tracking of markers and marker grids

Cameras integrated in handheld devices are well suited for optical tracking. The position and orientation of the device can be sensed quickly and with low delay relative to a single visual marker (Hansen, Eriksson and Lykke-Olesen 2005; Rohs and Zweifel 2005) or to a grid of markers (Hachet, Pouderoux, Guitton and Gonzato 2005; Rohs, Essl and Roth 2006). Tracking fixed markers with mobile devices enables *absolute positioning*: the physical marker establishes a frame of reference in which the interaction takes place. A single marker provides a relatively small physical space for tracking; a marker grid enables a larger tracking space. In the implementation presented in Rohs et al. (2006) the grid establishes an interaction space of up to 150 × 150 × 30 cm. However, the need for a fixed marker or marker grid can sometimes be an issue, because it limits mobility.

Next we briefly discuss the sensor technologies we consider in the design space taxonomy. Readers who are not interested in specific details but rather in the complete interaction instruments and their musical uses can skip to Section 2.

1.2.2. Optical movement detection

Integrated cameras can also be used for optical movement detection without markers (Rohs 2005; Wang, Zhai and Canny 2006). Optical movement detection is a *relative positioning* method and does not provide a fixed frame of reference. The physical interaction space is in principle unlimited, but the tracking background has to be slightly textured. A disadvantage is that the user's movement velocity is limited by the frame rate of today's phone cameras. If the user moves too fast, it is difficult to compute optical flow. Another problem is that the amplitude of movement of the device does not correspond in a one-to-one fashion to changes in the interface. The same device movement velocity at different background distances will thus yield different velocity measures. Linear as well as rotational movements can be sensed with this method.

1.2.3. Acceleration sensing

Accelerometers are a widely used sensor technology to detect motion. Their main advantage is that they are very cheap, come as small IC units and are

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

already showing up in commodity hardware. Several Nokia phone models, Apple's iPhone, and the Wii game console all contain 3-axis accelerometers. The main disadvantage is the lack of a reference frame other than gravity. Continued integration of acceleration to get velocity and then displacement integrates the noise of the sensor and leads to inevitable drift. Hence accelerometers cannot easily sense absolute motion in all directions, but are very well suited for sensing relative motion. Accelerometers are capable of sensing very fast motions, thus not limiting the human user's velocity of movement.

1.2.4. Magnetic field sensing

Magnetometers sense the magnetic field. These also come as small integrated units and are fairly easily accessible. They have, to the best of our knowledge, not made it into commodity mobile handheld devices. They are reasonably cheap and, once calibrated, rather accurate, as long as the immediate environment does not have strong electromagnetic interference (EMI). Magnetometers provide absolute rotational position.

1.2.5. Gyroscopes

Gyroscopes use the spinning-top effect to sense changes in angular motion. These are very high-fidelity sensors that come in reasonably small integrated units. They are, however, currently rather pricey, which may be the main reason why they are not seen in consumer hardware. Gyroscopes are robust and stable and provide precise readings on fast rotational gestures. The do not provide an absolute frame of reference.

1.2.6. Touch and multitouch screen

Touch screens have been very successful as an input technology for mobile devices. Touch screens for mobile devices are typically implemented as analogue resistive or capacitive surfaces. The technology is cheap, has low power consumption and high resolution, and allows for pen and finger input. Resistive touch screens are not affected by dust or water, which makes them suitable for mobile outdoor use. Some resistive surfaces can sense the amount of touch pressure.

Matrix analogue touch screens can sense two or more locations simultaneously. Apple's iPhone is able to sense multiple touch points and multitouch gestures. Thus it increases the flexibility of user input by moving away from a single point of input towards multifinger gestures and whole-hand input. Recent research tries to use the raw input data to drive physical models of widgets on the screen, allowing for a wide variety of hand gestures in order to generate the modelled forces and velocities.

1.2.7. Capacitive proximity sensing

Capacitive sensing uses the capacity between the skin of a hand or finger and a conductor as input. These devices are technologically very simple and cheap to build but have the problem that they are sensitive to electromagnetic interference as well as changes in the conductivity of the skin (for example due to sweat) or the air (due to change of humidity). Hence, they need to be calibrated and show somewhat different readings for different users due to variation in personal skin conductivity.

1.2.8. Microphones

Microphones are not movement-based sensors, as the ones we described above. Therefore they do not fit well into the design space diagram for movement-based sensors. It could be argued that microphones provide absolute input, because the input energy can be measured in an absolute way. Also, localisation is possible to a certain extent by tapping onto or near the microphone. Hence the microphone can be used as a position-sensing device in a limited way. When scratch sounds are sampled (the sliding sounds of two objects) then a microphone can be used to sense velocity by the frequency of the generated sound (Murray-Smith, Williamson, Hughes and Quaade 2008).

