An enactive approach to the design of new tangible musical instruments

GEORG ESSL† and SILE O'MODHRAIN‡

†Deutsche Telekom Laboratories & TU-Berlin, Ernst Reuter Platz 7, D-10587 Berlin, Germany
E-mail: Georg.essl@telekom.de
‡Sonic Arts Research Centre, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland
E-mail: sile@qub.ac.uk

In this paper, we propose a theoretical framework for the design of tangible interfaces for musical expression. The main insight for the proposed approach is the importance and utility of familiar sensorimotor experiences for the creation of engaging and playable new musical instruments. In particular, we suggest exploiting the commonalities between different natural interactions by varying the auditory response or tactile details of the instrument within certain limits. Using this principle, devices for classes of sounds such as coarse grain collision interactions or friction interactions can be designed. The designs we propose retain the familiar tactile aspect of the interaction so that the performer can take advantage of tacit knowledge gained through experiences with such phenomena in the real world.

1. INTRODUCTION

Interaction with objects in the world around us is a richly multi-sensory experience. Casting a pebble into a pond, we both see the ripples resulting from the disturbance of the water's surface and hear the impact of the stone on the water as a disturbance of the air. If we are close enough and the stone is big enough, we might also get wet. Furthermore, the interaction of stone and water makes certain information explicit – the size of the splash is correlated with both the size of the stone and the force with which it was thrown, and the sound it makes provides information about the depth of the water. Thus the physical laws that govern the behaviour of stones falling into water give rise to an event which is perceived via many sensory channels which each encode, in their different ways the complexity of the event. The perceptual system therefore has a number of representations of the event upon which to draw. In this paper, we suggest that it is possible to build a methodology for sound control upon commonalities between the behaviour of physical objects and that of sound objects which share many of their perceived physical properties.

1.1. Physics and the instrumental gesture

Many interactions with physical objects are physically or perceptually similar. For example, the experience of shaking a container of ice cubes shares many perceptual qualities with that of shaking a container of pebbles, or ball bearings, or boiled sweets. All are objects of a similar size and hardness, properties which give rise to similar auditory and haptic (inertial) percepts when they collide inside the container. Moreover, these similar physical properties also define the kinds of gestures that are possible. In the case of the example above, one can imagine reaching into the container and shuffling the objects or even removing some of them, or holding them in one's hand. Other classes of objects do not share these properties – one cannot for example remove a subset of the top of a table at will – the physics of the system constrains the desk to be not liftable in the same way as the ice cubes are. In this sense we may say that the physics of an interaction defines its gesture space.

Claude Cadoz, in developing a theory that could capture the notion of physical determinants of gesture space for musical instruments, defined the term *instrumental gesture* (Cadoz 1988; Cadoz and Wanderley 2000). For him, instrumental gestures are gestures which satisfy three requirements: they contain information conveyed to the audience (*semiotic*), they contain actions of the performer on the physical system (*ergotic*), and they encompass the perception of the physical environment or context by both the performer and the audience (*epistemic*). Thus, in the example above, the visual and auditory information of ice cubes colliding within a container when a performer shuffles them would be the semiotic component of the instrumental gesture. The actions of the performer are the ergotic part and the perceptions of the performer, for example the tactile sensations at the fingertips would be the epistemic component. In his definition, Cadoz strictly requires all three components to be present. Thus this definition would exclude the class of hands-free interactions as are used with gesture-based instruments such as the Theremin. For our purpose, Cadoz's notion of the instrumental gesture is interesting because, by his insistence on the simultaneous presence of the three components, he emphasises the complex interplay between the performer's actions and the sensory experiences for both the performer and the audience.

Organised Sound **11**(3): 285–296 © 2006 Cambridge University Press. Printed in the United Kingdom. doi: 10.1017/S135577180600152X

1.2. Related work

Our proposed design principle has its roots in the recognition of the importance of tangibility to instrument design (Rovan and Hayward 2000; O'Modhrain 2000). These ideas have already been expressed in different contexts earlier, for example within the design principles developed by Ryan (1991). As Ryan states, 'The trick is to put physical handles on phantom simulations' (Ryan 1991). For him, providing physicality, which later would be coined tangibility by Ishii and Ullmer (1997), is a way to make abstract sounds concrete.

Vertegaal, Ungvary and Kieslinger highlight the importance of tangibility for both the performer and audience by saying:

> Physical effort is an important musical parameter for both the artist and the audience:
>
> (1) Artists need to feel a piece as it is being created and performed. […]
>
> (2) The audience perceives the physical effort as the cause and manifestation of the musical tension of the piece […] (Vertegaal *et al.* 1996)

We see our main contribution to this earlier work as being one of anchoring design recommendations for tangible interfaces in a cognitive and sensorimotor framework, suggesting that rather specific variations in the relationship between the auditory and tangible aspects of an instrument interface are possible.

There are a growing number of converging proposals with respect to guidelines for new musical instrument design that are of interest here, though they do not strictly emphasise tangibility. The notion of intimacy between the performer and the instrument is emphasised by Fels (2004) and also Wessel and Wright (2001), going back to earlier considerations in this direction by Moore (1988). Intimacy, as Fels puts it, 'deals with the perceived match between the behaviour of a device and the operation of that device' (Fels 2004). Most comprehensive suggestions on an abstract level originate from the work of Wanderley (2001) and a comparably comprehensive view can also be found in the recent thesis by Jorda (2005). In Wanderley's work, the canonical mapping problem and strategies for post-design evaluation take a prominent role (see also Orio, Schnell and Wanderley 2001). Wanderley's main concern is the 'balanced analysis' of both gestural control and sound synthesis anchored in their relational mapping as well as in the development of evaluation methodologies as a pathway to design principles. Jorda promotes the notion of interface efficiency, parallelism, macro-control and transparency. Many of these positions come from the personal experience of ongoing design of musical instruments (Cook 2001; Jorda 2005). We do not seek to propose principles that work for all categories of interfaces and hence do not attempt to cover the scope of interfaces that are considered by Wanderley and Jorda. Rather we suggest a design principle in the context of tangibility that utilises cognitive and sensorimotor aspects explicitly.

