The Design Space of Sensing-Based Interaction for Mobile Music Performance

Georg Essl
Deutsche Telekom Laboratories
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
georg.essl@telekom.de

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
michael.rohs@telekom.de

ABSTRACT

Active music performance can mean composition and interpretation of composed music or improvisation. In this paper we discuss the design requirements to make mobile handheld devices into actively engaged expressive musical instruments for these different types of performance. Candidate sensors are discussed and compared within this design space.

1. INTRODUCTION

The goal of this work is to turn mobile handheld devices with integrated sensors into interfaces for active musical performance. In order to assess the design space of this goal we compare a range of available sensors for the suitability in typical western music performance practices. The sensors we consider are either already available in commodity mobile devices or are available in modular extensions to them.

Current mobile phones are more and more used and marketed as multimedia devices in general and portable music players in particular. Handset manufacturers see music capabilities as an important market segment[1] and increasingly develop and market their mobile phones as digital music players.[2] Today's phones have large storage capacities, include improved audio codecs for high-quality music reproduction, and provide development APIs to access multimedia capabilities.

However, today's phones only offer passive music playback. With the integration of sensors, like accelerometers, or by using the cameras integrated in many mobile phones, users can be more actively engaged in the mobile music experience. Sensing capabilities have been used for some time in mobile devices to enable new kinds of user interfaces [2, 6, 15]. We discuss how various kinds of sensors can detect a variety of gestures for musical expression with a mobile device. These interactions allow users to influence interpretation parameters of pre-recorded pieces of music, such as adding and modifying sound filters. They also allow to take more direct control and turn mobile devices into musical instruments. This work is part of this general project and implementations and user studies have been presented elsewhere [3, 12].

After presenting the design requirements relevant for music creation activities we describe suitable sensors and map them in a design space. We then discuss how these sensors can be applied in the mobile music design context and conclude with a short summary and ideas for future work.

[1] www.forum.nokia.com/music

[2] www.apple.com/iphone, en.wikipedia.org/wiki/IPhone

2. DESIGN GOAL

In western music performance culture [8] music creation activities often are separated into different roles and activities. We have a strong tradition of authored music. This role falls to the *composer*, who sets out the intent of the piece as well as its overall structure and progression. Often a general intent for interpretation is also notated. Sometimes the authored pieces of music are performed by the composer themselves. This is not necessarily so, and it's common that pieces are performed by multiple other people, who then fill the *performer* role. There are however different forms of performing music. An alternative route is improvised music, usually based on a commonly agreed upon set of rules and structures. For this reason we classify the activities of music creation and performance as *composition*, *interpretation* and *improvisation*. Like many classifications, the boundaries are fuzzy and in some cases a conceptual separation is not without ambiguity. For example, the pre-set rules and structures of improvisation could be thought of as "composed". We think that the categories are however still distinct enough in character to warrant their separation.

- Composition

  The main characteristic of a composition is that it is a deliberately pre-authored piece of music meant for later performance. In order to allow for the persistence of the intent of the musical idea, it is helpful to have means of displaying the structure of the intent in accessible form to others or the composer himself. In traditional western music the typical mechanism to do this is called the *score*. The score serves as a medium for communicating the ideas, but also as a way to store relevant information. Composition without score or storage except for the memory of the composer is thinkable. The first, for example in the case of tape pieces that were authored straight into playback representation without abstracted visual representation, does not really address the topic of live music performance that interests us here. The latter is a fusion of the composer with the performer removing, and by choice precluding, external representation and sharing of musical ideas outside hearing the music itself.