2. BUILDING MOBILE MUSICAL INSTRUMENTS

Enabling live music performance on mobile devices is critically defined by the capability of the device itself. This is very much akin to turning computers or laptops into live music performance devices. All these are general computing devices with specific system limitations and input and output modalities. But certainly there are critical differences between these systems. Mobile devices for the first time offer a kind of ubiquitous technological mobility. Laptops are portable yet it is a burden to carry one everywhere. Mobile phones are generally carried around, like wallets or keyings. Their form gives them an ongoing presence and availability that laptops have not reached. However, due to the form factor there are certain limitations imposed on mobile devices that laptops suffer to a much lesser extent. The size of the laptop is primarily defined by the size of a standard keyboard. Hence full and rapid keyboard input is a defining interaction paradigm for the laptop. At the same time, the laptop offers a sensibly large screen space, thus work well with the predominant paradigm of 2D graphical user interface with mouse interaction. Mobile phones lack the space for a full-size keyboard and either substitute it for a 12-key telephone dialling block or downsized and compacted versions of the standard keyboard. This limits textual

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

typing as an attractive input modality (MacKenzie and Tanaka-Ishii 2007).

Even the largest-screen mobile phones such as the iPhone (7.4 cm × 5.1 cm) are over a factor of 10 smaller than typical laptop screens (28.5 cm × 21.5 cm for 14-inch screens). This puts limits on the graphical user interface paradigms that are currently in place for laptop and desktop computers and in relation also alters the input methods.

However, even given the limitations mobile devices often offer more sensory input than laptops, by having built-in accelerometers, GPS sensors and cameras. Mobile phones certainly usually come with quality audio input/output (I/O) capabilities.

These differences in form factor are akin to the limitations of performance mobility by classical musical instruments. A piccolo or a clarinet is more mobile than an orchestral harp, timpani or a grand piano. By that reasoning it is easier to prefigure a mobile-phone-based new-music marching band than it is to hope to achieve this on a laptop platform.

The MobileSTK port of Perry Cook's and Gary Scavone's Synthesis Toolkit (STK) (Cook and Scavone 1999) to Symbian OS (Essl and Rohs 2006) is the first full parametric synthesis environment available on mobile phones. It was used in combination with accelerometer and magnetometer data in ShaMus (Essl and Rohs 2007) to allow purely on-the-phone performance without any laptop. STK also runs on iPhones with minimal changes. This means that one can indeed consider current mobile phone technology to be capable of serving as generic digital sound synthesis devices. Any digital sound can be played on them, but more interestingly it can in principle also render any interactive and parametric synthesis algorithm, assuming that enough computational power is available. This is currently still a practical hurdle. Contemporary mobile phones are not yet capable of rendering complex synthesis algorithms. However, many basic standard algorithms – such as FM synthesis, additive synthesis or concatenative synthesis – are possible.

The availability of accelerometers in programmable mobile phones such as Nokia's N95 or Apple's iPhone has created a technology that enables us to more fully consider mobile phones as meta-instruments for gesture-driven music performance. The main idea of the mobile phone as a meta-instrument is to provide a generic platform on which the composer can craft his or her artistic vision (Essl et al. 2008). We are currently at a stage where ensemble-based mobile-phone orchestras are possible (Wang, Essl, Penttinen 2008). The advent of the iPhone has brought a whole host of mobile-phone-based music-making applications like iBand, PocketGuitar or Scratch, showing a growing interest in mobile music-making commercially. Our interest is in mobile music-making as a frontier for new forms of musical expression, in an artistic and academic context. In this sense our outlook differs from many of the available commercial programs, which emulate existing instruments often limiting expressivity and novelty of expression.

2.1. CaMus: camera-based interactions

CaMus uses the built-in camera of smart-phones to enable motion-based interaction. The prototype was implemented on Nokia 6630 smart phones (Rohs et al. 2006; Rohs and Essl 2007).

We implemented two ways to allow motion detection via the camera. One is a two-dimensional marker recognition method that derives coordinates, height, rotation and tilt relative to a marker sheet. The other is a movement-detection approach which does not require specific optical reference material, but is more prone to error due to overly homogeneous or poor lighting conditions. The technology allowed for interactive navigation in a virtual space that was larger than the screen of the mobile phone. Due to the absolute coordinates offered by the marker sheet, position in the virtual space can be precisely identified. There is a limit to tilt gestures, which is the primary reason not to attempt free-motion gesture interactions with this technology. Instead, we explored the predominantly two-dimensional space created by the marker sheet as an interactive editing and manipulation environment. In a way, this emulates a live graphical score that can be manipulated on the fly. Sound sources and sound effects can be placed on different layers of a two-dimensional plane. The distances to these sources are sent as parameters to MIDI-based sound synthesis engines and sequencers, and affect height, rotation and tilt. This leads to a natural cross-mixing of sources and effects in the plane where relative position of sources define the strength of the mix.

