

---

# A Manual Actions Expressive System (MAES)

---

RAJMIL FISCHMAN

Music, School of Humanities, Keele University, Keele, Staffordshire, ST5 5BG, UK  
E-mail: r.a.fischman@keele.ac.uk

**This article describes a *manual actions expressive system* (MAES) which aims to enable music creation and performance using natural hand actions (e.g. hitting virtual objects, or shaking them). Gestures are fully programmable and result from tracking and analysing hand motion and finger bend, potentially allowing performers to concentrate on natural actions from our daily use of the hands (e.g. the physical movement associated with hitting and shaking). Work carried out focused on the development of an approach for the creation of gestures based on intuitive metaphors, their implementation as software for composition and performance, and their realisation within a musical composition through the choice of suitable mappings, sonic materials and processes.**

## 1. DESCRIPTION

The *manual actions expressive system* (MAES) described below addresses the problem of enabling music creation and performance using natural hand actions (e.g. hitting virtual objects, or shaking them). The motivation for this project is to exploit the potential of hand gestures to generate, shape and manipulate sounds as if these were physical entities within a larger structured environment of independent sonic material (i.e. other entities that act independently from the manipulated sounds). In order to achieve this, the actions of the hands must be convincing causes of the sounds produced and visible effectors of their evolution in time, which is a well-known requisite for expression in digital devices and interfaces. Specifically for MAES, the intention is to preserve strong links between the causality of gestures and everyday experience of the world, yielding sound that is a believable result of the performer's natural actions and enabling intimate control of that sound. This can be achieved by mimicking and/or adapting the mechanics of physical phenomena; for instance, gripping grains to bind them together into a smooth continuous sound and loosening the grip to release them as a sparser jumbled texture, shaking particles to produce collisions and so on. Furthermore, causality can still be maintained in hyper-real situations: for example, a particle container that changes size from the proportions of a hand-held receptacle to the dimensions of a room.

An important aim in this project concerns the achievement of expressive content and sonic sophistication

with simple hand gestures (Wanderley 2001), so that MAES can be used by individuals who do not have formal musical training, because the performing gestures are already ingrained in their neuromuscular system. Yet, MAES aims to enable virtuosity in the compositional structuring and articulation of sonic material at a level comparable to outputs produced in the electroacoustic studio: while the gestures remain simple, the mappings associated with these can be sophisticated. This is achievable by subsuming complexity within the technology, thus reducing specialised dexterity in order to produce sophisticated sound. In fact, simple gestures in the real world often set in motion complex processes: for instance, when we throw an object, its trajectory is governed by the combined effects of gravity, air friction, the momentum transferred to the object by the hand and so on. This is also true for sonic processes, such as in the mechanisms at work when hitting a bell, rubbing a roughed surface with a rod and so forth. Furthermore, subsuming complexity also enables sonic manipulation within a larger sonic environment in which, similarly to real life, we act within our independent surroundings and our actions modify the latter, but do not control it totally. Ideally, structure design and implementation of the environment, its mechanics and its functions can be created beforehand, while still allowing individual expression and performance freedom in a manner analogous to the design of videogames.<sup>1</sup> Thus, the participant is able to realise an individual instance of a performance within the constraints and affordances resulting from such structural design.

To a significant extent, the considerations above differ from the concept of *instrument*, which requires learning conventions specific to a particular device that are often outside common bodily experience (e.g. it is necessary to learn how to use a bow to produce sound on a violin, or to develop an embouchure to

<sup>1</sup>The degree of performance variability depends on the intelligence and sophistication of the system. In its current state, MAES is reactive, which means that it follows a structured sequence of subsections when prompted by the performer's actions. The individual characteristics of the realisation of each subsection (e.g. duration, specific spectromorphological attributes) depend on the user's actions and are therefore unique to each performance.

play a wind instrument). Instead, MAES aims to maintain gestural affinity with manual actions rather than engaging with specially built mechanisms. This also minimises the necessity for timbral consistency in comparison to instrument-driven metaphors, enabling use of a wide variety of spectromorphologies, only limited by the current capabilities of the processing engine and the imagination of the user. Therefore, the device is treated as a transducer of existing bodily skills that becomes as transparent as possible and, as technology develops, will disappear altogether. Furthermore, it is envisaged that future developments of systems of this type will shape and manipulate audiovisual and haptic objects, bringing the model even closer to the mechanics of videogame play. Nevertheless, although such objects are not yet implemented in MAES, the physicality of the metaphors employed implies tacit tactility and vision: hopefully, this should become apparent in the mapping examples and musical work discussed below (sections 4.4 and 5).

The conceptual approach in MAES focuses on mapping strategies and spectromorphological content: rather than investing time and effort on the creation of new devices this research emphasises the adaptation of existing technology for creative compositional use in gesture design, sound design and the causal match of gesture with sound. Therefore, in addition to the development of the software tools required for adaptation, most of the work has gravitated around the design and implementation of a sufficiently versatile mapping strategy underpinned by a corresponding synthesis and processing audio engine, and a viable compositional approach which is embodied and demonstrated in the resulting musical work: the mapping of gestures is as important as the corresponding selection of spectromorphologies and sound processes. In other words, the main concern shifts from technological development to actual content, ultimately embodied in the musical output. The use of buttons or additional devices (pedals, keyboards, etc.) in order to perform a work is avoided in order to prevent disruption of the sound shaping/manipulation metaphor, aiding to the smoothness of a performance and, since the captured data is generic, simultaneously reducing dependency on a specific device. The result is an interactive environment that facilitates the composition of works in which the performer is responsible for the articulation of part of the sonic material within a larger sonic field supported by the technology. Moreover, we can already observe the incipient mechanics of a videogame, in which the actions of the user prompt a response (or lack of response) from the technology.

## 2. BACKGROUND AND CONTEXT

The quest for intuitive interfaces appropriate for music performance is inextricably linked with issues

concerning *gesture* and expression.<sup>2</sup> Recent research has led to the discovery of important insights and essential concepts in this area, which have facilitated practice-led developments.

### 2.1. Gesture and expression

Decoupling of *sound control* from *sound production* (Sapir 2000; Wanderley 2001) facilitated the implementation of a new breed of digital performance devices. However, it also highlighted the potential loss of causal logic and lack of expression when performers' gestures cannot be associated to sonic outputs (Cadoz, Luciani and Florens 1984; Cadoz 1988; Cadoz and Ramstein 1990; Mulder 1994; Roads 1996; Goto 1999, 2005).

Gesture has been defined as all multisensory physical behaviour, excluding vocal transmission, used by humans to inform or transform their immediate environment (Cadoz 1988). It fulfils a double role as 'symbolic function of sound', and 'object of composition' whose validity can only be proven by the necessities of the creative process; often requiring trial and error development through its realisation in musical compositions (Krefeld 1990).

### 2.2. Mapping and metaphor

Causal logic is dependent on *mapping* – in other words, the correspondence between gestures or control parameters and the sounds produced (Levitin, McAdams and Adams 2002). Correspondence can be *one-to-one*, when one control parameter is mapped to one sound parameter; *convergent*, when many control parameters are mapped to a single sound parameter; *divergent*, when one control parameter is mapped to many sound parameters (Rovan, Wanderley, Dubnov and Depalle 1997) or a combination of these.<sup>3</sup> Mappings may be *modal*, when internal modes choose appropriate algorithms and sound outputs for a gesture depending on the circumstances, or *non-modal*, when mechanisms and outputs are always the same for each gesture (Fels, Gadd and Mulder 2002). Furthermore, gestures are most effective when mappings implement higher levels of abstraction instead of raw synthesis variables, such as brightness instead of relative amplitudes of partials (Hunt, Paradis and Wanderley 2003). This is achieved by adding additional modal mapping layers, which can be time-varying (Momeni and Henry 2006).

Mappings should be intuitive (Choi, Bargar and Goudeseune 1995; Mulder 1996; Mulder, Fels and

<sup>2</sup>Fels, Gadd and Mulder (2002: 110) define expression as 'the act of communicating meaning or feeling. Both player and listener, therefore, are involved in an understanding of the mapping between the player's actions and the sounds produced.'

<sup>3</sup>Often called *many-to-many*.

Mase 1997; Wessel, Wright and Schott 2002; Momeni and Wessel 2003), exploiting intrinsic properties of our *cognitive map* that enable tight coupling of physical gestures with musical intentions (Levitin et al. 2002). Successful gestures can incorporate expressive actions from other domains. This is desirable because spontaneous associations of gestures with sounds are the results of lifelong experience (Jensenius, Godoy and Wanderley 2005). This leads to the concept of *metaphor* (Sapir 2000), whereby electronic interfaces emulate existing gestural paradigms which may originate in acoustic instruments – such as the *eviolin* (Goudeseune, Garnett and Johnson 2001) or the *SqueezeVox* (Cook and Leider 2000) – or in generic sources – such as *MetaMuse*'s falling rain metaphor (Gadd and Fels 2002).

Metaphor facilitates *transparency*, an attribute of mappings that indicates the psychophysiological distance between the intent to produce an output and its fulfilment through some action. Transparency enables designers, performers and audiences to link gestures with corresponding sounds by referring to *common knowledge* understood and accepted as part of a culture (Gadd and Fels 2002): spontaneous associations of gestures with sounds and cognitive mappings are crucial components of this common knowledge.

Therefore, MAES aimed to develop strong metaphors through hand gestures embedded in common knowledge belonging to the cognitive map of daily human activity: as long as these are linked to appropriate spectromorphologies they have the potential to produce convincing mappings for gesture design.<sup>4</sup> Performers do not have to consider parameters and mapping mechanisms, but rather conceive natural actions akin to human manual activity (e.g. throwing and shaking objects, etc.); being reinforced by the multimodal nature of these actions.

### 2.3. Learnability versus virtuosity

Gestural interfaces should balance between a potential for virtuosic expression and *learnability* (Hunt et al. 2002).<sup>5</sup> Technologies requiring little training for basic use but allowing skill development through practice strike this balance, offering gentle learning curves and ongoing challenges (Levitin et al. 2002).