Name	Type	Resolution	Sample Rates	Range	Noise	Reliability
Accelerometer	Electromechanical	High (1 mg [7])	0.5-2 kHz	±6g	Low	High
Magnetometer	Electromagnetic	High (1 mGauss [7])	1 kHz	±2 Gauss	Medium	Medium
Gyroscope	Electromechanical	High (0.1 deg/s [7])	$\geq$ 1 kHz	±500 deg/s	Low	High
Marker/grid tracking	Optical	High [11]	15-30 Hz	150×150×30 cm	Low	High
Movement detection	Optical	Medium [11]	15-30 Hz		Medium	Low
Touch screen	Electromechanical	High		4×5 cm	Low	High
Capacitive proximity	Electrostatic	High	1 kHz [7]	0-10 mm	Low	Medium

Table 1: Comparison of the technical characteristics of sensors.

- Interpretation

  A performer who is given a representation of a musical piece to then play it, engages in an act of interpreting the piece. This task involves reading the intent of the composer, contributing independent intent and turning the existing representation into an expressive piece of music. Interpretation can be tight, with an attempt to try to reach the composer's intent as best as possible. It can also be loose, where the performer takes the flow of the piece and varies the detail, or the intent during the performance. If much of the original piece is used as a guiding structure, this then turns into improvisation which we treat separately. In order for this to be possible, the interpreter needs an accessible representation and means of creating expression and variation of the given material.

- Improvisation

  A performer who improvises, uses a given structure (even if very loose or ad hoc, like a tonal system and scale set or chord progression) to create music live and on the fly. Here the expressivity of the instrument, immediacy of the ability to turn ideas into musical results and feedback for both performer and audience are important.

In general we look to make mobile handheld devices into musical instruments, but we are aware that different types of choices to allow the needed interactivity supports or limits these different roles and activities in different ways.

The design goal is to find implementations that support these different activities either in separate implementations or in a common system.

3. CHARACTERISTICS OF SENSORS

There are excellent references available reviewing sensor technologies for the design of new interfaces for musical expression in detail [9]. 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.

- 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 relative to a single visual marker [5, 13] or to a grid of markers [4, 12]. 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 grid enables a larger tracking space. In our current implementation of [12] the grid extablishes an interaction space of up to 150×150×30 cm, in which devices determine their position with high accuracy and low latency. The method uses small markers with a data capacity of 14 bits each, arranged in a regular grid. The digital zoom capability, which is available in many camera phones, is used to extend the vertical interaction range. Digital zoom does not gain optical resolution, but essentially provides high-quality rescaling without involving the main processor of the device. This helps to keep the grid markers in an optimal size range for recognition. However, the need for a fixed marker or marker grid can sometimes be an issue, because it limits mobility.

- Optical movement detection

  Integrated cameras can also be used for optical movement detection without markers [11, 14]. The method described in [11] subdivides camera frames in blocks and looks for maximum correlations between successive images by trying different offsets. 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. The method does not work on uniform surfaces, like a gray carpet. A disadvantage is that the user's movement velocity is limited by the low frame rate of today's camera phones. To compute relative movement there has to be some overlap between successive images. If the user moves too fast, there is no overlap and no result can be computed. Another problem is that the amplitude of movement of the device does not correspond in a one-to-one fashion to changes the interface. The detected movement velocity depends on the distance of the background to the lens. The same device movement velocity at different background distances will thus yield different velocity measures.

  Like all vision-based systems, optical tracking and optical movement detection depend on sufficient lighting. With current CCD sensors the operable lighting range is quite broad. A special characteristic of optical tracking is that the camera image can be interpreted by both human and machine and can thus convey the semantics of an operation to the user.

- Acceleration sensing

  Accelerometers are a widely used sensor technology to detect motion. Their main advantage is that they are

Name	Context	Constraints	Cost	Commodity
Accelerometer	Jolt-free environments	Drift	Low	Beginning
Magnetometer	Low EMI	Requires calibration	Low	Low
Gyroscope	Any	N/A	High	Very Low
Marker tracking	Sufficient lighting	Distance, range	Medium	High
Movement detection	Sufficient lighting	Velocity, drift	Medium	High
Touch screen	Any	Screen size	Medium	High
Capacitive proximity	Low EMI	Drift	Low	Low

Table 2: Comparison of the further characteristics of sensors.

very cheap, come as small IC units and are already showing up in commodity hardware. For example the Nokia 5500[3] mobile camera phone contains 3-axis accelerometers and the Wii game console[4] use the same technology. Apple's iPhone will contain accelerometers to automatically align the screen depending on the direction the device is held. The main disadvantages are the lack of a reference frame other than gravity. Continued integration of acceleration to get velocity and then displacement also integrates all noise of the sensor and leads to inevitable drift. Hence accelerometers cannot easily be used for absolute motion. They are, however, very good for sensing relative motion.