A number of musical instruments have been designed in recent history which are very close to the principle we want to emphasise here. These can roughly be divided into three categories. One such category takes traditional instruments and augments them by adding sensing technologies that offer access to aspects of the instrumental gesture. The most comprehensive work in this area is that of Machover and his colleagues in their development of so-called Hyperinstruments (Machover 1992). The HyperBow (Young 2002), for example, offers access to a wide variety of parameters of bowed string performance including bow pressure, bow velocity and bow tilt. Similarly the HyperPuja controller designed by Diana Young allows for detection of the rubbing speed and pressure of the stick used in Tibetan singing bowl performances (Young and Essl 2003).

The second category of relevance here are those instruments which replicate rather than augment traditional instruments. Typically this replication process removes or alters some aspect of the traditional instrument. For example, the vBow (Nicols 2003) is a replication of the bowed string, but rather than keeping the sounding string, this is replaced with a force-feedback mechanism and a physically-derived audio model, so that the physical interface has been decoupled from the sounding of a physical string. The ePipe replicates tone-holed wind instruments like recorders or bag-pipe chanters (Cannon, Hughes and O'Modhrain 2003). Here the actual physical air flow has been removed and the action of the finger on the hole is detected using delicate fine-grain capacitive sensing. In a similar approach, Gillespie's haptic display mechanism for simulating piano key action removes the physical system that produces sound in the real piano, replacing it with a force feedback display and a synthesised audio output (Gillespie 1996). Perry Cook's PhISEM shaker percussion controllers also fall into this category (Cook 2001). While all of these instruments do replace the physical system that is the sound-producing mechanism, they all acknowledge the tight coupling between the sound and the feel of an instrument as a fundamental part of the instrumental gesture, and retain the physicality of the original interaction.

A variation of this category are instruments that are based on replication, but allow a variation to the performance characteristic due to the freedom introduced by the newly employed technology. An example of this type of instrument is the Tooka, designed by Felse and Vogt (2002), where a wind-instrument has been reconstructed but modified to allow for two-person play.

A third related category contains instruments that reappropriate instrumental gestures, that is to say instruments that use the physical properties of one mechanism to control another. Such instruments exploit similarities in the physics of families of sound-production mechanisms, and are most successful where the physics of the sound-producing mechanisms of the original and surrogate instruments are related through natural physical laws. At their simplest, MIDI keyboards that can be used to play piano or harpsichord patches are examples of such a reappropriation. Pianos and Harpsichords are both examples of keyboard instruments that use ballistic control – i.e. the player cannot affect a note once a key has been pressed. However, controllers exist which push this idea much further. The SqueezeVox family of instruments by Cook and Leider (2000), for example, makes use of a controller that reappropriates the gestures used in accordion playing to control a model of the human vocal tract. Here the underlying sound-producing mechanisms are far more complex, but are yet related by virtue of their physics. The dynamics of the breathing apparatus can be recognised in the dynamic behaviour of the air flow in the bellows of an accordion and in the movement of air across a reed or flap. It is this opportunity to reappropriate instrumental gestures, to take advantage of the similarities in opportunities afforded by related physical mechanisms, that has been the motivation for the body of work described below.

There are also a large number of abstract tangible interfaces for music control which have no parallels in existing instrument design. The Squeezables (2001) and the Sonic Banana (Singer 2000) are two examples of these kinds of devices. Their tangible properties are determined not by any physically motivated laws that link sound and touch, but by a desire to provide meaningful tangible responses to gestures such as squeezing and bending, which in turn can be mapped to the control of such musical parameters as timbral density and pitch bend, respectively. Thus, while not strictly mapped to the behaviour of an existing physical system, the result is a controller that explores the appropriateness of certain tangible interactions with sound itself.

1.3. Enaction in the context of musical instrument design

The close coupling of action and perception has also received much attention in philosophy and psychology, most notably in the theory of enaction. While the concept of enaction has a rich and diverse history and contesting definitions (Pasquinelli 2004), we will here define enaction as the necessary and close link between action and perception.

Actions necessitate concurrent and consequent perceptions, and perceptions guide and inform actions. Thus the concept of enaction is inevitably dependent upon embodied knowledge, the kind of knowledge that is derived from being and acting in the world, with all its physical properties and constraints. Varela, Thompson and Rosch, who put in place much of the foundations of the theory of enaction, described the crucial relationship between embodiment and enaction thus:

> By the term embodied we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context. By using the term action we mean to emphasise once again that sensory and motor processes, perception and action, are fundamentally inseparable in live cognition. ... the enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided. (Varela *et al.* 1991)

What concerns us here, then, is the consideration of enactive knowledge in the context of musical instrument design and how perceptually guided action defines the 'feel' and playability of a musical instrument. This can be true at a number of levels: Certain *a-priori* knowledge of physical principals (e.g. bangability, blowability, etc.) may inform initial expectations, while experience of interacting with the instrument will result in the acquisition of tacit knowledge about how to play it, i.e. the embodiment of specific knowledge – how fast to move a bow, the presence of sufficient rosin to control this motion, and so on.

In considering how concepts of enaction relate to performance, one question which must be considered is how the musician's working model of the dynamics of a system as complex as a musical instrument is built up. Moreover, how can enaction, in Bruner's broader definition of 'knowing by doing' (Bruner 1968), account for the musician's ability to predict the outcome of actions which they may never have had to produce before – a nuance in performance that is produced *ad libitum*. The theory of enaction suggests that prior knowledge of integrated sensorimotor experiences and their ongoing support by repeated interaction is important in forming an answer to this question.

The specific physical configuration of a musical instrument ultimately defines what senses will be involved in the experience of playing the instrument. For the performer, engrossed in the physical world, the question is, then: In what way does a specific integration of sensory perception and motor action correspond to a controllable, even enjoyable musical performance experience? This is a difficult question, but we propose retaining the familiarity of actions in the physical world as a first practical answer.

Designing new instruments ultimately involves a change in the relationship between performed actions and perceived responses on the part of both performer and listener. The challenge is to find ways to introduce such change while allowing for the persistence of certain enactive knowledge, i.e. in a manner that continues to allow knowledge of real physical systems that support the tight coupling of auditory and tangible percepts to inform the design of entirely new instruments.