MAES enhances learnability by enabling the design of gestures that are already ingrained in the human cognitive map that constitutes a manual

<sup>4</sup>Gesture–sound linking does not apply exclusively to direct sound recognition (*source bonding*; Smalley 1997), but also to second and higher orders of surrogacy (Smalley 1986); where source bonding is associated with energy profiles, spectral similarity, and psychological and cultural conditioning.

<sup>5</sup>'Learnability' is 'the amount of learning necessary to achieve certain tasks' (Vertegaal, Ungvary and Kieslinger 1996)

repertoire. For instance, a set of instructions for the performance of a musical passage (i.e. a score) may consist of the following sequence of manual actions:

- Extend hand to front, grab virtual particle container and shake circularly for 3 seconds.
- Pause for 3.5 seconds.
- Shake for 5 seconds.
- Near the end: slow down, lower hand and open it slowly.

These indications are performable by musicians and non-musicians alike. Of course, it is important to ensure that the sonic results correspond to these actions: for instance, shaking could be mapped to audio grains being articulated according to the velocity of the hand, and so on.

MAES also enables progression towards virtuosity. However, it is important to stress that, since one of the aims is to enable music performance using simple gestures, virtuosity in the achievement of timbral variety through mapping and sound processing is prioritised over instrumental manual dexterity: in other words, the emphasis is on the development of compositional sophistication. For this purpose the software allows the user to configure complex mappings in order to design gestures and change these configurations as the performance progresses, rather than implementing a 'hard-wired' setting; offering a large number of possible available combinations of mapping primitives and interchangeability of the spectromorphologies controlled by these mappings. This enables the expansion of the inventory of actions available (see section 4.3).

### 2.4. Effort

The perception of physical effort enhances expressivity (Vertegaal, Ungvary and Kieslinger 1996; Krefeld 1990; Mulder 1994). Regardless of whether actual effort is real or virtual, it enlarges motion by projecting it and expresses musical tension through the musician's body language (Vertegaal et al. 1996). Although MAES does not implement actual effort into the gestures, this is already implied in our *cognitive map* as a result of daily experience (e.g. the muscular activity involved in throwing an object).

## 3. EXISTING INTERFACES

It is impossible to list all existing interfaces in this article.<sup>6</sup> We will therefore concentrate on systems that are relevant to the development of MAES. Following the latter's aims and approach, meaningful contextualisation focuses on guiding metaphors, corresponding mapping

<sup>6</sup>See Jordá (2005), Miranda and Wanderley (2006) and Torre (2013).

strategies, spectromorphological content, timbral control and their realisation in composition.

Historically, Michel Waisvisz's and Laettia Sonami's work has been influential to this day. Waisvisz's *The Hands* control of MIDI signals favours triggering (e.g. recorded samples) rather than continuous control of sonic attributes (Krefeld 1990; Torre 2013). Although there is little documentation on their functioning, the structure of the actual devices and Waisvisz's performances indicate an *instrumental approach* rather than hand-action sound-shaping (e.g. Waisvisz 2003). While also controlling MIDI, Sonami employs the *Lady Glove* (which includes a foot sensor) differently (Bongers 2007; Sonami 2010, 2013; Torre 2013): together with the use of filtering processes, she achieves expression by controlling large numbers of short snippets by means of concurrent mechanisms (hands mutual distance, hand-foot distance, orientation, finger bend, etc.) which, in addition to rhythmic control, allow her to shape sounds timbrally.<sup>7</sup> This is strengthened by the theatricality of her gestures.

MAES was influenced by *Sound Sculpting's* 'human movement primitives' metaphors<sup>8</sup> (Mulder, 1996; Mulder, Fels and Mase 1999), and the use of gestures for multidimensional timbral control. Both reinforce the manual dexterity innate in humans; the premise that, within limits, audio feedback can replace force-feedback; and an implicit sensorial multimodality (Mulder et al. 1997). However, *Sound Sculpting's* timbral control focused on FM parameters, providing narrower scope for variety and differing from MAES's regard for spectromorphologies as surrogate to hand actions' physical effects.

Rovan (2010) uses a right-hand glove comprising force-sensitive-resistors, bend sensors and an accelerometer, together with an infrared sensor manipulated by the left hand. Music structure articulation shares common ground with MAES, whereby the user can control durations of subsections within a predetermined order (Rovan advances using left-hand actions whilst MAES generally subsumes preset advancement within sound-producing gestures). On the other hand, Rován's metaphors are different from MAES's manual actions, suggesting an instrumental approach.<sup>9</sup>

Essl and O'Modhrain's tangible interfaces implement hand action metaphors (e.g. friction). Similar to MAES, these address 'knowledge gained through experiences with such phenomena in the real world' (2006: 285). However, because of their reliance on tactile feedback, metaphor expansion depends on the implementation of additional hardware controllers. Also, sonic output relies on microphone capture and

a limited amount of processing. MAES avoids these constraints through the virtualisation of sensorimotor actions supported by ingrained cognitive maps, and mapping flexibility coupled with synthesis/processing variety (e.g. interchangeable outputs that maintain surrogacy links to gestures): this compensates to an extent for the lack of haptic feedback.

*SoundGrasp* (Mitchell and Heap 2011; Mitchell, Madgwick and Heap 2012) uses neural nets to recognise *postures* (hand shapes) reliably, implementing a set of modes which are optimised for sound capture and post-production effects, and are complemented by a synthesiser and drum modes. In comparison, MAES enables more flexible mode change through programmable gesture-driven presets that are adapted to the flow of the composition, but hand-shape recognition is more rudimentary.<sup>10</sup> This reflects a significant difference in approaches and main metaphors: SoundGrasp's sound-grabbing metaphor is inspired by the sonic extension of Imogen Heap's vocals, which could be conceptualised as an extended instrument paradigm that may also trigger computer behaviours. This differs from MAES's focus on the control of spectromorphological content within an independent environment. Finally, SoundGrasp's mapping is more liberal than MAES's regarding common knowledge causality of gestures: while some of the mappings correspond to established cognitive maps (e.g. sound-grabbing gesture to catch a vocal sample, angles controlling rotary motion), others are inconsistent with MAES's metaphor (e.g. angle controlling reverberation). *SoundCatcher's* metaphor (Vigliensoni and Wanderley 2009) works similarly to SoundGrasp, but without posture recognition. It implements looping and/or spectral freezing of the voice: hand positions control loop start/end points. A vibrating motor provides tactile feedback for the distance from the sensors and microphone. Comparisons between MAES and SoundGrasp are also applicable to SoundCatcher.

*Pointing-At* (Torre, Torres and Fernstrom 2008; Torre 2013) measures 3D orientation (*attitude*) with high accuracy. A bending sensor on the index finger behaves as a three-state switch (as opposed to MAES's continuous bending data in all five fingers). MAES and *Pointing-At* share a compositional approach that combines the computer's environmental role with sound controlled directly by the performer's gestures; subsuming complexity in the technology but being careful not to affect the transparency of the interface. The approach to software design is also similar, using MAX<sup>11</sup> in combination with external objects purposely designed to obtain and interpret data from the controller. This enables a flexible mapping strategy, as evidenced in

<sup>7</sup>For example, resembling a granular engine in *Why \_\_ dreams like a loose engine (autoportrait)* (Sonami 2010).

<sup>8</sup>For example, claying and carving.

<sup>9</sup>See *COLLIDE* (Rovan 2009).

<sup>10</sup>Measuring individual finger bend within specified deviation tolerances.

<sup>11</sup>Cycling74. <http://cycling74.com> (accessed on 28 February 2013).

different pieces.<sup>12</sup> However, such changes require the implementation of purpose-built patches, as opposed to MAES's inbuilt connectable mappings between any data and sound processing parameters within a single patch. This also accounts for a difference in the approach to the integration of gesture design within a composition: Torre's design normally consists of a sequence of sub-gestures which form a higher-level complex throughout macrolevel sections in the piece, while in MAES mappings change from preset to preset, using gestures in parallel or in rapid succession, normally resulting in a larger number of modes.<sup>13</sup> This has an obvious impact on the structuring of musical works.<sup>14</sup> Also, Pointing-At allows for a liberal choice of metaphors which, similarly to SoundGrasp, combine mappings corresponding to hand-action cognitive maps (e.g. sweeping sound snippets like dust in *Agorá*) with more arbitrary ones (e.g. hand roll controlling delay feedback in *Molitva*). Technical differences between Pointing-At and MAES also influence gesture design and composition. For instance, Pointing-At facilitates spatialisation within a spherical shell controlled by rotation movement thanks to accuracy and a full 360-degree range of attitude data. On the other hand, position tracking motion inside the sphere is less intuitive.<sup>15</sup> In contrast, better accuracy and ranges for position tracking than orientation favour mapping of the former in MAES, therefore allowing coverage of points inside the speaker circle, including changes in proximity to the listener. Also, the use of continuous bend values for each finger in MAES enables different mappings<sup>16</sup> from those generated by bending used as a switch in Pointing-At (e.g. crooking and straightening the finger to advance through loop lists in *Mani*). Finally, theatrical elements and conceptual plots that aid gesture transparency are common to compositional approaches developed for both Pointing-At<sup>17</sup> and MAES.

The *P5 Glove*<sup>18</sup> has been used for music on several occasions<sup>19</sup>, mainly controlling MIDI<sup>20</sup> and/

or sample loops.<sup>21</sup> An exception is Matthew Ostrowsky's approach, which shares similarities with MAES: he focuses on creating gestalts using a P5 driven by MAX (DuBois 2012), achieving tight control of continuous parameters and discrete gestures.<sup>22</sup> The metaphors also manipulate virtual objects in a multidimensional parameter space, employing physical modelling principles (Ostrowski 2010). On the other hand, Ostrowski focuses on abstract attributes rather than seeking causality within a physical environment. Malte Steiner controls CSound parameters and graphics: documentation of a performance excerpt (Steiner 2006) suggests control of sonic textural material in an instrumental manner, while the control of graphics responds to spatial position and orientation.