2.1.1. Visual grid tracking technology

Interaction takes place on a grid of visual markers, which are derived from visual codes. The grid represents a large workspace, of which different parts can be accessed with a camera phone by simply placing it over the relevant area (see figure 2). The phone display acts as a window into the virtual workspace. The grid defines a coordinate system that provides an absolute frame of reference for the spatial interaction (see figure 2). Printed on paper, it typically extends over a DIN A4 or A3 sheet, but the size of the sheet is in principle flexible. The prime determinant of coordinates are the size of a single black-and-white cell, but sub-pixel resolution can be achieved by detecting the position of that cell within the camera image, allowing camera-resolution-bound precision of coordinate detection. Details of the algorithm can be found in Rohs et al. (2006). In the interaction space

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

established by the marker grid, we can precisely determine the position and orientation of the phone at a rate of up to 10 frames per second with our prototype device (a Nokia 6630 Symbian phone), depending on the complexity of the rendered virtual workspace. On newer devices this rate improves. The grid has to allow for a smooth continuous detection of position and orientation during movement at high precision. It also provides the basis for a perspective rendering of the virtual space as shown in figure 3. Multiple detected markers within the field of view are combined to additionally increase the stability of detection.

2.1.2. Optical flow

The optical movement detection algorithm turns the camera phone into an optical mouse. The algorithm is a refinement of the method described by Rohs et al. (2006). It detects relative linear movement of the phone in the display plane and relative rotational movement of the phone around the optical axis, representing three degrees of freedom. The algorithm operates on the video stream at the frame rate offered by the camera, which can be 10 fps or higher. The algorithm subdivides each frame into a block image, computes cross-correlations between successive pairs of block images for a range of different shift and rotation offsets, and looks for the maximum correlation. Details of the algorithm can be found in Rohs and Essl (2007).

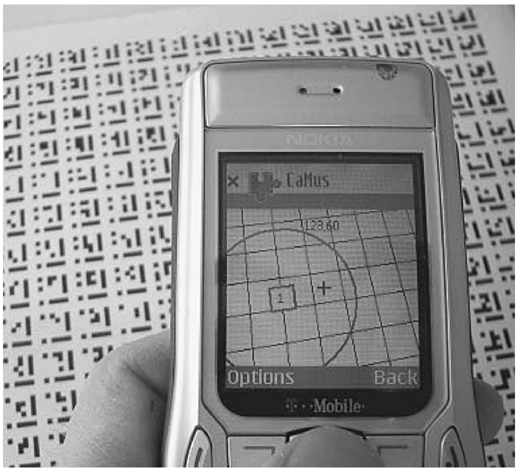

Figure 2. CaMus uses the built-in camera over a two-dimensional marker sheet to allow motion in a virtual space.

2.1.3. Interactivity of CaMus

CaMus as an interactive system has specific characteristics. In particular, the virtual space in which the interactions are placed has an absolute reference. This allows configuration of the virtual space to be stored and recovered, leading to a permanence that can be used for composition, authoring and scoring. Targets, filter effects and sounding sources can be placed, stored and recalled at absolute positions in the virtual two-dimensional space. This makes the virtual space into the equivalent of a score, with similar characters of permanence. The relation of sources to sounding results in CaMus is very general. While the motion in the plane is continuous, one can discretise the relationship and hence arrive at performance of discrete musical pitches. One can say that the system does not prescribe the particular musical style, though it favours continuous motion and hence continuously changing sound. Furthermore the system allows what we would traditionally call interpretation of a composed piece. The composer defines the setup of the two-dimensional virtual space. The performer can then follow the score and use it as a base for an interpretation with a fine level of nuance. For the audience the type of visible motion and reach are important. The positions of the marker sheet serve as an absolute reference but also limit the range of gestures that are possible with this technology. By replacing the marker sheet with the optical flow method, one immediately gets drift problems, due to the loss of an absolute reference while freeing up the gesture repertoire. The speed of recognised motion and the response time is defined by the frame rate of the camera, which is between 10 to 30 frames per second. This is good for slow to moderately fast motion but too slow for very rapid changes. Because

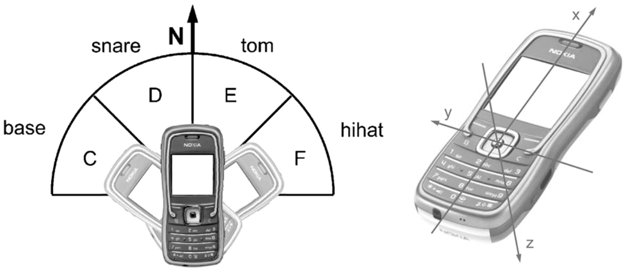

Figure 3. (Left) ShaMus is an accelerometer-based interaction model that can be augmented by magnetometer sensing to allow angular positions to be used for pitches or different types of sounds. (Right) The magnetometer allows for angular-position-based performance.