We call our approach the hypothesis of weak sensorimotor integration. The hypothesis assumes that the real world supports some amount of flexibility in the coupling of action and sensory response. Whether one plays with rocks, shuffles leaves, or moves one's hand through water, there is in each case a tight coupling between actions performed and the tactile and acoustic responses to these actions. Our assumption of flexibility suggests that, if a set of actions can be grouped according to some shared physical behaviour (shuffleability, stirrability, etc.), then we may be able to freely substitute sonic responses to actions performed on one material with sonic responses to the same actions performed on another. The potential power of this approach is that the laws governing the physics of the group of actions are potentially supported for the user by some element of enactive knowledge, knowledge acquired by interacting with similar properties of objects in the real world.

2. BUILDING ENACTIVE INSTRUMENTS: DESIGN BY VARIATION

At the heart of our design of tangible controllers for sound synthesis is a recognition that there exist a class of sounds which are produced by our actions on objects in the world. Thus, dragging, dropping, scraping and crushing give rise to correlated touch and sound events (Rocchesso and Fontana 2003). As noted earlier, such events also bear many signatures of other physical characteristics of the materials and actions involved. However, it is possible to imagine a further class of events where the feel of an object and the sound it produces are less strongly correlated – for example, when playing with pebbles in one's hand, the haptic sensation one feels is that of the pebbles against the hand, while the sound of the interaction stems from the colliding of pebbles within the hand. This loose correlation between feel and sound is appropriate for this experience and in its looseness provides an opportunity to extend the range of plausible couplings between the haptic and auditory experience of the event.

The basic mechanism we employ to achieve this flexibility in the coupling between the sound and feel of the instrument is to take advantage of situations where such interactions have a natural acoustic component – friction between surfaces, collisions between objects, etc. In our designs, we capture this acoustic information and process it to derive relevant parameters for later control of alternative sound synthesis models. The fact that the physical properties of the interface are the common element that relates the class of sounds we wish to control ensures that the dynamics of the interaction are appropriately preserved. However, it should be pointed out that the choice of acoustic sensing to capture gesture data is not essential to the ideas presented. Any sensor technology that provides information about the dynamics of the interaction can be employed. For example, motion sensing, visual collision detection or other methods could be employed. We focus in our work on the use of acoustic signals for the simple reason that the signal is immediately related to the parameters we seek, sensor technology is cheap and easy to incorporate, and the signal substitution happens with exactly the dynamic characteristics we seek without being biased by the sensing technology. In the following sections, we describe in detail the implementation of a number of prototype instrument designs. Two are for coarse grain collision interactions, one is for fine-grain brittle interactions, and the final example is concerned with friction-based interactions such as rubbing.

It should be noted here that the notion of similarity between musical instruments with respect to their physical and musical properties is a significant component within the body of work on classification of musical instruments.[1] While traditionally musical instruments were classified top-down by some chosen *a-priori* categories, the work of Elschek and Stockmann has shifted attention to the detailed description of the properties of musical instruments by as many aspects as possible. They seek to find groupings by inspecting these properties across instruments and seek to understand which changes of properties make them essentially similar or essentially different. Hence rather than following a preset top-down classification, the classification is sought bottom up, emerging from the descriptive properties (Kartomi 1990: 198–203). Our proposal can be understood to fall into this category. We seek emerging properties that form commonalities within an instrumental space. However, contrary to Elschek's work, we emphasise aspects of perception and cognition in the presence of action, and the work is motivated by discovering essential categories of perception similarity among all involved senses.

2.1. Variations on the physical interface

2.1.1. PebbleBox

PebbleBox is a musical interface designed for granular sound synthesis. The granular synthesis is made tangible by providing the performer with physical grains that can be manipulated by hand. While the performer can choose the sound grains to be used, the main performative control is in the temporal structure and the dynamic content of the grains. Hence the parameters

[1] The history of the proposed categorisations are comprehensively reviewed by Kartomi (1990).

we seek to extract are those of temporal collisions of grains. Mainly we seek two parameters: the amplitude or force of a collision and a measure of the type of objects that collided. Spectral content depends on shape, size and collision position of an object, and hence we derive a measure of spectral content. These parameters are derived for each clearly identified temporal event and hence we additionally get a temporal texture, which itself is a strong perceptual cue (Warren and Verbrugge 1984).

The pebbleBox consists of a foam-padded container, which holds a number of non-brittle objects. We typically use polished rocks as available for gardening decoration. The number of rocks is chosen to be sufficient to fill one layer of the container. One version of the device is depicted in Figure 1. The interface supports a wide range of gestures and actions. The rocks can be shuffled or stroked or a single rock can be picked up and used to strike others. Alternatively, a number of rocks can be picked up and dropped into the box.

2.1.2. DaGlove

DaGlove is very similar in concept to PebbleBox. Environmental sounds are picked up through a microphone embedded in the glove. The sound is then processed to detect event triggers, which are in turn used to synthesise new sounds. The main difference is the locus of the sensor with respect to the user. In the case of DaGlove, the microphone is embedded in the palm area of an open-finger glove. Thus the user can still have access to the detailed tangible properties of the objects they are handling in the real environment, while allowing us to capture the sound of the interaction to drive the sound synthesis model. To accomplish this, we used a wireless microphone to allow the user free motion in the range of the base station. Containers with coarse-grain objects such as PebbleBox can also be performed with this device so allowing the locus of interaction to move seamlessly from inside the container to an interaction within the hand of the user. This greatly extends the range of gestures that can be supported. The complete design can be seen in Figure 2.

2.1.3. CrumbleBag

CrumbleBag is another musical interface for granular sound synthesis. But rather than dealing with collisions of non-brittle objects, this interface makes tangible the behaviour of brittle and deformable objects. The performer is given a bag filled with grains. Complex grabbing actions can be performed creating sound from braking and other sounding phenomena from the objects inside. Again, more vigorous grabbing will induce louder sound events. Also, spectral information will carry details about the sounding mechanism. Hence these parameters are picked up and made available.