*Nuvolet* tracks hand gestures via *Kinect*<sup>23</sup> as 'an interface paradigm ... between the archetypes of a musical instrument and an interactive installation' (Comajuncosas, Barrachina, O'Connell and Guaus 2011: 254). It addresses sound shaping, realised as navigation through a concatenative synthesis (mosaicking) source database. Although it differs significantly from MAES by adopting a path metaphor and use of a single audio technique, it shares two important concerns: higher abstraction level via intuitive mappings (e.g. control of spectral centroid) and a game-like structuring of a performance through pre-composed paths that the user can follow and explore, avoiding known issues related to the sparseness of particular areas of the attribute space. Similarly to *Nuvolet*, *The Enlightened Hands* (Vigliensoni Martin 2010) map position to concatenative sound synthesis. Axes are mapped to spectral centroid and loudness. Visuals are also controlled manually. While the issue of sparseness is identified and prioritised for future research, there is no indication that this has already been addressed in the project. *Mano* (Oliver 2010) provides an inexpensive yet effective method of tracking hand shapes using a lamp on a dark surface; offering detailed continuous parameter control. Its approach follows theories of cognitive theory embodiment, promoting simple mappings that arise from interrelated complex inputs.

The *Thummer Interface* (Paine 2009) benefits from the versatility and abundant data provided by the Nintendo *Wii Remote*.<sup>24</sup> Although it uses an instrument metaphor, conceptual premises leading to its development shed light on wider issues related to the design of digital controllers, such as transparency and high-level mapping abstraction: Thummer uses four predominant physical measurements (pressure, speed,

<sup>12</sup>*Molitva* (Torre 2012a), *Mani* (Torre 2012b) and *Agorá* (Torre 2012c)

<sup>13</sup>For example, the composition described in section 5 comprises 67 modes, each corresponding to a preset.

<sup>14</sup>In broad terms, Pointing-At facilitates macrostructural compositional planning with more improvisational freedom while MAES favours more pointillist micro- and meso-structural compositional planning with occasional improvisation.

<sup>15</sup>A spherical shell metaphor is at the heart of the aesthetics and structuring principles of *Agorá*.

<sup>16</sup>See section 4.4.2.

<sup>17</sup>See the programme notes for *Mani*, a musical interpretation of a political debate (Torre 2013).

<sup>18</sup>Launched in 2002 by Essential Reality. Now distributed by Virtual Realities (<http://www.vrealities.com/P5.html>) and Amazon.com ([http://www.amazon.com/Essential-Reality-P5-Gaming-Glove/dp/B00007JNFE/ref=sr\\_1\\_1?ie=UTF8&qid=1355234772&sr=8-1&keywords=p5+glove](http://www.amazon.com/Essential-Reality-P5-Gaming-Glove/dp/B00007JNFE/ref=sr_1_1?ie=UTF8&qid=1355234772&sr=8-1&keywords=p5+glove)) (accessed on 13 December 2012).

<sup>19</sup>A list is available at [http://scratchpad.wikia.com/wiki/P5\\_Glove](http://scratchpad.wikia.com/wiki/P5_Glove): Musical (accessed on 25 February 2013).

<sup>20</sup>For example, Richard Boulanger's *Hearing Voices* (2005) and *In the Palms of Our Hands* (2005).

<sup>21</sup>See *GlovePIE* (Kenner 2010).

<sup>22</sup>Apparently by mapping position and orientation: no detailed documentation available.

<sup>23</sup>Microsoft. <http://www.xbox.com/en-GB/Kinect> (accessed on 17 December 2012).

<sup>24</sup><http://www.nintendo.com/wii> (accessed on 6 February 2013).

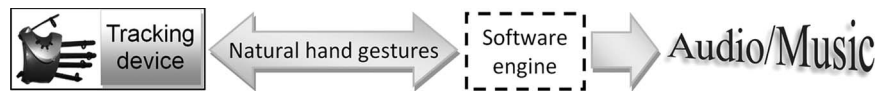


Figure 1. MAES block diagram.

angle and position) to control five spectromorphological parameters (dynamics, pitch, vibrato, articulation and attack/release) considered to be fundamental in the design of musical instruments. Therefore, it subsumes complex data mappings in the technology in order to achieve more tangible metaphors; an approach followed in MAES. Both Thummer and MAES implement configurable mappings and groupings between controller data and the sound production engine according to user needs. Finally, the ‘comprovisation’<sup>25</sup> approach described by Paine shares the concept of a (relatively liberal) structured individual trajectory, akin to videogame play in MAES’s approach.

Beyer and Meier’s system (2011) subsumes complexity in the technology in order to allow user focus on simple actions, similarly to MAES. Users with no musical training compose according to their preferences within a known set of note-based musical genres and styles. However, this project differs in its note-based metaphor and its emphasis on learnability by novice users over development of virtuosity, requiring hard-wired mappings – as opposed to MAES’s configurable mappings.

Other interfaces share less common ground with MAES but are listed here for completeness: *Power-glove* (Goto 2005) and *GloveTalkII* (Fels and Hinton 1998) use glove devices. *GRIP MAESTRO* (Berger 2010) is a sensor-augmented hand-exerciser measuring gripping force and 3D motion. *Digito* (Gillian and Paradiso 2012) implements a note-based modified keyboard paradigm. *Phalanger* (Kiefer, Collins and Fitzpatrick 2009) tracks hand motion optically, controlling MIDI.<sup>26</sup> *Couacs* (Berthaut, Katayose, Wakama, Totani and Sato 2011) relies on first-person shooter videogame techniques for musical interaction.

#### 4. IMPLEMENTATION

Figure 1 provides a convenient way of conceptualising MAES, consisting of a tracking device controlled by specialised software for the creation of musical gestures.

##### 4.1. Tracking device

Since the project focused on content, it was important to choose a device that would minimise the technical

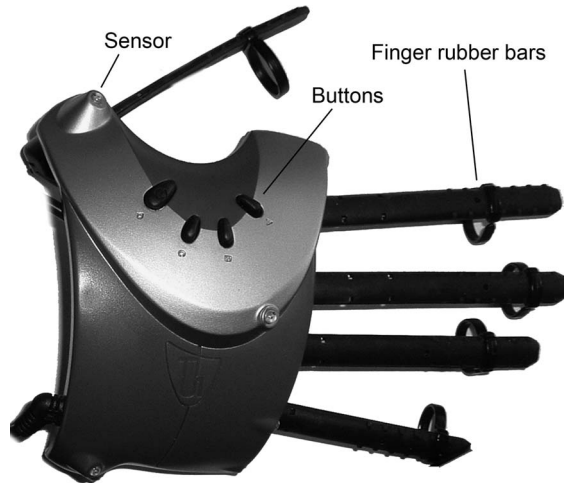


Figure 2. P5 Glove.

effort invested in its adaptation. The *P5 Glove* (Figure 2) captures the necessary data required to devise convincing gestures. It is affordable<sup>27</sup> and therefore within reach of the widest possible public, supporting reasonable expectations of affordability in future technology. While being an old device<sup>28</sup> it provides:

- tracking of three-dimensional translation and rotation, and finger bend,
- sufficient sensitivity and speed,<sup>29</sup>
- detection within a wide spatial range calibrated by each user.

However, the original manufacturer’s software library did not exploit its capabilities, working within a narrow spatial range and reacting sluggishly. Fortunately, McMullan (2003, 2008) and Bencina (2006) developed C libraries that access the glove’s raw data: these were used to implement alternative tracking functions within an external MAX object. Also, looseness of the plastic rings used to couple the rubber bands to the fingers resulted in slippages that affected the reliability and repeatability of finger bend measurements. This was significantly remedied with adjustable Velcro attachments placed between the base of the finger and the original rings.

<sup>25</sup>Paine describes ‘comprovisation’ in the following way: ‘aesthetic decisions pertaining to timbral space are determined in the composition process but the navigation of those potentials into a temporal form occurs through structured improvisation’ (2009: 152).

<sup>26</sup>See <http://www.youtube.com/watch?v=sVC4GR13F90> (accessed on 8 March 2013).

<sup>27</sup>Online price in 2013: \$60–\$80.

<sup>28</sup>Released in 2002. Full P5 at [http://scratchpad.wikia.com/wiki/P5\\_Glove:Specs](http://scratchpad.wikia.com/wiki/P5_Glove:Specs) (accessed on 26 February 2013).

<sup>29</sup>Although using USB 1.1, it provides reliable measurements at 15 millisecond intervals.

## 4.2. Software

In addition to a user interface implemented in MAX, the software consists of:

1. the processing package,
2. the external object *P5GloveRF.mxe* (Fischman 2013), and
3. the mapping mechanism.

### 4.2.1. Synthesis and processing

Synthesis and processing modules are interconnectable by means of a patchbay emulating matrix, and consist of the following:

1. Sound sources
  - 1.1. Three synthesisers (including microphone capture)
  - 1.2. Two audio file players (1–8 channels)
2. Spectral processes<sup>30</sup>
  - 2.1. Two spectral shifters (Fischman 1997: 134–5)
  - 2.2. Two spectral stretchers (Fischman 1997: 134–5)
  - 2.3. Two time stretchers (Fischman 1997: 134–5)
  - 2.4. Two spectral blur units (Charles 2008: 92–4)
  - 2.5. A bank of four time-varying formants<sup>31</sup>
3. Asynchronous granulation<sup>32</sup>
  - This includes the control of sample read position, wander and speed (time-stretch); grain density, duration, transposition and spatial scatter; and cloud envelope.
4. QList Automation
  - MAX QLists allow smooth variation of parameters in time according to breakpoint tables.
5. Spatialisation
  - A proprietary algorithm implements spatialisation in stereo, surround 5.1 and two octophonic formats, including optional Doppler shift. The matrix patchbay enables the connection of the outputs of any of the synthesisers, file players and processes to ten independent spatialisers. The granulator features six additional spatialisers and the file players can be routed directly to the audio outputs, which is useful in the case of multichannel files that are already distributed in space.<sup>33</sup>

### 4.2.3. MAX external object

*P5GloveRF.mxe* fulfils the following functions:

1. Communication between the glove and MAX

2. Conversion of raw data into position, rotation, velocity, acceleration and finger bend
3. Utilities, such as glove calibration, storing shapes, tracking display

The MAX patch includes optional low-pass data filters that smooth discontinuities and spikes. So far, these have only been used for orientation and, more seldom, to smooth velocities.