- 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).

- Gyroscope

  Gyroscope sensors use the spinning top effect to sense change 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 and are also rarely used in research projects. These are very robust and stable and provide very good reading on rotational gestures on the device, independent of position.

- Touch screen

  Touch screens have been very successful as an input technology for mobile devices. Touch screens for mobile devices are typically implemented as analog resistive surfaces. The technology is cheap, has low power consumption, 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 analog touch screens can sense two or more locations simultaneously. Apple's iPhone is able to sense multiple touch points and multi-touch gestures. Ergonomic problems include stress on human fingers and limited interaction range.

- Capacitive proximity sensing

  Capacitive sensing uses the capacity between the skin of a hand or finger to a conductor of the sensor as input. These 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. Once calibrated they do give high resolution proximity sensing. Despite their very low cost they are not typically found in consumer electronic devices.

In our actual implementation we looked at the Shake[5] [7] device. It's a small device designed to incorporate a range of high-fidelity sensors for rapid prototyping of mobile interactions. The core unit contains 3-axis accelerometers, 3-axis magnetometers, a vibration motor for vibro-tactile display, and switch and capacitive sensing abilities. All analog data is sampled to 12 bit resolution at 1 kHz.

The layout of our design space of sensor technologies is inspired by [1] and adapted to the requirements of mobile music performance. Whereas the design space presented in [1] 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.

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 no outer frame of reference. 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. 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*. 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 it is limited by the size of the screen, and for proximity sensors

[3] `forum.nokia.com/devices/5500`

[4] `wii.com`, `en.wikipedia.org/wiki/Wii`

[5] `www.samh-engineering.com`

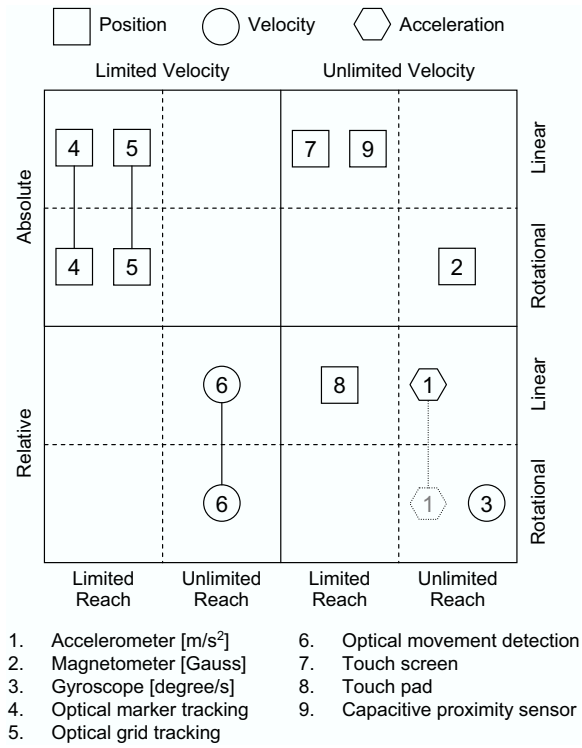

Figure 1: Design space of sensors for mobile music performance.

the limit is the maximum sensing distance.

We do not use the *direct/indirect* dimension [1], because there is no strong notion of spatial coincidence or separation in auditive output. It is generally difficult to precisely locate sounds. Audio output could more naturally be described as *environmental* or *ambient* feedback. As discussed in [1] 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.