A foam-rubber material was chosen as the outer bag, which was padded with a layer of felt. One microphone was placed inside the bag. The bag size is big enough to be manipulated easily by either one or two hands. A plastic bag is used to contain the actual filler material and can easily be replaced. We considered cereal, coral shell and Styrofoam fillings in our realisation. Figure 3 shows CrumbleBag and some filler material.

Typical performance gestures are grabbing and kneading. For example, footsteps on brittle material can easily be performed by skilled grabbing gestures. For additional detail on these designs, see O'Modhrain and Essl (2004).

Figure 1. The PebbleBox.

Figure 2. DaGlove.

Figure 3. The CrumbleBag with cereal, coral and Styrofoam fillings.

2.1.4. Scrubber

Sounds resulting from sliding actions are addressed by the Scrubber interface. Sliding often results in a particular class of sound based on friction. These friction-induced sounds are usually sustained through the time of the action and hence have a rather different character than the impulsive and short-term sounds that result from collisions or fracture. As a representative of a familiar device which corresponds to sliding hand actions we chose a white-board eraser, which also has a familiar form factor compared to brushes, sponges and other hand-held devices which are used in sliding motions. This device is then used on some surface, and the variation in the specific feel of the sliding action is defined by the surface. As with PebbleBox, tangible properties of the interaction are provided by the real-world context, in this case by the nature of the surface being rubbed.

For this interface, a whiteboard eraser was gutted and its interior was replaced by a custom-made silicon filler, containing a pipe-shaped cavity to hold the sensor and associated electronics. Two microphones were inserted into the cavity and a force-sensing resistor was fixed to the bottom side of the silicon. The silicon core was then wrapped in felt and inserted into the eraser casing. The assembly of the interface can be seen in Figure 4. Sliding actions form the typical performance gesture for this interface. These can be slow and regular as with the stroking of a cat or the sliding of doors, or fast and irregular, like the striking of a match. More detail on the design can be found in Essl and O'Modhrain (2005).

Figure 4. The Scrubber (note that this work is not sponsored, nor does it endorse this particular eraser brand).

2.2. Variations of the sound of the instrument

A crucial design choice in all these devices is the abstraction of the sound of a physical situation and the possibility of replacing it by another sound. In principle one could think of many technological means for detecting the physical behaviour which is relevant for sounding mechanisms. Our choice, to detect gestural nuance in performance by analysing an audio signal derived from the physical interaction with the interface, was based on the richness of gestural nuance that was present in the audio signal for these interactions and the fact that we could use well-understood analysis methods to extract salient parameters and performance gestures. The parameters we extract are then used to control alternative sound models.

In addition to the benefit of the close mapping in dynamical behaviour between the physical and the replaced sound, this has the added benefit of making use of simple and inexpensive sensor technology, with convenient interface to commodity computers through standard microphone or line-in inputs. The main design drawback is that the sound of the physical phenomenon of interest cannot always be guaranteed to be separated from other environmental sounds. In practical terms this wasn't a major problem as simple design choices help minimise the potential bleeding of other sound sources into the signal.

The use of the audio signal as a means to devise new interfaces for musical expression has a long history. Oliveros has used live processing of audio in her performances for many decades. The Expanded Instrument System provides control over delay times, delay feedback, pitch transformation, and other features through real-time audio processing (for a recent review of the technology, see Gramper and Oliveros 1998). Many technological advances were inspired by this desire to use the audio signal to derive control parameters to manipulate sound. For example, the classic paper by Puckette, Apel and Zicarelli (1998) was an important stepping stone by providing a widely accessible processing object for PD and Max/MSP.

Work by Ciufo (2003) and Jehan and Schoner (2001) are recent examples of this ongoing line of investigation. Ciufo emphasises audio analysis that focuses on timbre changes over time and incorporates them in hybrid performances of physical instruments (for example, prepared electric guitar) and an audio-stream analysing computer. Jehan and Schoner also focus on timbral extraction, but propose perceptually informed algorithms to do so.

Our work differs from this line of research in the sense that we use the features extracted from the audio stream purely as control parameters for completely unrelated sound, and that we emphasise temporal characteristics of interactions. In fact, the sensing of the interaction parameters could be replaced by non-acoustic sensing without impacting the principle. For us, using the audio signal is a technologically easy and convenient way to address the concern.

2.2.1. Sonic event detection

The process of deriving relevant control parameters starts with a model of the characteristic sounds created by the controlling objects. Once the characteristics are fixed, a method for extracting the parameters using signal processing techniques is developed.

PebbleBox and CrumbleBag are interfaces that are concerned with so-called granular sounds. These are sounds that are impulsive in nature, hence are characterised by a rapid onset and a fairly speedy exponential decay. Scrubber is concerned with friction sounds which are usually sustained in nature. These sounds are characterised by a sustained volume between the start of the action and its end. The onset must not necessarily be rapid, neither should the decay. These are two different requirements and hence it is necessary to define different strategies to extract parameters for control. However, in both cases our approach is to pay attention to features of the overall envelope for event detection, hence the signal will be averaged for a short window to generate the envelope. Onsets are exceptions to this; they are detected immediately from the signal to remove latency imposed by an averaging window.

The task of the parameter extraction for grainy sounds is to find the onset moment, the onset amplitude, and to enforce the separation of oscillatory features from onsets. The onset moment is found by a thresholding procedure: if a signal exceeds a pre-defined threshold (see Figure 4), a granular event is noted, and then the next local maximum is found. To reject further peaks which lie within the time at which separate events can be discerned by the ear in the time domain, further peaks will be disregarded for a specified duration d_r (cf Figure 4). The transition between temporal and spectral hearing happens between frequencies of 10–20 Hz or alternatively for a temporal range of 0.05–0.1 seconds. The duration is picked to reject frequency features and accept events that are perceived temporally. This is motivated by our desire to capture the perceptual cues in the temporal domain that group sounds, as discovered by Warren and Verbrugge (1984). A related thresholding scheme used on a spectrally composed signal has originally been proposed by Puckette and co-workers (1998). One further parameter is also derived from the signal. Zero crossings are calculated following the onset for a predefined number of samples to give an estimate of spectral information. Zero crossings are known to be a correlate of the spectral centroid, which is the spectral centre of gravity (Peeters and Rodet 2002; Panagiotakis and Tziritas 2004). Given the fidelity of the microphone used, we found that the dynamic range of the amplitude as extracted from the microphone signal was not a confining factor. While zero crossing is only a rough spectral measure, it provided an obvious correlate between the original dynamics and new sound. Performers only ever commented about perceptual latency which we had in the system in early implementations due to the limitations of the sound processing hardware. A low latency external soundcard helped address this problem.