## 4.3. Mapping approach

The patch provides configurable mappings via a matrix instead of implementing a hard-wired approach.<sup>34</sup> It is possible to establish correspondences between 15 streams of tracked data and 18 continuous processing parameters, a physical model for throwing/sowing particles, two soundfile triggers, and a preset increment (Figure 3). This one-to-one mapping produces a set of primitives that can be used independently to establish basic correspondences and also combined simultaneously into divergent and many-to-many mappings to generate more complex metaphors, offering a large number of possibilities.<sup>35</sup>

Furthermore, correspondences can include a number of conditions: for example, Figure 4 displays conditions for mappings between position X and density 1. Also, mappings of continuous parameters can be *direct* – when increments (decrements) in the source parameter cause corresponding increments (decrements) in the target parameter – or *inverse* – when increments (decrements) cause corresponding decrements (increments). Continuous parameters can have more than one condition: for instance, in Figure 4 the mapping between position X and density 1 is also conditional on the value of position Y being greater than 0.5.

Finally, all these mappings can be stored in the patch's presets. Therefore, it is possible to change mapping modes and gestures used from preset to preset as often as required. Moreover, since it is possible to map tracked data and set conditions for preset increments, there is no need to use other means to increment presets and change mappings; this avoids disruption to the metaphor of direct shaping of sounds through manual actions, aids the transparency of the technology and enables an organic interaction between it and the performer.

The following section illustrates the flexibility of the mapping mechanisms in the generation of more

<sup>30</sup>Implementation using jitter matrices based on Charles (2008).

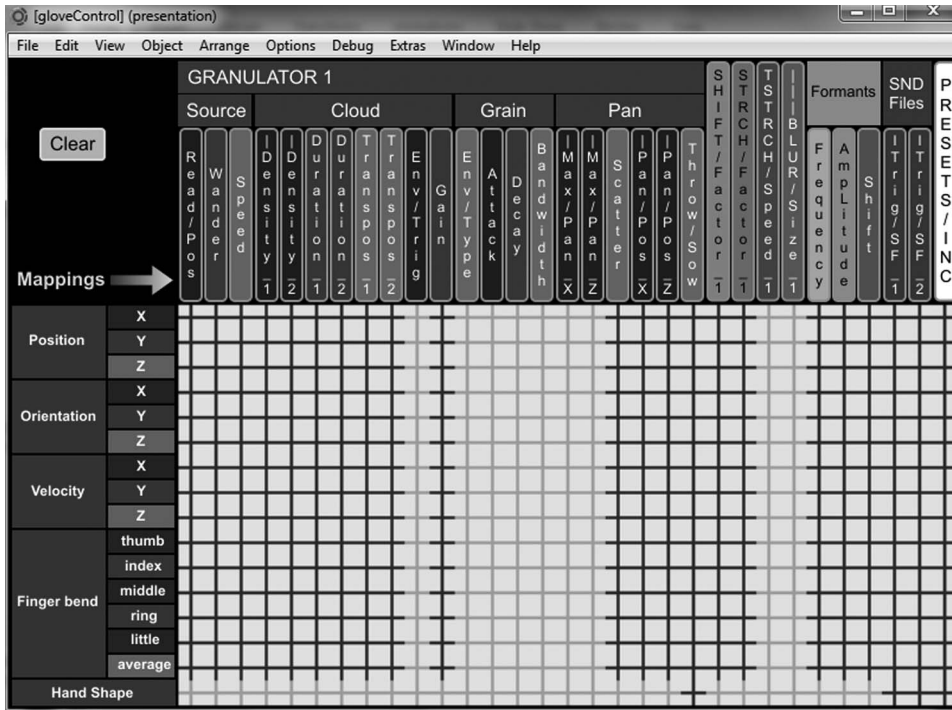
<sup>31</sup>See <http://en.wikipedia.org/wiki/Formant> (accessed on 24 December 2012).

<sup>32</sup>De Poli, Piccialli and Roads (1991: 137–86) explain asynchronous granular synthesis.

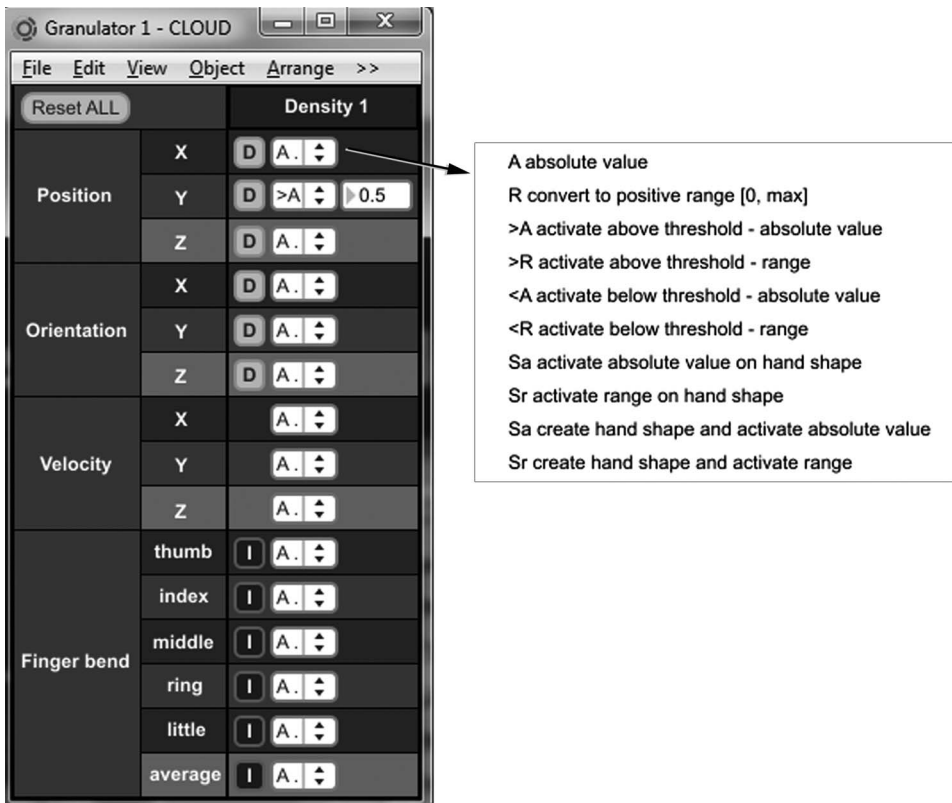
<sup>33</sup>Altogether, 18 independently spatialised streams.

<sup>34</sup>With the exception of a throw/sow gesture which could not be implemented by other means, since it required configurable time warp, velocity factors, gravity, sampling period and release duration.

<sup>35</sup>This allows  $16^{18}$  possible mapping combinations (including unmapped states) for continuous parameters. Additionally, hand-shape recognition can be mapped to throw/sow, soundfile triggers and preset increments, multiplying this number by  $17^4$ : in total  $(16^{18} \times 17^4) - 1$  combinations.



**Figure 3.** Mapping matrix. Tracked parameters appear in the left column. Sound parameters, audio file triggers and preset increments appear on the top row. Fuzzy columns indicate that mappings are not available (due to 32 bit limitations). Fuzzy entries in the bottom row disable hand shape mappings that do not make sense.



**Figure 4.** Mapping conditions for correspondences between position X and density 1.



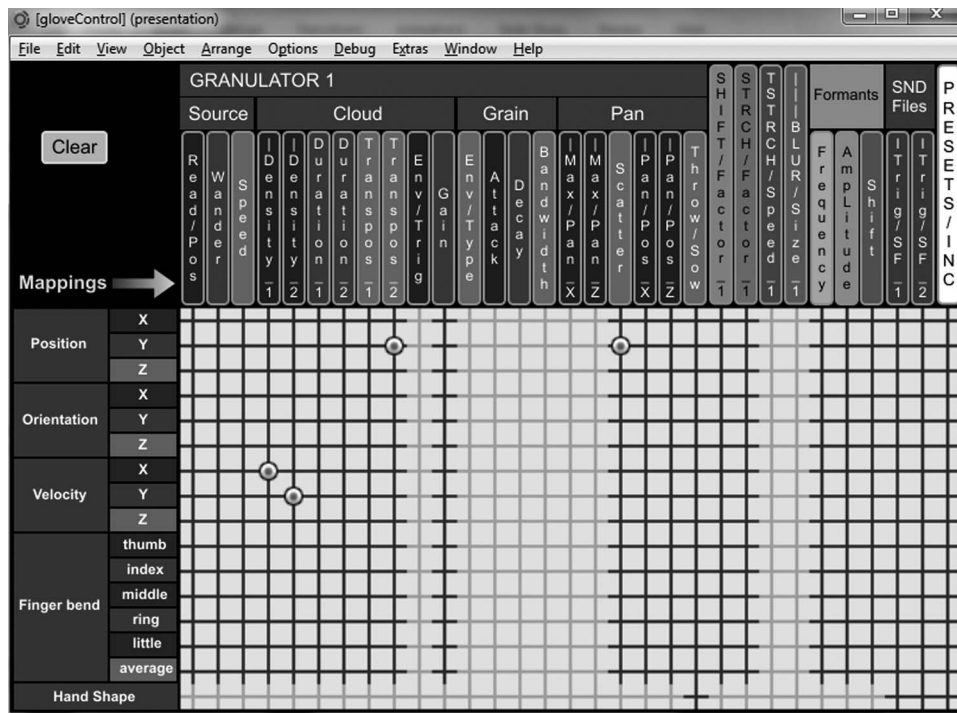


Figure 5. Shaking particles: mapping.

complex metaphors by combining primitives and setting up conditions. The generation of such metaphors is at the heart of this project's aim to achieve expression through simple gestures while subsuming technological complexity.