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

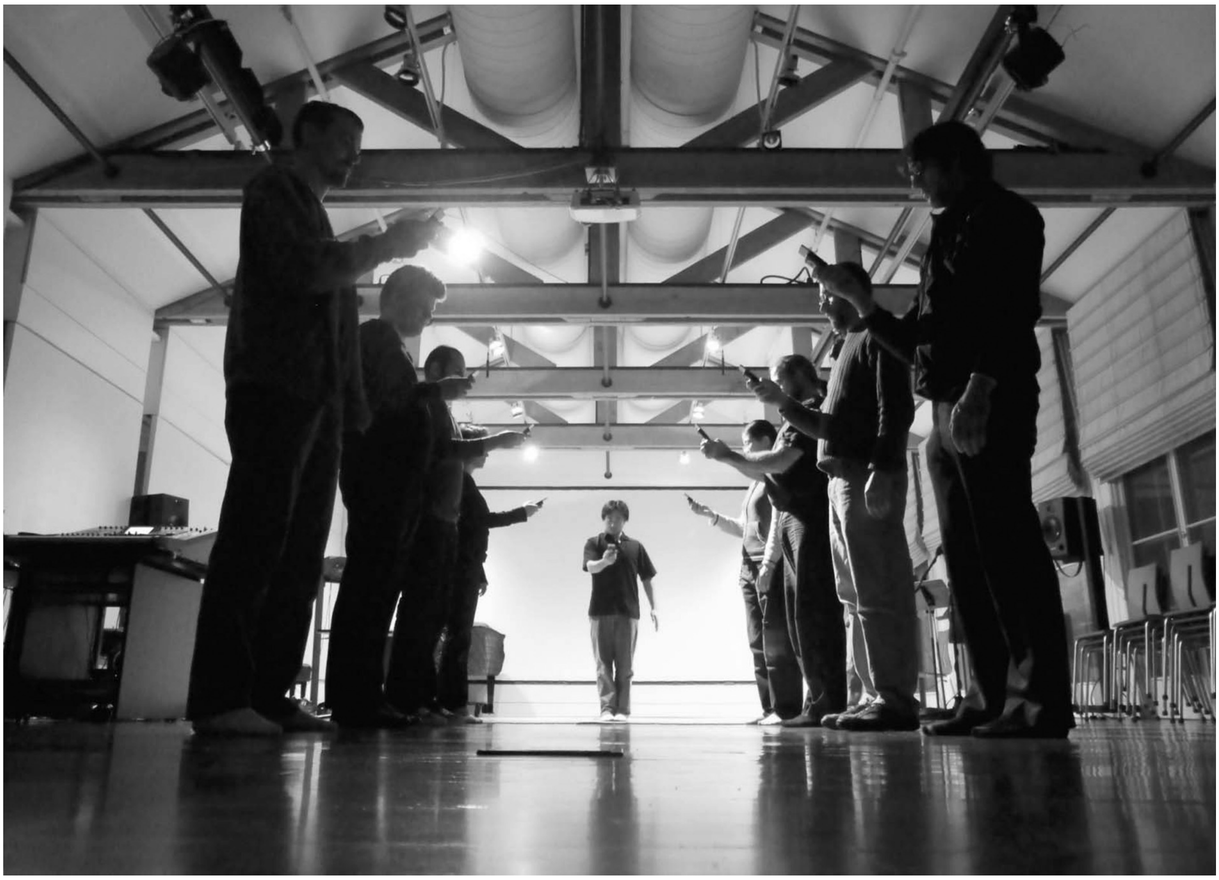

Figure 4. The Mobile Phone Orchestra of CCRMA conducted by Ge Wang in January 2008.

of the technology's need to reference of the marker sheet, the expressive gestures of the performer, as they are conveyed to the audience, are limited and the speed of recognition contributes to this perception. We used the system primarily to remix prescored musical pieces and found the type of interactivity possible with the system to work well for this type of use. Generically, the system favours slowly evolving continuous musical compositions that can be pre-authored, recalled and manipulated during performance.

2.2. ShaMus: motion-sensor based interactions

One of the main limitations of CaMus was the virtual tethering of the system to a marker sheet and the limit of free hand gestures due to the limits of the tilt angle detection of the system. This led to a muted expressive gesture for both the performer and the audience. A natural candidate sensor to get around these limitations is the accelerometer, as it is cheap and offers direct tilt readout without angular restrictions. The main downside of accelerometers is the inability to get reliable position information, as integration of acceleration leads to accumulation of noise in the signal and hence to uncontrollable drift. ShaMus explores the use of accelerometers for gesture-based interactions with mobile devices. With the Nokia 5500 Sport phone and the N95 phones, accelerometer sensors became available in commodity telephones. Figure 3 shows a typical arrangement of the coordinate system of the accelerometers. Many new phones that have come on to the market since then come with or can offer accelerometers for interaction, including the iPhone and the HTC Dream.

ShaMus allows the mapping of accelerometer data to realtime rendering of synthesis algorithms on the phone itself. Hence, the mobile phone becomes both untethered from any input supporting material such as marker sheets, but also unlinked from an external sound rendering system. This kind of technology is very easy to set up and disseminate and led to the formation of the first repertoire-based mobile phone ensemble (figure 4) founded at the Center of Computer Research in Music and Acoustics at Stanford University (Wang, Essl, Penttinen 2008).