The moment of onset for sustained segments of friction sounds is found the same way as before. A threshold defines when an event starts, but rather than finding one individual peak amplitude, we extract an ongoing amplitude envelope, which describes the dynamic evolution of a sustained sound. When the envelope falls back below the threshold, the event is assumed to have ended, and the duration between these two points gives the event time. Again zero crossings are counted within a moving window of fixed size to give a correlate of spectral information. Additionally, if two channels (i.e. two microphones) are used, the inter-channel delay between the two onsets can be used to estimate the moving speed and direction of motion of the scrubber.

We extract two types of features: temporal dynamics, and spectral content. The methods we employ for extracting temporal dynamics are well known and quite simple. These are the most salient properties we are interested in to determine the dynamics of the interaction and are inspired by the importance of the temporal features to perceptual grouping, as discovered by

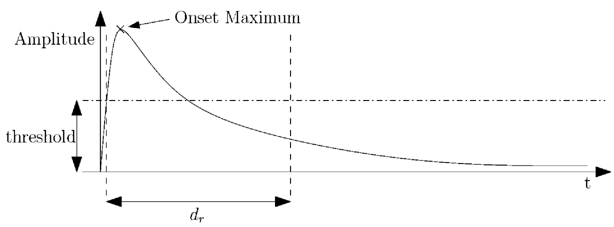

Figure 5. Threshold-based grainification scheme. The curve displays an amplitude envelope of an event. d_r is the retrigger delay, preventing detection of new onsets.

Warren and Verbrugge (1984). The spectral information is in some sense a convenient and accessible supplemental to this information. We found zero crossings to be sufficient for our purpose while at the same time being efficient. More detailed information could possibly be extracted about spectral information (see Puckette, Apel and Zicarelli 1998; Ciufo 2003). More sophisticated methods usually introduce higher computational demands and potentially additional delay, if larger Fourier-transform windows are required for finer spectral resolution.

2.2.2. Sound response synthesis

Depending on which of the control paradigms described above we are using, we may expect one of two sets of control parameters. For granular events we detect onset time, peak amplitude, and spectral content. For friction events we detect onset time, temporal amplitude envelope, offset time, and spectral content. How these parameters are used is in principle open and they could indeed be used in a way completely disconnected from their original meaning. Our purpose, however, is to vary sounding responses of physical interactions where the overall characteristics of the sounds are comparable, but the details have been varied. Hence in general we strive for synthesis algorithms which take the detected parameters and resynthesise them in a comparable fashion. Onset time should therefore lead to an onset of a sounding event, the peak and envelope amplitudes should control the peak and the envelope of the resulting sound, and the spectral content should create meaningful spectral variation in the resulting sound. In this way we ensure that all nuances in the gestures used when manipulating the interface objects are retained and passed through to control the synthesised audio stream.

We have therefore defined a certain core mapping strategy that we try to enforce on any particular synthesis method used which, as we have described, is dependent on finding physically related behaviours that have common gestural signatures. This greatly reduces the generic problem in the design of new interfaces for musical expression of relating sensory parameters to synthesis parameters, known as the mapping problem (Rovan, Wanderley, Dubnov and Depalle 1997; Hunt, Wanderley and Kirk 2000; Hunt, Wanderley and Paradis 2002)

However, the specific sound synthesis technique we use for each mapping remains open. We have tried two solutions, one based on wavetable playback using basic wavetable manipulation, the other of the class of parametric physically informed models as developed by Perry Cook and readily available as part of the synthesis toolkit STK (Cook 1997, 1999, 2002; for discussion on the connection of the second with the model, see O'Modhrain and Essl 2004; Essl and O'Modhrain 2005). Here we will focus on discussing the wavetable- based method.

The primary idea is to extract the temporal dynamics of an interaction. This information is usually related to specific aspects of the physical scenario, for example collision of objects for granular interactions or friction sound loudness envelope for friction interactions. This dynamical information can then be used to drive the synthesis of arbitrary new sounds while retaining the dynamical properties of the original interaction.

For the granular case, these would be any one of a number of impulsive or short-lived sounds which correspond to a specific short onset moment that induces energy. Examples we have tried include collections of coins, ice cubes, crushed cans, and droplets of water. Friction sounds, on the other hand, are characterised by subtle changes in the content of their spectra over time, corresponding to a continuous influx of energy. Examples of this class of sounds which we have implemented include sliding doors, bowed strings, and objects sliding across tables.

While these represent very different sounding behaviours, the basic mechanism of wavetable playback is the same: with an onset, a wavetable playback starts, the overall, or evolving volume being controlled by the control parameter. In the case of the granular sound, the end of the playback is naturally defined by the length of the sound grain and requires no further control. In the case of the sustained sound, the end is defined by the offset moment and playback has to continue until that time. Finally, in both cases the information about spectral content has to be used to modify the sound. As the zero crossing is a correlate of spectral centroid, which in turn is a descriptor of overall location of the spectral content, it can be seen as a descriptor of how high or low a sound is perceived overall. Hence we link this roughly to a higher- or lower-pitched playback of a grain. The variability of detected frequencies in various events along with amplitude changes is usually enough to make the playback appear rather natural for grainy sounds, even if only one recorded instance is used for the wavetable synthesis. In the case of sustained sounds, the repetition of changing temporal features in the sound tends to make repetition more perceptible. To address the potential of auditory fatigue, a number of events are used and randomly chosen for playback, hence replacing a grain with a dictionary of grains from which one instance is chosen for each event. In our implementation we used up to twelve different grains, though the typical number was around five.