#### 4.4. Mapping examples

The examples will be described in terms of metaphors: the reader is encouraged to use these to form a mental image of the equivalent physical actions and resulting sound.

##### 4.4.1. Shaking particles (Movie example 1)

The performer holds and shakes a receptacle containing particles that produce sound when they collide with each other and with the edges of the container, as they are being shaken: the faster the shaking speed, the larger the number of collisions and corresponding sounds. When shaking stops there is silence. This metaphor is extended into hyper-reality by creating a correspondence between the particles' register (and perhaps their size) and the height at which the container is held: when the hand is lowered, frequency content is correspondingly low, and it goes up as the hand is raised.

Figure 5 shows the mappings used in this example:

1. X and Y velocities control the granulation density limits: when the hand moves rapidly between left and right (X), and between top and bottom (Y), the density is higher and we hear more collisions.

As the hand slows down we gradually hear fewer collisions. When the hand is still (velocity = 0) the grain densities are 0 and there is no sound.

2. Y position is mapped divergently onto the second transposition limit and the grain scatter: the higher the hand is held the higher the transposition (and the corresponding top register of the grains) and the wider their spatial scatter.

##### 4.4.2. Gripping/scattering a vocal passage (Movie example 2)

The performer traps grains dispersed in space, condensing these into a voice that becomes intelligible and hovers around the position of the fist that holds it together. As the grip loosens grains begin to escape and spread in space, and the speech slows down; disintegrating and becoming unintelligible.

Figures 6(a) and 6(b) show the mapping and conditions used in this example. Finger-bend average is mapped divergently onto the sample's reading speed, and the grains' duration limits, transposition limits and scatter. The following conditions apply:

1. Mappings onto the duration limits are direct (Figure 6(b)) so that maximum finger bending corresponds to the longest grains: in this case, maximum duration limits of 90 and 110 milliseconds were set elsewhere in the patch (not shown in the figure) in order to obtain good grain overlap and a smooth sound.

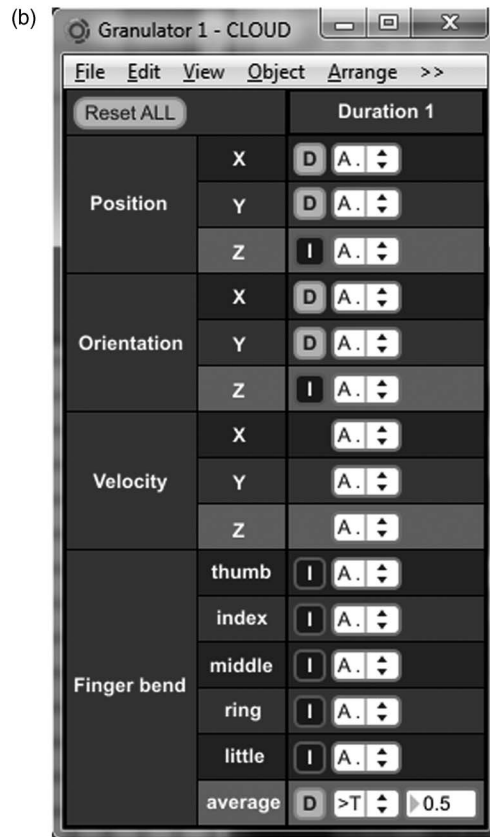
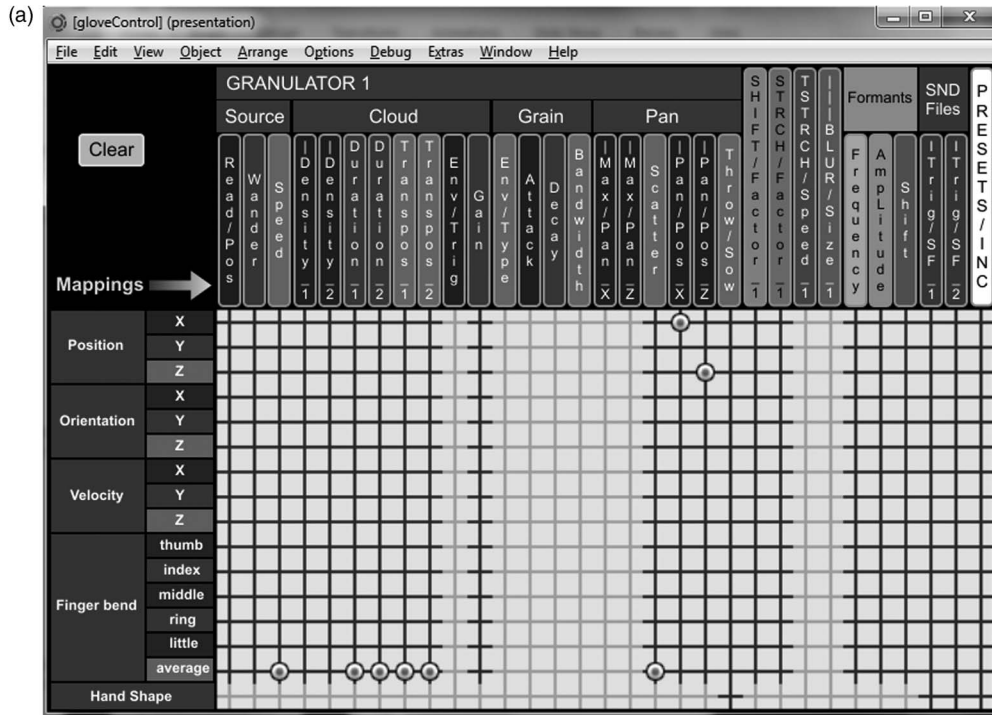
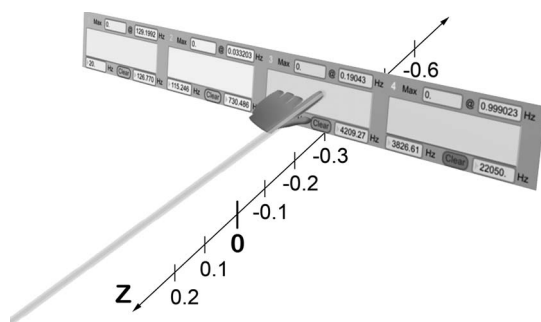


Figure 6. Gripping/scattering a vocal passage: (a) mapping; (b) conditions.

2. Finger bend only takes effect when it is greater than 0.5 (finger bend range is 0 to 1). This avoids grain durations that are too short: in this case, the

minimum durations will correspond to a bend of 0.5, yielding half of the maximum values above; in other words, 45 to 55 milliseconds.



**Figure 7.** Perforating a formant veil: the Z axis determines the position of the veil ( $z = -0.3$ ).

Nevertheless, we will hear shorter grains due to transposition by resampling.

3. Mappings onto transposition limits are inverted: full-range finger bend is mapped without conditions onto corresponding ranges of 0 to 25 and 0 to 43 semitones. As the fist tightens finger bend increases and transposition decreases towards 0; when the latter value is reached, the voice is reproduced in its natural register.

#### 4.4.3. *Perforating a veil (Movie example 3)*

The metaphor here is an imaginary soft barrier that prevents sound from breaking through, analogous to a fragile veil that blocks the passage of light. The barrier consists of a bank of formants that stop all frequencies. But in the same way the hand perforates a light blocking veil, the performer punches holes in the formant filters, allowing frequencies corresponding to the horizontal position of the holes to break through (Figure 7). Frequencies ascend from left to right, and the height determines the amplitude of each corresponding frequency. Finally, the veil has a depth position specified by the Z axis: the hand penetrates the veil when its position is deeper than that of the veil,<sup>36</sup> which in this case is  $-0.3$ . Therefore, if the Z position of the hand is greater than  $-0.3$ , there will be no perforation and no new frequencies will be heard. Conversely, if the hand's depth is less than  $-0.3$ , a frequency proportional to its X position and amplitude proportional to its Y position will be heard.

Figure 8(a) shows the mappings for this example, connecting the X position to frequency and the Y position to amplitude. Figure 8(b) shows the condition for the mapping of position X, which will only take effect when the Z position is less than  $-0.3$  (highlighted with a white rectangle). The same condition applies to mapping the Y position to amplitude.

<sup>36</sup>Z decreases with depth from 1 to  $-1$ .

## 5. MUSICAL WORK: *RURAQ MAKI*

Gesture is an essential component of music composition and is inextricably linked to the latter's intrinsic processes and necessities (Cadoz 1988). Thus the validity of a system for musical expression can only be corroborated by its effectiveness in the creation and performance of musical works. For this reason, the last stage of this project consisted of the creation and performance of a composition entitled *Rurraq Maki* (Fischman 2012).<sup>37</sup> MAES facilitated the construction of manual gestures by means of combinations of its mapping primitives, including the examples above. The implementation of tangible gestures allows the performer to generate and manipulate sounds, shaping the latter and interacting with MAES in a manner similar to that of videogame play, in which the technology is driven by the user according to rules that vary depending on the current state of the game – in other words, the musical work.

*Rurraq Maki* is pre-composed and performed according to a score (Figure 9). It is therefore repeatable and recognisable as the same composition from performance to performance in the same way scored music is, providing for ample interpretative variety and including short sections when the performer can improvise. Movie example 4 consists of a video recording and stereo mixdown of an eight-channel performance: the timings of the illustrative passages described below refer to this recording.