2.2.1. Accelerometer conditioning

Luckily, in contemporary mobile devices the quality of the integrated accelerometers has drastically improved, which was not true for early models (Essl and Rohs 2007). The main need for conditioning of

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

the received signal remains the sensitivity to fine disturbances on the sensor, which can be motion from jitters in the hand or other disturbances such as the steps of passers-by or traffic. Hence, there is a need to take out high-frequency content, which can be achieved with cheap low-order low-pass filters.

2.2.2. *Interactivity in ShaMus*

The accelerometer data itself offers a wide array of interactions. The sensor data can be directly mapped to synthesis algorithms, leading to tilt-based interactions. But the sensor data can also be used to drive gesture-recognition software. That way, for example, striking gestures can be implemented. Due to Newton's law there is an immediate relationship between acceleration and force, so any physically based interaction that relates to force can easily be implemented. This means that very expressive effective gestures can be used. There is no inherent limitation to reach, so the performer can use all motions that are ergonomically possible or pleasant. The speed of accelerometers is very fast (~ 1000 Hz) meaning that for musical purposes there is no perceptible delay or impact on performance speed. Accelerometers can be used for very fine and rapid motions. The main limitation lies with the lack of a reliable position reference. This favours music that does not rely on fixed references, such as smooth modifications of sounds that are already fixed otherwise. In ShaMus we tried to cope with this limitation by adding custom sensor technology (SHAKE) to detect magnetic field strength through magnetometers. This allows detection of absolute angular position relative to the earth's magnetic field (see figure 3 (right)). With this technology it is possible to take angular position into account and hence implement air-drumming or pitched play by striking different positions in the air. However the problem with this approach is that magnetic field sensors have not established themselves as part of commodity mobile phones.

Accelerometer data can be sampled at very high rates: 100–1000 Hz is typical. This means that even very rapid changes can be detected without noticable delay. Magnetometers show similar response rates. The main limiting factor on the easy use of magnetometers is the dependency of geographic location on the absolute magnetic field data, hence requiring calibration every time the physical location is changed significantly. Overall this is, however, a very expressive way to allow continuous play. Due to the difficulty of getting an absolute reference, pitched play is difficult and it is hard to define virtual scores for this technology.

2.3. Keys as interaction modality

Most mobile devices come with an array of buttons and keys. These offer an obvious and natural mode of interaction. For one, keys can be directly linked to synthesis engines and serve as a pitched keyboard. They can also be used to change the mapping of other sensors to synthesis results or manipulate parameters on the fly. While technologically buttons are rather trivial, these remain important interaction elements for musical instruments. They are well suited to control discrete quantities such as musical pitches and they allow for impact-like onsets such as striking or plucking. But keys can also be used in indirect ways to modify the meaning of other sensors. For example a key press can change the the fundamental frequency or spectrum of a continuously playing sound controlled smoothly by accelerometers. It is easy to store and recall key interactions, and hence to compose virtually for them.

2.4. MiMus: microphone-based interactions

Mobile phones are designed to take sound as a primary input to allow for voiced communication. But the audio signal captured by the microphone can be used in a broader sense, as a general input sensor (O'Modhrain and Essl 2004; Essl and O'Modhrain 2005; Misra, Essl and Rohs 2008). Defining characteristics of the microphone as sensor is the overall high fidelity and large dynamic range. It is a sensor capable of picking up ranged signals. Aside from literal recording of audio signals, one can also derive semantic or gestural information: for example, one can detect blowing or striking gestures. The complete configuration of the MiMus setup can be seen in figure 5. These types of interactions have already made it into commercial products. For example Sonic Lighter, an iPhone application, allows a flame's sound and motion to be manipulated by blowing into the microphone.

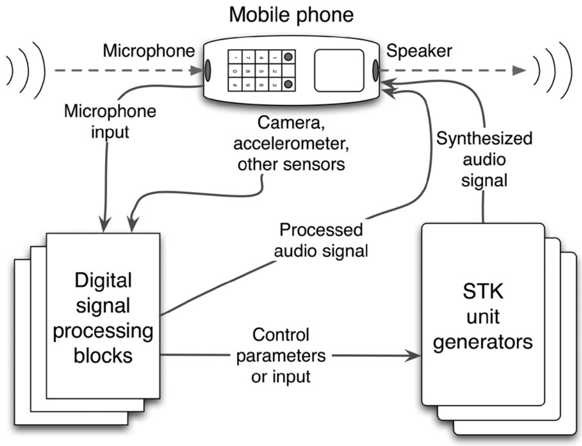

Figure 5. MiMus is a microphone-based interaction model that uses the microphone input in abstracted form for musical performance.