The completed process in the case of granular sounds and friction sounds can be seen in Figures 7 and 8. Figure 7 displays the result of PebbleBox, whereas Figure 8 shows the signal of the Scrubber. The top signal shows the reading of a single channel microphone as placed inside the respective device. The bottom signal depicts the resulting synthesis driven by the parameters

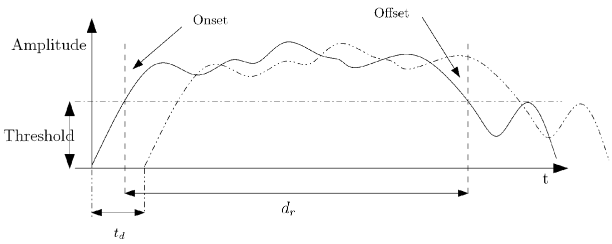

Figure 6. Threshold-based grainification scheme. The curve displays an amplitude envelope of an event. d_r is the event duration above a set threshold.

extracted from the top signal. One can clearly see the rough overall correspondence between the signals, but the rather significant difference in the details. In the granular case of Figure 6, a hammer grain was chosen that has a fairly long ringing time. Hence the amplitude envelope following an onset is much more prolonged. But notice that the temporal distribution of events exceeding the threshold depicted in the top part of the figure is the same; also the amplitude is comparable. The correspondence of the spectral behaviour cannot be seen in Figure 7.

The situation is similar for the Scrubber signal of Figure 8. Temporal and rough amplitude features are comparable in both signals. Here both onsets and offsets as defined by the signal are matched. Observe how the specific amplitude envelopes differ as the sound recording chosen for the resynthesised rendering has its own rather varied envelope, which is superimposed on the envelope extracted from the original.

Within these technical constraints, the suggested methodology keeps an open space for composing the sound of the instrument. For example, sounds that in any way resemble granular behaviour can be composed into PebbleBox and DaGlove. Hence the particular feel of the device is, despite the suggested confinements, open, and the composer or performer can choose desirable qualities of sound within this space. As any arbitrary real-time synthesis method can be used, whether it is wavetable playback or a parametric synthesis algorithm, performers can achieve a wide array of sonic responses. As the sensing is decoupled from the playback, the sensing doesn't impact the qualities of the played sound but controls only its temporal characteristics.

We see the main function of changing the content of PebbleBox or the filler material of CrumbleBag as ways to compose the tactile experience of the interface, though this choice also has an impact on the sonic performance of the interface by influencing what temporal patterns one will observe or what spectral information will be present. For example, breaking cornflakes have a different dynamics from Styrofoam fillings, and in this sense the choice of material can be used to modify the instrument towards a specific artistic intent.

2.3. Gesture and performance

In all the described interfaces, the important common theme is that a specific physical scenario is taken, and then subjected to variation. These variations can be tactile, or can be sonic. The gestures that are thinkable

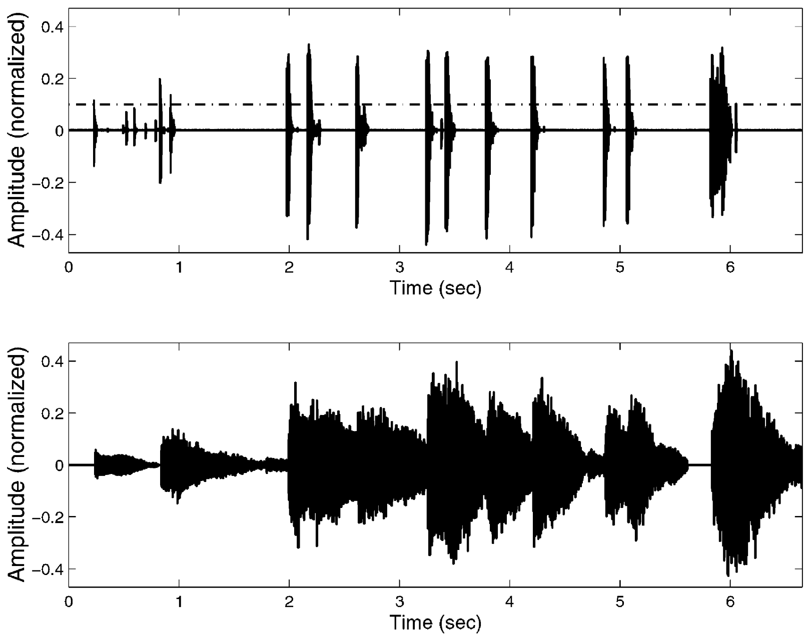

Figure 7. Recorded signal of the PebbleBox (top) and granulated response using a Hammer grain (bottom) of the complete granulation process.

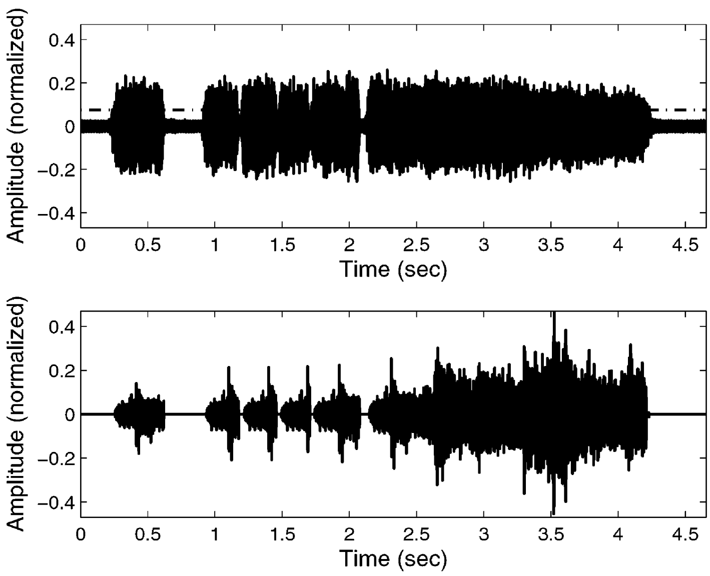

Figure 8. Recorded signal of the Scrubber (top) and frictified response using a sliding garage door sample (bottom) of the complete frictification process.