Following this project's approach, the performing gestures remain simple: conscious attempts were made to keep these strongly rooted in cognitive maps of daily human activity and to link them to appropriate spectromorphologies, in order to preserve a connection between their causality and everyday experience of the world.<sup>38</sup> This concerns the establishment of relatively close levels of surrogacy between common-knowledge hand actions and their aural effect. Furthermore, such links are also extrapolated onto the hyper-real. For example, in the beginning of the piece (0'07''–0'24'', rehearsal mark 3 in Figure 9) a particle container is shaken using the same mappings as movie example 1 (section 4.4.1.). While the particles are shaken, the performer can make them exceed a humanly sized container to the proportions of the speaker system, increasing their spatial scatter by adding vertical movement to the shaking (i.e. shaking diagonally): the higher the hand the more scattered the particles. As a result, the perspective of the audience is shifted from a situation in which the particle container is viewed from a distance to being inside the container.

<sup>37</sup>In the Quechua language spoken in the central Andes of South America *Rurraq Maki* means 'handmade'.

<sup>38</sup>Including efforts to subsume preset advancement in the gestures producing sonic outputs.

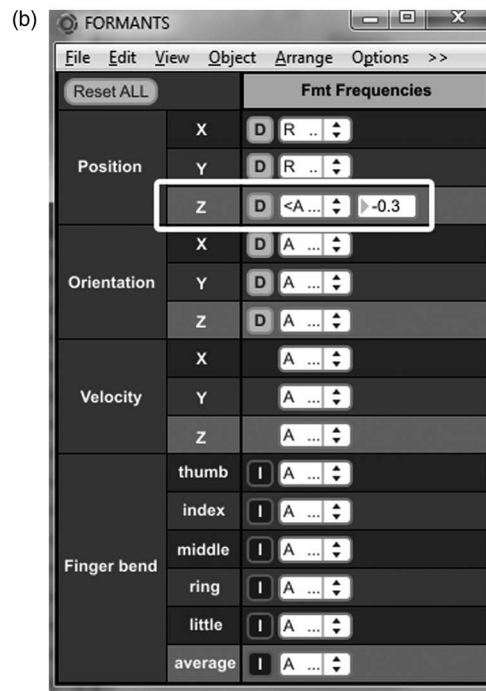
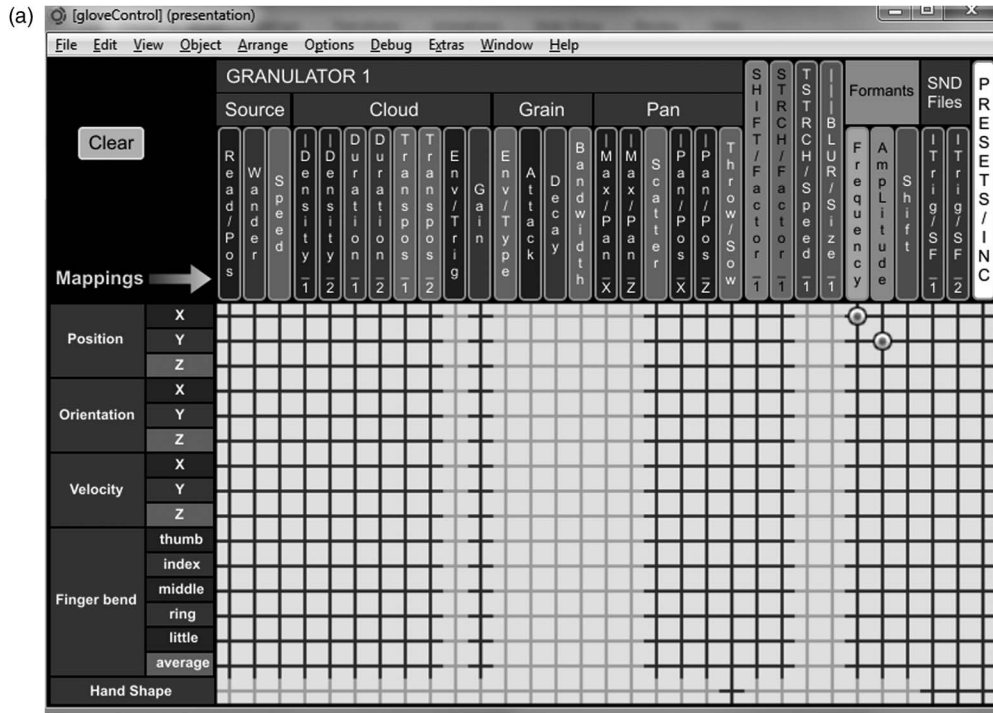


Figure 8. Perforating a veil: (a) mapping; (b) conditions.

During the process of learning the work, it was discovered that the approach described above aided memorisation: the metaphors became a mnemonic device similar to those used by memorisation experts, providing a sequence of actions similar to a plot for the realisation of the work (Figure 9).

### 5.1. Performer control versus sonic environment

The paradigm of sound manipulation within a larger structured environment of independent sonic material was realised through three main mechanisms:

1. direct control of continuous processing parameters through performer gestures,

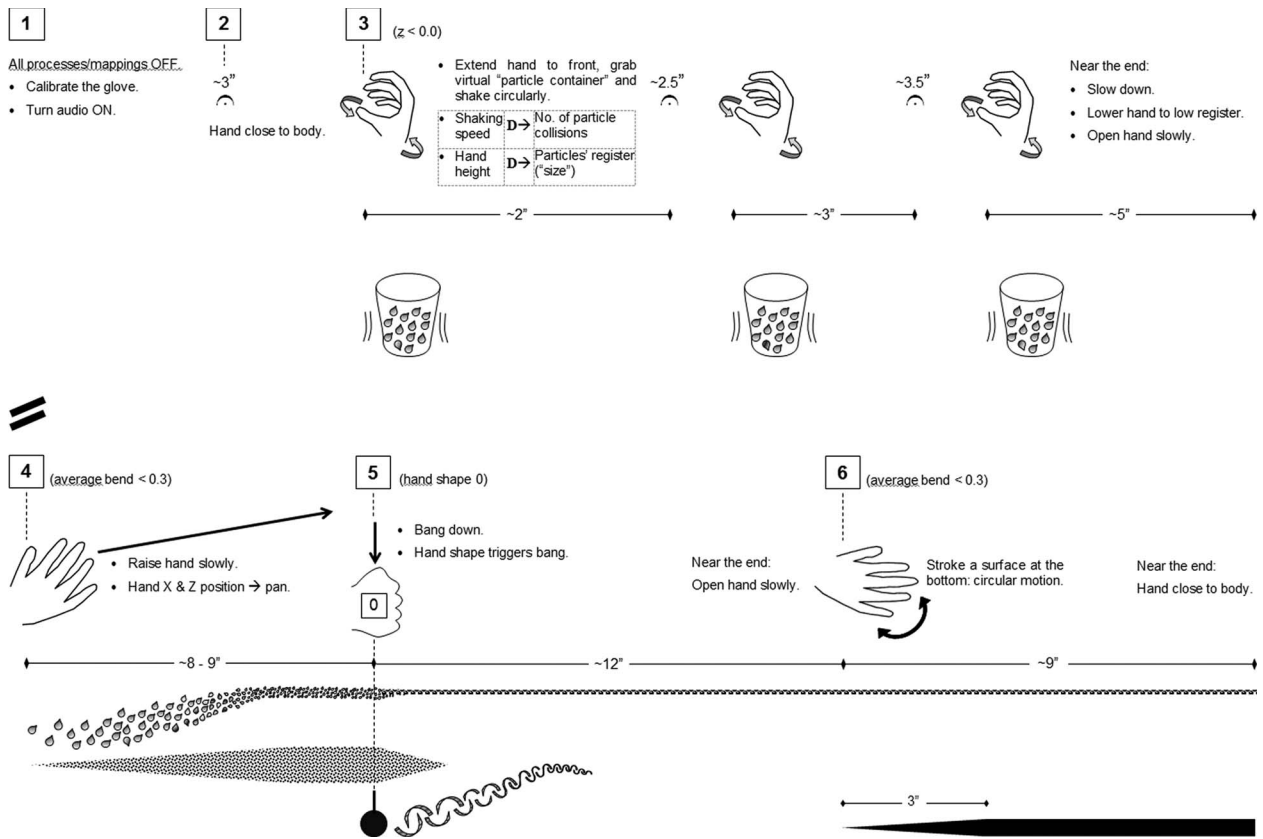


Figure 9. *Ruraq Maki* score, page 1.

2. gestures triggering pre-composed materials, and
3. fully automated time-varying processes assuming an environmental role.

Also, *ancillary* gestures were scored in order to aid metaphor identification by audiences. Unlike *functional* gestures, these do not affect the mechanics of sound production. However, they enhance performance expression, fulfilling similar functions to gestures in other genres; for instance, rock guitarists' exaggerated circular arm movements on downbeats, pianists' use of the whole upper body to emphasise cadences, undulating movements when playing cantabile and so on.

**5.1. Direct control of processing parameters**

This is the most intimate level of control, directly implementing the metaphor of sound shaping and manipulation as a physical entity. Examples are found throughout the piece, including the case of shaking particles illustrated above. At 1'05''-1'23'' and 8'04''-8'10'' the fingers are used as if they were a metallic flap sweeping through a curved row of bars, producing a corresponding sound in which the loudness increases as the index is straightened, the speed of the sweep is controlled by the hand's left-right velocity and panning follows the position of the hand in the horizontal plane. At 2'51''-3'14'' the

performer unfolds particles tentatively until they weld seamlessly into a single surface by stretching/bending the fingers (the surface becomes more welded as fingers stretch). This is accompanied by two ancillary gestures consisting of raising/lowering the hand and smooth motion in the horizontal plane leading eventually to a low palm-down position, as if stroking the welded surface. To realise this metaphor, average finger bend is mapped directly to granular transposition and scatter, and panning follows the position of the hand in the horizontal plane.