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

2.4.1. Interactivity of MiMus

The quality of microphone-based interactions is intimately linked to the quality of the audio I/O hardware of the device and the complexity and properties of the audio analysis algorithms employed. Ideally, one wants a low-latency full-duplex audio I/O hardware which allows access to the incoming digitised audio signal at the individual sample level or at small buffer sizes, and at the same time allows seemless low-latency playback of audio. In reality, the audio I/O layer of mobile devices can vary drastically. While the audio layer of Nokia 6630 offered buffer sizes of just 320 bytes, Nokia 5500 sport phones required a minimum of 4096 bytes, with over an order of magnitude increase in associated latency.

Which audio analysis algorithms can be realistically employed depends on two aspects: (1) the computational power of the device and (2) the inherent delay of the algorithm itself. The computational power of the device determines how much computation can be achieved within the time span between two audio buffers needing filling. The limitation of this factor can be expected to diminish with newer phones.

Microphone signals can be pitched if absolute information (loudness, spectral content) is used, but it can also be continuous as is the case with changing amplitude envelopes, for example. In general, microphone signals are as general as audio itself and the act of recording serves as the simplest instance of "scoring" the microphone. The sensor is very fast so there is no impact of performance speed. However, it is not always obvious that blowing into a mobile device is a meaningful activity, so the audience may not consider the effective gesture of playing a mobile device as virtual wind instrument as actual.

2.5. Fendrix: multitouch-screen-based interactions

The addition of the multitouch screen to the iPhone is the main advancement over earlier smartphones. The multitouch screen offers a rather rich set of interaction types. For one, two-dimensional local and moving positions can be detected for multiple fingers. Also there is explicit support for timed tapping on the screen. Hence one gets not only continuous input via the moving coordinates of touch points but also discrete impulsive interactions via tapping on the screen at any location.

Fendrix adds this rich interaction space to the acceleration-based interactions of ShaMus. Coordinates, touch onset and offset can be mapped to synthesis parameters. There are numerous examples of applications already available on iPhone's application store, which emulate traditional instruments. Most of these try to capture the interaction paradigm of that instrument via the multitouch screen. There are piano applications which allow multikey play, though without "aftertouch", that is the ability to use strike intensity. Some guitar applications offer the shape of a stringed guitar neck and hence faithful fingering. There are a number of percussive instruments available. However there are few examples which use the seemless continuous nature of the display. An example in this category would be a turntable emulation (Scratch) which allows smooth motion of a virtual turntable surface.

Our interest lies with offering the multitouch sensor as a generic input modality, that can also allow new forms of musical expression that are not anchored in the emulation of traditional musical instruments. Hence in Fendrix we take the input space of the multitouch screen literally and map coordinates to synthesis parameters without relying on analogies to existing musical instruments (see figure 6 (left)). This is akin to musical performance on touchpads like the Lemur or Wacom tablets (Arfib, Couturier, Filatriau and Kessous 2006). One can then go and give meaning to the geometric space. Fendrix allows multiple fingers to be used. An example mapping is to associate with each finger a "voice". This is similar to a piano. However one can move the fingers freely on the touchpad and hence traverse parameters of the synthesis algorithm freely, which is not possible on the piano. In this sense it is more closely related to the play on a fretless string instrument like a violin. However instead of moving in a one-dimensional parameter space (the position on a string), we move in a two-dimensional parameter space (the position on a plane). Furthermore pitches are still also discretely separated on stringed instruments by the number of strings themselves, and fingers usually only occupy one string. We have all fingers occupy the same space and in fact one can decide that the parameters are interchangeable. If one swaps the positions of the fingers on the screen their roles reverse.

To be more specific, each finger acting on the multitouch screen creates a set of parameters. As depicted in figure 6 (right), we can index these parameters by the finger, hence getting coordinates x_n, y_n and s_n, with n being an integer describing the finger being put down; x_n and y_n are plainly the coordinates on the touch screen; s_n is a state vector associated with the touch event. It can encode that a touch event started, that a touch is currently moving or if a touch event stopped. It can also encode rhythmic information such as how many taps have been counted at this coordinate point. The visual display can be used in various ways in conjunction with the touch interaction. For one, it can provide visual elements that suggest what affordances are possible at various positions on the screen (e.g. by displaying a keyboard or strings or other familiar musical instrument elements), or it can be used to give visual feedback about incoming touch events. In Fendrix we chose the latter option, and traces