4. COMPARISONS OF SENSORS IN THE DESIGN CONTEXT

We seek to find sensor technologies for various types of music creation and performance. These are composition, interpretation and improvisation.

- Composition

  For composition, we need persistent storage and an absolute reference so intent can be transported from the composer to the performer. Of the sensor technologies we have discussed, optical tracking of markers and touch screens provides the most immediate absolute reference frame and a convenient way to set up scoring and authoring.

  We have implemented two versions of the optically based system in a project called CaMus [12], one using a marker grid to give an absolute position reference and one using optical flow for movement sensing. The former was chosen to allow score-like precomposition and storing behavior. Sounding and sound manipulation elements can be placed in a virtual plane and manipulated. The essential state can be fixed and restored. Objects retain their relative position if not interacted with. This allows for both pre-authoring (composition) and interpretation (variation on pre-authored music).

  Magnetometers and gyroscope can be thought to allow composition. In undisturbed fields, magnetometers give a good reference with respect to the earth's magnetic field and provide compass information to the interface. This can be used for rotational gestures or as an absolute reference for rotational gestures. The main problem with this absolute reference is that it is not the same as a performer would usually expect. Most performers see their own orientation as the right frame of reference and not the externally imposed one of the magnetic field. Regardless one can envision using this to store information relative to the reference and hence compose to this setup. For example the composer could choose to fix the meaning of north or east. Gyroscopes are reasonably stable but in conjunction with a rotational reference (as for example given by a magnetometer) can be much improved. Gyroscopes hence have all the good properties of magnetometers but do not have the necessary limitation of an external reference.

  Accelerometers and optical flow suffers from drift problems making the establishment of a reference difficult. Hence they are not really immediately suitable for a compositional setup. Capacitive proximity sensing is similar. Here environmentally induced drift makes it difficult to fix a clear absolute reference. Accelerometers can be used to measure tilt relative to gravity. However this tilt sensing is only accurate if there are no other relative motions in the direction of the gravitational field. For practical purposes the tilt sensing of accelerometers is sensibly stable to give reproducible references for composition.

- Interpretation

  The presence of pre-authored content is a is prerequisite for allowing interpretation. Hence we can either take those sensor technologies that allow this, or augment such system with additional sensing channels to provide additional expressivity. High frame or sample rate sensors provide the best potential for expressivity. Hence accelerometers, gyroscopes, capacitive sensing and also touch screens provide good means for expression. Optical systems are currently limited somewhat by their frame rate.

  Hence for a strongly interpretive system either a touch screen implementation alone, or a compositional system augmented by fast sensors are thinkable.

- Improvisation

  For improvisation we do not need to fix content, but need to give the performer both control through feedback and expressivity.

  Accelerometers, gyroscopes and magnetometers allow a wide range of gestures, that additionally are not me-

diated by the geometry of the sensor itself, as is the case with touch screens. This makes these sensors attractive candidates for hand-held mobile interface designs for musical improvisation. While optical flow provides the same freedom in gesture, the technology is limited by its frame rate.

Touch screens can be very expressive and are already in use in a multi-touch context. The type of gestures are fixed to the screen, which can also be seen as an advantage of defining a performance space.

Capacitive sensing is already used in the Theremin and other instrument designs (see [9] for a review) but in general these are hard to control due to the environmental drifting problem and the lack of haptic feedback.

All these requirements limit the options for a universal instrument for musical expression, that is an instrument which would be equally suitable for composition, interpretation and improvisation. As the single most suitable technology, touch screens seem most appropriate. This is not to say that this would be ideal. As O'Modhrain has shown, affordance is an important component in performance and expressivity of musical instruments [10]. The affordance of acting on a touch screen is somewhat limited. This favors the use of accelerometers or other sensors that use the affordance imposed by the weight of the mobile device itself to mediate performance.