A structured improvisation section at 4'43''-5'50'' implements a gripping/scattering mechanism similar to that depicted in movie example 2 (section 4.4.2.), but applied to a processed quasi-vocal sample. Panning follows the position of the hand in the horizontal plane, which is clearly noticeable when there is minimum grain scatter (hand fully closed). The structure of the improvisation includes a plot through which a number of ancillary gestures are enacted: for instance, when attempting to grab the particles at the beginning a pitched buzz begins to form, at which point the performer releases the grip quickly as if reacting to a small electric shock; while panning the sound when it is firmly gripped, he follows the trajectory with the gaze as if displaying it to an audience, and so on.

Piercing of a formant veil similarly to movie example 3 occurs at 3'47''–3'56''; however, in this case, the veil is limited to the left half of the horizontal space by applying the additional condition  $x < -0.1$  to the mappings of formant amplitude and frequency. At 4'20''–4'37'', wiggling the fingers controls the articulation of a rubbery granular texture through inverse mapping of finger-bend average to grain density: changing finger bend by wiggling changes the density and, because this is an inverse mapping, fully bending the fingers produces zero density, resulting in silence.

At 9'53''–10'15'' the performer gradually grabs a canopied granular texture, releasing fewer and fewer grains as the hand closes its grip until a single grain is released in isolation, concluding the piece. Here, average finger bend is mapped inversely to density, similarly to the case of the rubbery texture controlled by wiggling. However, because of the spectromorphology used (highly transposed, shortened duration grains) the resulting metaphor is different.

## 5.2. Triggering pre-composed materials

Triggered materials consist of pre-composed audio samples and/or processes generated via QLists. They bridge between directly controllable and fully automated processes, depending on the specific gestures and associated spectromorphologies chosen during the mapping process. For example, hitting gestures at 6'22''–7'19'' are easily bound into cause/effect relationships with percussive instrumental sounds, becoming direct manipulation of sonic material by the performer.<sup>39</sup> Conversely, the accompanying percussive rhythm beginning at 6'56'' is fully environmental, since it is not associated with gestures that trigger a shaker (*quijada*) and, at 7'09'', a cowbell (*campana*). However, there are no clear boundaries between fully automated and triggered situations. For instance, at 6'41'' the performer hits an imaginary object above him, resulting in a strong attack continued by a texture that slowly settles into a steady environmental rhythm.<sup>40</sup> At 0'32'' we encounter an intermediate situation between full controllability and environment: the simultaneous closure of the fist (functional gesture) and a downwards hit (ancillary gesture) trigger one of the main cadences used throughout the piece, consisting of a low-frequency bang with long resonance mixed with a metallic texture. The bang has a causal link to the performer's

<sup>39</sup>This is similar to hitting a real object: after the hit, the performer has no control over the sound's evolution, which might result in a short spectromorphology (e.g. woodblock) or a longer resonance (e.g. bell). Yet, we establish a clear causality between the performer's gesture and the sonic output.

<sup>40</sup>This process, generated by a granulator QList, also has real life analogies, such as the effects of hitting a hornet's nest.

gesture but the texture does not, fulfilling an environmental role.

There is one type of mapping which is neither a trigger nor a continuous control paradigm: throw/sow implements momentum transfer from the user to audio grains, affecting their direction and velocity in the speaker space and simulating a throw. This happens at 7'37'', 7'44'' and 7'51'': in the original surround performance grains are launched towards the audience, traversing the space towards the back until they fade in the distance.

The examples above illustrate compositional rather than technological design, emphasising gesture's role as object of composition. For instance, compare the previous example with the triggering of a high-pitched chime by means of a flicker at 3'19''. If this flicker had triggered a low bang, or if a closed fist hit had triggered a chime, this would have not only affected musical meaning but also called for a different interpretation of the causality of the triggering gestures.<sup>41</sup>

## 5.3. Fully automated processes

These sustain the interactive role of the technology reacting to the performer's actions: when a gesture advances a preset, the technology can initiate QLists and/or play pre-composed soundfiles. The accompanying percussive rhythm at 6'56'' is an example of the latter, while an emerging smooth texture leading to a local climatic build-up at 0'24''–0'32'' is an example of a QList controlling both spectral stretchers. At 9'41'', a QList controls the granulator to produce a texture that changes from longer mid–low-frequency grains to the canopied granular texture which is subsequently grabbed by the performer at 9'53''.

## 6. DISCUSSION

The realisation of *Ruraq Maki* suggests that MAES provides a reasonably robust approach, a necessary and sufficient range of tools and techniques, and varied mapping possibilities in order to create expressive gestures.

As expected from any system, MAES has shortcomings; specifically:

1. Tracking of orientation angles is inaccurate. Therefore, orientation mappings are only useful when they do not require precision; for instance, when we are more interested in a rough fluctuation of a parameter.
2. Slippage of the finger rubber bands: although using Velcro is reasonable, this could be improved

<sup>41</sup>Of course, such interpretation might be intended during composition: for instance, in order to suggest a surreal or humorous situation.

significantly; for instance, using gloves made of stretching fabric that fit snugly on the hand. Nevertheless, it is worth remembering that the *P5* was used in order demonstrate that it is possible to implement an expressive system employing existing technology, with the expectation that future technological development will yield more accurate controllers, wireless communication and, eventually, more accurate tracking without bodily attached devices.

These shortcomings do not prevent MAES's usage by other composers and performers. The software is available under a public licence both as executable and as a source for further adaptation, modification and development. Although MAES is optimised for use with a digital controller, the synthesis and processing implementation can fully function without it. Furthermore, the patch allows straightforward replacement of the external object *P5GloveRF* controlling the glove by other MAX objects, such as objects that capture the data of other controllers. It is hoped that this will offer a wide range of possibilities which enable music-making within a wide spectrum of aesthetic positions, hopefully contributing to the blurring of popular/art boundaries and challenging this traditional schism.

## 7. FUTURE DEVELOPMENTS

MAES is the first stage of a longer-term research strategy for the realisation of *Structured interactive immersive Musical experiences (SiiMe)*, in which users advance at their own pace, choosing their own trajectory through a musical work but having to act within its rules and constraints towards a final goal: the realisation of the work. A detailed description of the strategy is given in Fischman (2011: 58–60), but its main premises follow below.

Firstly, in order to enable non-musically trained individuals to perform with MAES, it will be useful to implement a *videogame score* that provides performance instructions through rules and physics of videogames. For instance, it is not difficult to imagine a graphics interactive environment that prompts the actions required to perform the beginning of *Ruraq Maki* (Figure 9), requiring the user to reach a virtual container and shake it in a similar manner to movie example 1 (section 4.4.1.). This would also provide visual feedback which, although less crucial in the case of continuous control (Gillian and Paradiso 2012), would be extremely valuable in the case of discrete parameters and events, including thresholds and triggers.

Further stages of development include:

- increasing the number and sophistication of the audio processes and control parameters available,

- expanding to other time-based media (e.g. visuals, haptics),
- enabling the system to behave according to the particular circumstances of a performance (e.g. using rule-based behaviour or neural networks),
- implementing a generative system that instigates musical actions without being prompted (e.g. using genetic algorithms),
- implementing a multi-performer system incorporating improved controllers and devices,<sup>42</sup> and
- evolving the mechanics of musical performance by blurring the boundaries between audiences and performer, including music-making between participants in virtual spaces, remote locations,<sup>43</sup> and so on.

## Acknowledgements

This project was made possible thanks to a Research Fellowship from the Arts and Humanities Research Council (AHRC), UK; grant No. AH/J001562/1. I am also grateful to the reviewers for their insightful comments, without which the discussion above would lack adequate depth and detail, and the contextualisation of this project would have suffered.

## REFERENCES

- Bencina, R. 2006. *P5 Glove Developments*. <http://www.sil-mulus.org/p5glove> (accessed on 24 December 2012).
- Berger, M. 2010. The GRIP MAESTRO: Idiomatic Mappings of Emotive Gestures for Control of Live Electroacoustic Music. *Proceedings of the Conference on New Interfaces for Musical Expression*. Sydney. [http://www.nime.org/proceedings/2010/nime2010\\_419.pdf](http://www.nime.org/proceedings/2010/nime2010_419.pdf) (accessed on 14 January 2013).
- Berthaut, F., Katayose, H., Wakama, H., Totani, N. and Sato, Y. 2011. First Person Shooters as Collaborative Multiprocess Instruments. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Oslo, 44–7. [http://www.nime.org/proceedings/2011/nime2011\\_044.pdf](http://www.nime.org/proceedings/2011/nime2011_044.pdf) (accessed on 8 March 13).
- Beyer, G. and Meier, M. 2011. Music Interfaces for Novice Users: Composing Music on a Public Display with Hand Gestures. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Oslo, 507–10. <http://www.nime2011.org/proceedings/papers/M23-Beyer.pdf> (accessed on 20 January 2013).
- Bongers, B. 2007. Electronic Musical Instruments: Experiences of a New Luthier. *Leonardo Music Journal* 17: 9–16.

<sup>42</sup>For instance, *Kinect*, *Cyclops* (Singer 2010) and *ENACTIV* (Spasov 2009–10, 2011), XSENS (<http://www.xsens.com>; Skogstad, de Quay and Jensenius 2011; de Quay and Skogstad 2010a, 2010b, 2011), bio-signals (e.g. Ortiz, Coghlan, Jaimovitch and Knapp 2011).

<sup>43</sup>Pioneered by *LiveJam* (Waters 2000), improvements in latency make it possible to envisage realistic music-making between performers in remote locations; for instance LOLA (Low Latency Audiovisual System, <https://community.ja.net/groups/arts-and-humanities/article/lola-low-latency-audio-visual-streaming-system>; accessed on 24 December 2012).