with a specific instrument, however, remain a common theme to that particular instrument and are mediated by the physical situation that was initially chosen as a starting point. To make this concrete, PebbleBox was designed to allow for performance of sounds that live in analogy with coarse-grain object ensembles that engage in dynamic collision. The tangible interface is indeed a collection of coarse-grain objects and the sensed parameters are the temporal dynamics between the collisions of these objects. Hence the kinds of gestures that are possible as well as meaningful are exactly the kinds of gestures that lead to collisions of these objects. Given these restrictions, the performer is free to then choose their performance style and gestures. PebbleBox, in particular, has been used by many performers in demonstrations at MediaLab Europe, and at Deutsche Telekom Laboratories. We observed a number of common gestures that performers chose. The most prominent gestures were shuffling gestures, which ranged from scooping gestures to stroking gestures. Pouring gestures were also popular. A number of rocks would be picked up and then dropped into the box. Some performers would pick up a stone and hit other stones to get individual control over a sonic response. Others preferred to pick two stones for the same purpose. While most performers played the complex ensemble behaviour that is suggested by the many pebbles, some would pick individual rocks and perform specific rhythmic patterns. Many performers did not understand the technical implementation through the audio channel, while some discovered this fact and exploited it by yelling into the box or by tapping stones against the outer casing of the box for additional performance styles.[2]

3. CONCLUSIONS

The musical instruments described here are illustrative of a design philosophy. At its core lies the maintenance of familiar sensorimotor experiences in interactive settings corresponding to sounding phenomena. The flexibility of the design is introduced by allowing for variation of sensations and particularities of motor actions of the setting. Hence we fix a rough scenario and vary the detail. The results of this design philosophy are instruments which work well within a specific class of interactions and their sounding responses. The two types of scenarios we have chosen to illustrate this were short-time impulsive sound as a result of collisions or fracture or related short-time sonic events and scenarios where the sound is sustained, usually associated with friction. These scenarios are associated with tactile expectations which are met by providing a physical interface that provides a tactile setting matching the class of interactions modelled.

But we are just at the beginning of understanding when a sensorimotor experience is natural and believable for a performer. We do not yet know what particular perturbations of the percepts available to the various sensory channels are permissible, or in what way the motor action has to correspond to a prior experience of a sonic response. This stands in stark contrast to the intricate and fine-tuned experience of expert performers of established musical instruments. These individual

[2]Since completing this manuscript, we have investigated the space of sonic variability in more formalised experimental settings. First results concerning aspects such as the likeability, believability and the sense of control with PebbleBox in relation to other haptic displays, e.g. a Phantom device, can be found in Essl, Magnusson, Eriksson and O'Modhrain (2005).

Figure 9. The design of PebbleBox for the Exhibition 'Touch Me: Design and Sensation', Victoria & Albert Museum (photo: Peter Kelleher, V&A Photo Studio).

experiences have not yet been understood well enough to lead to broad design principles.

Casual observation indicates that the principle of loose sensorimotor intergration upon which our designs are based works well for a range of variations in sound. These observations come from repeated individual demonstrations for visitors to Media Lab Europe in Dublin during the year of 2004. A larger version of PebbleBox has since been curetted for the Touch Me Exhibit at the Victoria & Albert Museum and was used by visitors during the summer of 2005.

Ultimately we hope that this exploratory work will help us to define some principals for the design of physically based controllers for classes of tangible musical instruments that embody Cadoz's notions of instrumental gesture, while at the same time being flexible and extensible for a wider range of physically inspired mappings between touch and sound.

REFERENCES

Bruner, J. 1968. *Processes of Cognitive Growth: Infancy*. Worcester, MA: Clark University Press.

Cadoz, C. 1988. Instrumental gesture and musical composition. *Proc. of the Int. Computer Music Conf. (ICMC-88)*, pp. 1–12. Cologne, Germany, San Francisco: ICMA Press.

Cadoz, C., and Wanderley, M. 2000. Gesture – music. In M. Battier and M. Wanderley (eds.) *Trends in Gestural Control of Music*, pp. 71–93. Paris: Editions IRCAM.

Cannon, C., Hughes, S., and O'Modhrain, S. 2003. EpipE: exploration of the uilleann pipes as a potential controller for computer-based music. *Proc. of the 2003 Int. Conf. on New Interfaces for Musical Expression – NIME03*, pp. 3–8. Montreal.

Ciufo, T. 2003. Design concepts and control strategies for interactive improvisational music systems. In *Proc. of MAXIS*. Leeds, UK: Sheffield Hallam University.

Cook, P. R. 1997. Physically Informed Sonic Modeling (PhISM): synthesis of percussive sounds. *Computer Music Journal* **21**(3): 38–49.

Cook, P. R. 1999. Toward physically-informed parametric synthesis of sound effects. *Proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-99)*, pp. 1–5. New Paltz.

Cook, P. R. 2001. Principles for designing computer music controllers. In *Proc. of the CHI'01 Workshop: New Interfaces for Musical Expression (NIME-01)*. Seattle, WA.

Cook, P. R. 2002. *Real Sound Synthesis for Interactive Applications,* Chapter 13, pp. 149–68. Natick, MA: A. K. Peters, Ltd.

Cook, P. R., and Leider, C. 2000. SqueezeVox: a new controller for vocal synthesis models. In *Proc. of the Int. Computer Music Conf.* Berlin, Germany.

Essl, G., Magnusson, C., Eriksson, J., and O'Modhrain, S. Towards evaluation of performance, control and preference in physical and virtual sensorimotor integration. *Proc. of the 2nd Int. Conf. on Enactive Interfaces*. Genova, Italy.

Essl, G., and O'Modhrain, S. 2005. Scrubber: an interface for friction-induced sounds. *Proc. of the Conf. for New Interfaces for Musical Expression*, pp. 70–5. Vancouver, Canada.

Fels, S. S., and Vogt, F. 2002. Tooka: explorations of two person instruments. *Proc. of the Int. Conf. on New Interfaces for Musical Expression (NIME02)*, pp. 116–21. Dublin, Ireland.