5. CONCLUSIONS

Music is an essential part of our everyday live. Mobile phones with integrated sensors, like cameras and accelerometers, will be ubiquitously available and can support new kinds of music performance experiences. They enable users to share their musical experiences with their friends and take a more active part in music performance.

In this paper we compared a range of available sensors for the suitability in typical western music performance practices. Visual tracking and touch screens are attractive candidates for compositional devices, whereas analog sensors like accelerometers, gyroscopes, or magnetometers lead to sensitive and high fidelity interpretive and improvisational performance.

As future work we plan to integrate our existing camera tracking system with available continuous sensing to increase the expressiveness and create a system that is both suitable for authoring and interpreting music on the fly and allows the performer to engage in free and expressive improvisation.

6. REFERENCES

[1] R. Ballagas, M. Rohs, J. G. Sheridan, and J. Borchers. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1):70–77, 2006.

[2] S. Benford, H. Schnädelbach, B. Koleva, R. Anastasi, C. Greenhalgh, T. Rodden, J. Green, A. Ghali, T. Pridmore, B. Gaver, A. Boucher, B. Walker, S. Pennington, A. Schmidt, H. Gellersen, and A. Steed. Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Trans. Comput.-Hum. Interact.*, 12(1):3–30, Mar. 2005.

[3] G. Essl and M. Rohs. Mobile STK for Symbian OS. In *Proceedings of the International Computer Music Conference*, pages 278–281, New Orleans, November 2006.

[4] M. Hachet, J. Pouderoux, P. Guitton, and J.-C. Gonzato. TangiMap: A tangible interface for visualization of large documents on handheld computers. In *GI '05: Proceedings of the 2005 Conference on Graphics Interface*, pages 9–15, Waterloo, Canada, 2005.

[5] T. R. Hansen, E. Eriksson, and A. Lykke-Olesen. Mixed interaction space: Designing for camera based interaction with mobile devices. In *Proceedings of CHI '05: extended abstracts on Human factors in computing systems*, pages 1933–1936, 2005.

[6] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz. Sensing techniques for mobile interaction. In *Proceedings of User Interface Software and Technology (UIST)*, pages 91–100. ACM Press, 2000.

[7] S. Hughes. *Shake – Sensing Hardware Accessory for Kinaesthetic Expression Model SK6*. SAMH Engineering Services, Blackrock, Ireland, 2006.

[8] M. J. Kartomi. *On Concepts and Classifications of Musical Instruments*. The Univeristy of Chicago Press, Chicago, 1990.

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

[10] S. O'Modhrain. *Playing By Feel: Incorporating Haptic Feedback into Computer-Based Musical Instruments*. PhD thesis, Stanford University, 2000.

[11] M. Rohs. Real-world interaction with camera phones. In H. Murakami, H. Nakashima, H. Tokuda, and M. Yasumura, editors, *Second International Symposium on Ubiquitous Computing Systems (UCS 2004), Revised Selected Papers*, pages 74–89, Tokyo, Japan, July 2005. LNCS 3598, Springer.

[12] M. Rohs, G. Essl, and M. Roth. CaMus: Live music performance using camera phones and visual grid tracking. In *Proceedings of the 6th International Conference on New Instruments for Musical Expression (NIME)*, pages 31–36, Paris, France, June 2006.

[13] M. Rohs and P. Zweifel. A conceptual framework for camera phone-based interaction techniques. In H. W. Gellersen, R. Want, and A. Schmidt, editors, *Pervasive Computing: Third International Conference, PERVASIVE 2005*, pages 171–189, Munich, Germany, May 2005. LNCS 3468, Springer.

[14] J. Wang, S. Zhai, and J. Canny. Camera phone based motion sensing: interaction techniques, applications and performance study. In *UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 101–110, New York, NY, USA, 2006. ACM Press.

[15] S. Zhai and V. Bellotti. Introduction to sensing-based interaction. *ACM Trans. Comput.-Hum. Interact.*, 12(1):1–2, 2005.