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

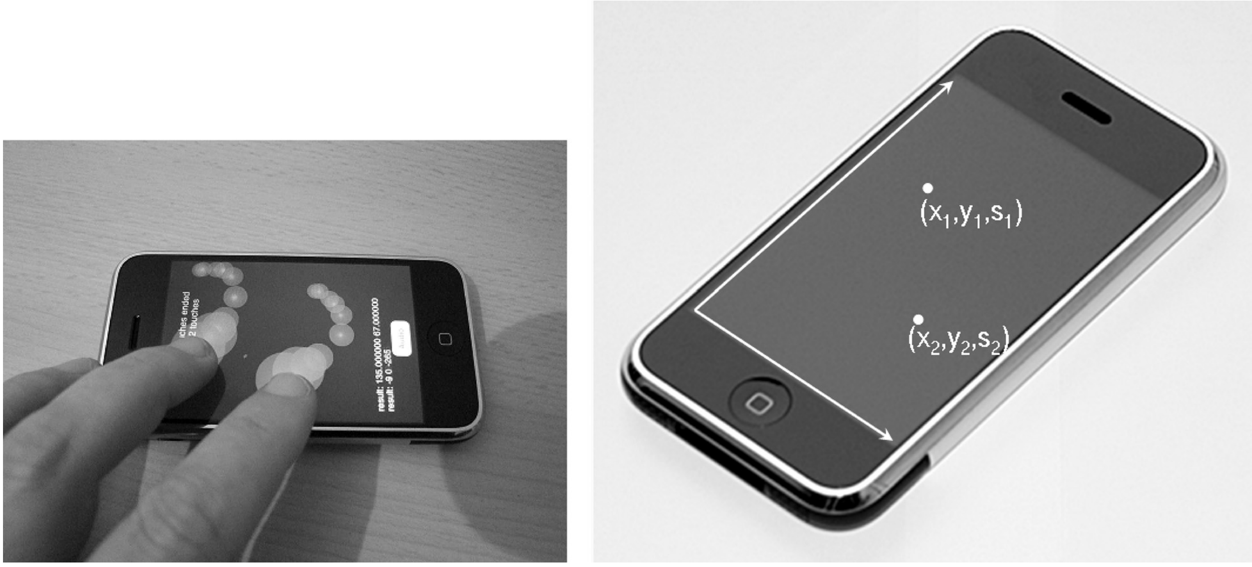

Figure 6. (Left) Fendrix is a multi-sensor instrument implemented on the iPhone. Multiple fingers can trigger multiple continuously controlled voices. (Right) Coordinate space of the iPhone multitouch interface.

of touch interactions will augment the gestures performed on the touch screen (see figure 6).

A key component of interaction with the iPhone is that other interesting sensor modalities are concurrently available. In Fendrix we also use the built-in three-axis accelerometer data for further input to synthesis algorithms. It is important to note that physically these two modalities are not completely decoupled. Motion but especially impact on the touch-screen will cause raw accelerometer data changes. This can be countered by filtering or masking methods. However, from a perspective of interaction design one can also link the meaning of one sensor with the meaning of another. For example, tilt can non-linearly interact with touch-screen coordinates, or can be discretised to offer the selection of different configurations of the touchscreen parameter mapping. Such cross-linked input behaviour is by no means unusual in musical instruments. Specifically in wind instruments, the fingering and blowing pressure are interrelated.

Fendrix incorporates the desirable features of the expressivity of continuous gestures of accelerometers with the specificity and absolute reference in 2D space that we earlier discussed with visual tracking of marker sheets. Hence Fendrix allows both continuous expressive gestures and a space for scoring and premeditated composition for both continuous sound and pitched music.

3. CONCLUSIONS

Contemporary mobile phones offer an increasing array of sensor technologies that can be used for rich interactions for musical performance. In this paper we traced the development of the literature on enabling interactive musical performance on mobile phones since 2005 until 2009, including recent developments with the emerging multitouch capability of mobile phones such as the iPhone. Alongside their sensory capabilities, the acoustic playback of these devices becomes more powerful, and complex synthesis algorithms become feasible. This makes an interesting generic use of mobile phones as ubiquitous and novel interactive musical devices possible. Contemporary sensors allow a broad musical use of these platforms, both for free-form improvisations and for composition, scoring and interpretation.

An interesting and important aspect of mobile technology is networking. We do not yet address this aspect, as many technological aspects that are important for performative use are in their infancy. Specifically, network performance needs easy and seamless connectivity as well as participation; ad hoc joining and leaving of local networks is as yet cumbersome due to security concerns. Yet we believe that networking will play an important role in mobile music-making. We also leave out locative performance, mostly because we believe that this is an inherently different form of music-making. Details about this emerging field is reviewed in Gaye et al. (2006). Finally, we still have a long way to go in terms of diversity and depth of systems available. As we argued elsewhere, we need many different ways to allow expressivity (Essl, Wang and Rohs 2008). But there is no doubt that already now mobile phones offer interesting new ways of making interactive music.

REFERENCES

Arfib, D., Couturier, J. M., Filatriau, J. J. and Kessous, L. 2006. What Sounds Can I Do with a Joystick and a

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press

Tablet? *Proceedings of the 2nd International Symposium on Gesture Interfaces for Multimedia Systems (GIMS2006)*, Leeds UK.

Ballagas, R., Rohs, M., Sheridan, J. G. and Borchers, J. 2006. The Smart Phone: A Ubiquitous Input Device. *Ieee Pervasive Computing* **5**(1): 70–7.

Behrendt, F. 2005. *Handymusik. Klangkunst und 'mobile devices'*. Epos. Available online at: www.epos.uos.de/music/templates/buch.php?id=57.

Cook, P. and Scavone, G. 1999. The Synthesis ToolKit (STK). *Proceedings of the International ComputerMusic Conference*, Beijing.