Gamper, D., and Oliveros, P. 1998. The Expanded Instrument System: recent developments. In *Proc. of the Int. Computer Music Conf.* Ann Arbor, MI.

Gillespie, R. B. 1996. *Haptic Displays of Systems with Changing KinematicConstraints: The Virtual Piano Action,* Ph.D. thesis. Stanford, CA: Stanford University.

Hunt, A., Wanderley, M. M., and Kirk, R. 2000. Towards a model for instrumental mapping in expert musical interaction. *Proc. of the Int. Computer Music Conf. (ICMC-00)*, pp. 209–12. Berlin, Germany.

Hunt, A., Wanderley, M. M., and Paradis, M. 2002. The importance of parameter mapping in electronic instrument design. *Proc. of the 2002 Conf. on New Instruments for Musical Expression (NIME-02)*, pp. 149–54. Dublin, Ireland.

Ishii, H., and Ullmer, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proc. of CHI'97*. Atlanta, GA: Association of Computing Machinery.

Jehan, T., and Schoner, B. 2001. An audio-driven, spectral analysis-based, perceptual synthesis engine. In *Proc. of the 110th AES Convention*. Amsterdam, The Netherlands: Audio Engineering Society.

Jorda, S. 2005. *Digital Lutherie – Crafting Musical Computers for New Musics' Performance and Improvisation,* Ph.D. thesis. Universitat Pompeu Fabra, Barcelona.

Kartomi, M. J. 1990. *On Concepts and Classifications of Musical Instruments*. Chicago: The University of Chicago Press.

Machover, T. 1992. Hyperinstruments – a composer's approach to the evolution of intelligent musical instruments. In L. Jacobson (ed.) *Cyberarts: Exploring Arts and Technology*, pp. 67–76. San Francisco: Miller Freeman Inc.

Moore, F. R. 1988. The dysfunctions of MIDI. *Computer Music Journal* **12**(1): 19–28.

Nicols, R. 2002. The vBow: development of a virtual violin bow haptic human-computer interface. *Proc. of the 2002 Conf. on New Interfaces for Musical Expression (NIME-02)*, pp. 29–32. Dublin, Ireland.

O'Modhrain, S. 2000. *Playing by Feel: Incorporating Haptic Feedback into Computer-Based Musical Instruments,* Ph.D. thesis. Palo Alto, California: Stanford University.

O'Modhrain, S., and Essl, G. 2004. PebbleBox and CrumbleBag: tactile interfaces for granular synthesis. *Proc. of the Conf. for New Interfaces for Musical Expression*, pp. 74–9. Hamamatsu, Japan.

Orio, N., Schnell, N., and Wanderley, M. M. 2001. Input devices for musical expression: borrowing tools from HCI. In *Proc. of the CHI'01 Workshop New Interfaces for Musical Expression (NIME-01)*. Seattle, WA.

Panagiotakis, C., and Tziritas, G. 2004. A speech/music discriminator based on RMS and zero-crossings. Retrieved online on 20 January 2004 at http://www.csd.uch.gr/~tziritas/papers/sound_seg_class_00.pdf

Pasquinelli, E. 2004. *Enaction. A State of the Art*. Unpublished manuscript.

Peeters, G., and Rodet, X. 2002. Automatically selecting signal descriptors for sound classification. *Proc. of the Int. Computer Music Conf.*, pp. 455–8. Göteborg, Sweden.

Puckette, M. S., Apel, T., and Zicarelli, D. D. 1998. Real-time audio analysis tools for Pd and MSP. In *Proc. of the Int. Computer Music Conf.* Ann Arbor, MI.

Rocchesso, D. and Fontana, F. (eds.) 2003. *The Sounding Object*. PHASAR Srl, Florence, Italy, 2003. Available online at http://www.soundobject.org/SObBook/SObBook_JUL03.pdf

Rovan, J. B., and Hayward, V. 2000. Typology of tactile sounds and their synthesis in gesture-driven computer music performance. In M. M. Wanderley and M. Battier (eds.) *Trends in Gestural Control of Music*, pp. 389–405. Paris, France: IRCAM.

Rovan, J. B., Wanderley, M. M., Dubnov, S., and Depalle, P. 1997. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Proc. of Kansei – The Technology of Emotion Workshop*. Genova, Italy. Available online at http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/kansei_final.pdf

Ryan, J. 1991. Some remarks on musical instrument design at STEIM. *Contemporary Music Review* **6**(1): 3–17.

Singer, E. 2003. Sonic Banana: a novel bend-sensor-based MIDI controller. In *Proc. of the Int. Conf. on New Interfaces for Musical Expression*. Montreal, Canada.

Varela, F., Thompson, E., and Rosch, E. 1991. *The Embodied Mind*. Cambridge, MA: MIT Press.

Vertegaal, R., Ungvary, T., and Kieslinger, M. 1996. Towards a musician's cockpit: transducers, feedback and musical function. *Proc. of the 1996 Int. Computer Music Conf.*, pp. 308–11. San Francisco, CA: International Computer Music Association.

Wanderley, M. M. [n.d.] Gestural control of music. *International Workshop on Human Supervision and Control in Engineering and Music*. Kassel, Germany.

Warren, W. H., and Verbrugge, R. R. 1984. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology* **10**(5): 704–12.

Weinberg, G., and Gan, S. -L. 2001. The Squeezables: toward an expressive and interdependent multi-player musical instrument. *Computer Music Journal* **25**(2): 37–45.

Wessel, D., and Wright, M. [n.d.] Problems and prospects for intimate musical control of computers. *Computer Music Journal* **26**(3): pp. 11–22. Cambridge, MA: MIT Press.

Young, D. 2002. The Hyperbow controller: real-time dynamics measurement of violin performance. *Proc. of the 2002 Conf. on New Interfaces for Musical Expression (NIME-02)*, pp. 65–70. Dublin, Ireland.

Young, D., and Essl, G. 2003. HyperPuja: a Tibetan singing bowl controller. *Proc. of the 2003 Int. Conf. on New Interfaces for Musical Expression*, pp. 9–14. Montreal, Canada.