Essl, G. and O'Modhrain, S. 2005. Scrubber: An Interface for Friction-induced Sounds. *Proceedings of the Conference for New Interfaces for Musical Expression*, Vancouver, Canada, 70–5.

Essl, G. and Rohs, M. 2006. Mobile STK for Symbian OS. *Proceedings of the International Computer Music Conference*, New Orleans, 278–81.

Essl, G. and Rohs, M. 2007. ShaMus – A Sensor-Based Integrated Mobile Phone Instrument. *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen.

Essl, G., Wang, G. and Rohs, M. 2008. Developments and Challenges turning Mobile Phones into Generic Music Performance Platforms. *Proceedings of the Mobile Music Workshop*, Vienna.

Gaye, L., Holmquist, L. E., Behrendt, F. and Tanaka, A. 2006. Mobile Music Technology: Report on an Emerging Community. *NIME '06: Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, 22–5.

Geiger, G. 2003. Pda: Real Time Signal Processing and Sound Generation on Handheld Devices. *Proceedings of the International Computer Music Conference*, Singapore.

Geiger, G. 2006. Using the Touch Screen as a Controller for Portable Computer Music Instruments. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'06)*. Paris, France.

Hachet, M., Pouderoux, J., Guitton, P. and Gonzato, J.-C. 2005. TangiMap: A tangible Interface for Visualization of Large Documents on Handheld Computers. *GI '05: Proceedings of the 2005 Conference on Graphics Interface*, Waterloo, Canada, 9–15.

Hansen, T. R., Eriksson, E. and Lykke-Olesen, A. 2005. Mixed Interaction Space: Designing for Camera Based Interaction with Mobile Devices. *Proceedings of CHI '05: Extended Abstracts on Human Factors in Computing Systems*, 1933–6.

MacKenzie, I. S. and Tanaka-Ishii, K. (eds.) 2007. *Text Entry Systems: Mobility, Accessibility, Universality*. San Francisco: Morgan Kaufmann Publishers.

Miranda, E. R. and Wanderley, M. M. 2006. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. Middleton, WI: AR Editions.

Misra, A., Essl, G. and Rohs, M. 2008. Microphone as Sensor in Mobile Phone Performance. *Proceedings of the International Conference for New Interfaces for Musical Expression (NIME–08)*, Genova, Italy.

Murray-Smith, R., Williamson, J., Hughes, S. and Quaade, T. 2008. Stane: Synthesized Surfaces for Tactile Input. *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, ACM, New York, 1299–1302.

O'Modhrain, S. and Essl, G. 2004. CrumbleBag: Tactile Interfaces for Granular Synthesis. *Proceedings of the International Conference for New Interfaces for Musical Expression (NIME)*, Hamamatsu, Japan.

Rohs, M. 2005. Real-World Interaction with Camera Phones. In H. Murakami, H. Nakashima, H. Tokuda and M. Yasumura (eds.) *Second International Symposium on Ubiquitous Computing Systems (UCS 2004), Revised Selected Papers*, LNCS 3598, Tokyo, Japan, 74–89.

Rohs, M. and Essl, G. 2007. Camus2 – Collaborative Music Performance with Mobile Camera Phones. *Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE)*, Salzburg, Austria, 13–15.

Rohs, M. and Zweifel, P. 2005. A Conceptual Framework for Camera Phone-Based Interaction Techniques. In H. W. Gellersen, R. Want and A. Schmidt (eds.) *Pervasive Computing: Third International Conference, PERVASIVE 2005*, LNCS 3468, Munich, Germany, 171–89.

Rohs, M., Essl, G. and Roth, M. 2006. CaMus: Live Music Performance using Camera Phones and Visual Grid Tracking. *Proceedings of the 6th International Conference on New Instruments for Musical Expression (NIME)*, 31–6.

Schiemer, G. and Havryliv, M. 2006. Pocket Gamelan: Tuneable Trajectories for Flying Sources in Mandala 3 and Mandala 4. *NIME '06: Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, 37–42.

Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S. and O'Modhrain, S. 2005. GpsTunes: Controlling Navigation via Audio Feedback. *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, Salzburg, Austria.

Tanaka, A. 2004. Mobile Music Making. *NIME '04: Proceedings of the 2004 conference on New Interfaces for Musical Expression*, 154–6.

Tanaka, A., Valadon, G. and Berger, C. 2007. Social Mobile Music Navigation using the Compass. *Proceedings of the International Mobile Music Workshop*, Amsterdam.9

Wang, G., Essl, G. and Penttinen, H. 2008. MoPhO: Do Mobile Phones Dream Of Electric Orchestras? *Proceedings of the International Computer Music Conference (ICMC–08)*, Belfast, Northern Ireland.

Wang, J., Zhai, S. and Canny, J. 2006. Camera Phone Based Motion Sensing: Interaction Techniques, Applications and Performance Study. *UIST '06: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, New York, 101–10.

https://doi.org/10.1017/S1355771809000302 Published online by Cambridge University Press