- Cadoz, C. 1988. Instrumental Gesture and Musical Composition. *Proceedings of the 1988 International Computer Music Conference*. Cologne/San Francisco: ICMA, 1–12.
- Cadoz, C., Luciani, A. and Florens, J. 1984. Responsive Input Devices and Sound Synthesis by Simulation of Instrumental Mechanisms: The Cordis System. *Computer Music Journal* 8(3): 60–73.
- Cadoz, C. and Ramstein, C. 1990. Capture, Representation and ‘Composition’ of the Instrumental Gesture. *Proceedings of the 1990 International Computer Music Conference*. Glasgow/San Francisco: ICMA, 53–56.
- Charles, J.F. 2008. A Tutorial on Spectral Sound Processing Using Max/MSP and Jitter. *Computer Music Journal* 32(3): 87–102.
- Choi, I., Bargar, R. and Goudeseune, C. 1995. A Manifold Interface for a High Dimensional Control Space. *Proceedings of the 1995 International Computer Music Conference*. Banff/San Francisco: ICMA, 385–92.
- Comajuncosas, J.M., Barrachina, A., O’Connell, J. and Guaus, E. 2011. Nuvolet: 3D Gesture-Driven Collaborative Audio Mosaicing. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Oslo, 252–4. <http://www.nime2011.org/proceedings/papers/G11-Comajuncosas.pdf> (accessed on 23 January 2013).
- Cook, P. and Leider, C. 2000. Making the Computer Sing: The SqueezeVox. *Proceedings of the XIII Colloquium on Musical Informatics*. L’Aquila.
- De Poli, G., Piccialli, A. and Roads, C. 1991. *Representation of Musical Signals*. Cambridge, MA: The MIT Press.
- de Quay, Y. and Skogstad, S. 2010a. *Dance Jockey*. <http://www.youtube.com/watch?v=m1OffxIArrA> (accessed on 19 December 2012).
- de Quay, Y. and Skogstad, S. 2010b. *Dance Jockey v0.3*. <http://www.youtube.com/watch?v=HaD9MJzW59s> (accessed on 19 December 2012).
- de Quay, Y. and Skogstad, S. 2011. *Dance Jockey @ NIME*. <https://www.youtube.com/watch?v=PgE6YxhHgVnM> (accessed on 19 December 2012).
- DuBois, R.L. 2012. *Interviews with Matthew Ostrowski: The Glove*. <http://cycling74.com/2012/10/16/interviews-with-matthew-ostrowski> (accessed on 25 February 2013).
- Essl, G. and O’Modhrain, S. 2006. An Enactive Approach to the Design of New Tangible Musical Instruments. *Organised Sound* 11(3): 285–96.
- Fels, S. and Hinton, G. 1998. Glove-TalkII: A Neural-Network Interface which Maps Gestures to Parallel Formant Speech Synthesiser Controls. *IEEE Transactions on Neural Networks* 9(1): 205–12. <http://www.cs.toronto.edu/~hinton/absps/glovetalkii.pdf> (accessed on 20 December 2012).
- Fels, S., Gadd, A. and Mulder, A. 2002. Mapping Transparency Through Metaphor: Towards More Expressive Musical Instruments. *Organised Sound* 7(2): 109–26.
- Fischman, R. 1997. The Phase Vocoder: Theory and Practice. *Organised Sound* 2(2): 127–45.
- Fischman, R. 2011. Back to the Parlour. *Sonic Ideas – Ideas Sónicas* 3(2): 53–66.
- Fischman, R. 2012. *Ruraq Maki* (2012). <https://vimeo.com/55093629> (accessed on 17 December 2012).
- Fischman, R. 2013 (forthcoming). *MAES*, MAX patch and *P5GloveRF.mxe* external object.
- Gadd, A. and Fels, S. 2002. MetaMuse: Metaphors for Expressive Instruments. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2002*. Dublin. <http://hct.ece.ubc.ca/publications/pdf/gadd-fels-NIME02.pdf> (accessed on 20 December 2012).
- Gillian, N. and Paradiso, J. 2012. Digo: A Fine-Grain Gesturally Controlled Virtual Musical Instrument. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2012*. Ann Arbor. [http://www.eecs.umich.edu/nime2012/Proceedings/papers/248\\_Final\\_Manuscript.pdf](http://www.eecs.umich.edu/nime2012/Proceedings/papers/248_Final_Manuscript.pdf) (accessed on 17 January 2013).
- Goto, S. 1999. The Aesthetics and Technological Aspects of Virtual Musical Instruments: The Case of the Super-Polm MIDI Violin. *Leonardo Music Journal* 9: 115–20.
- Goto, S. 2005. Virtual Musical Instruments: Technological Aspects and Interactive Performance Issues. *HZ Journal* 6. <http://hz-journal.org/n6/goto.html> (accessed on 24 December 2012).
- Goudeseune, C., Garnett, G. and Johnson, T. 2001. Resonant Processing of Instrumental Sound Controlled by Spatial Position. *New Instruments and Musical Expression workshop, SIGCHI ’01*. Seattle. <http://zx81.isl.uiuc.edu/camilleg/nime01.pdf> (accessed on 14 December 2012).
- Hunt, A.D., Paradis, M. and Wanderley, M. 2003. The Importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research* 32(4): 429–40. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.3098&rep=rep1&type=pdf> (accessed on 14 December 2012).
- Jensenius, A.R., Godoy, R. and Wanderley, M.M. 2005. Developing Tools for Studying Musical Gestures within the Max/MSP/Jitter Environment. *Proceedings of the 2005 International Computer Music Conference*. Barcelona/San Francisco: ICMA, 282–5.
- Jordá, S. 2005. *Digital Luthery: Crafting Musical Computers for New Musics, Performance, and Improvisation*. PhD dissertation, Universidad Pompeu Fabra (Barcelona).
- Kenner, C. 2010. *GlovePIE*. <http://www.glovepie.org> (accessed on 25 February 2013).
- Kiefer, C., Collins, N. and Fitzpatrick, G. 2009. Phalanger: Controlling Music Software With Hand Movement Using A Computer Vision and Machine Learning Approach. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2009*. Pittsburgh: 246–9. <http://sro.sussex.ac.uk/2202/1/ChrisKieferNIME2009Paper.pdf> (accessed on 8 March 2013).
- Krefeld, V. 1990. The Hand in the Web: An Interview with Michel Waisvisz. *Computer Music Journal* 14(2): 28–33.
- Levitin, D.J., McAdams, S. and Adams, R.L. 2002. Control Parameters for Musical Instruments: A Foundation for New Mappings of Gesture to Sound. *Organised Sound* 7(2): 171–89.
- McMullan, J. 2003. *libp5glove*. P5 Glove library for low level access of raw data. [http://noisybox.net/computers/p5glove/libp5glove\\_svn\\_20050206.tar.gz](http://noisybox.net/computers/p5glove/libp5glove_svn_20050206.tar.gz) (accessed on 12 December 2012).
- McMullan, J. 2008. *P5glove*. <http://noisybox.net/computers/p5glove/> (accessed on 24 December 2012).



- Miranda, E.R. and Wanderley, M.M. 2006. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. Middleton, WI: A-R Editions.
- Mitchell, T. and Heap, I. 2011. SoundGrasp: A Gestural Interface for the Performance of Live Music. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Oslo. <http://www.nime2011.org/proceedings/papers/M13-Mitchell.pdf> (accessed on January 2013).
- Mitchell, T., Madgwick, S. and Heap, I. 2012. Musical Interaction with Hand Posture and Orientation: A Toolbox of Gestural Control Mechanisms. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2012*. Ann Arbor. [http://www.eecs.umich.edu/nime2012/Proceedings/papers/272\\_Final\\_Manuscript.pdf](http://www.eecs.umich.edu/nime2012/Proceedings/papers/272_Final_Manuscript.pdf) (accessed on 11 January 2013).
- Momeni, A. and Henry, C. 2006. Dynamic Independent Mapping Layers for Concurrent Control of Audio and Video Synthesis. *Computer Music Journal* **30**(1): 49–66.
- Momeni, A. and Wessel, D. 2003. Characterizing and Controlling Musical Material Intuitively with Geometric Models. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2003*. Montreal. [http://www.nime.org/proceedings/2003/nime2003\\_054.pdf](http://www.nime.org/proceedings/2003/nime2003_054.pdf) (accessed on 14 December 2012).
- Mulder, A. 1994. Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer. *Proceedings of the First Brazilian Symposium on Computer Music: Caxambu, Minas Gerais: 243–50*. Online: <http://www.xspasm.com/x/sfu/vmi/BSCM1.pdf> (accessed on 24 December 2012).
- Mulder, A.G.E. 1996. Getting a GRIP on Alternate Controllers: Addressing the Variability of Gestural Expression in Musical Instrument Design. *Leonardo Music Journal* **6**: 33–40.
- Mulder, A., Fels, S. and Mase, K. 1997. Empty-Handed Gesture Analysis in Max/FTS. *Proceedings of Kansei: The Technology of Emotion, AIMI International Workshop*. Genova. <http://hct.ece.ubc.ca/publications/pdf/mulder-fels-mase-1997.pdf> (accessed on 14 December 2012).
- Mulder, A., Fels, S. and Mase, K. 1999. Design of Virtual 3d Instruments for Musical Interaction. *Graphics Interface '99*: 76–83.
- Oliver, J. 2010. *The MANO Controller: A Video Based Hand Tracking System*. <http://www.jaimeoliver.pe/pdf/MANO-joliverl.pdf> (accessed on 20 December 2012).
- Ortiz, M., Coghlan, N., Jaimovitch, J. and Knapp, B. 2011. Biosignal-Driven Art: Beyond Biofeedback. *Sonic Ideas – Ideas Sónicas* **3**(2): 43–52.
- Ostrowski, M. 2010. *Congerries 2*. <https://soundcloud.com/#harvestworks/matthew-ostrowski-congerries-2> (accessed on 25 February 2013).
- Paine, G. 2009. Towards Unified Design Guidelines for New Interfaces for Musical Expression. *Organised Sound* **14**(2): 142–55.
- Roads, C. 1996. *The Computer Music Tutorial*. Cambridge, MA: The MIT Press.
- Rovan, J.B. 2009. *COLLIDE: For Custom Glove Controller, Interactive Computer Music & Interactive Video (2002)*. <http://vimeo.com/2729606> (accessed on 27 February 2013).
- Rovan, J.B. 2010. *Butch Rován*. [http://www.soundidea.org/rovan/research\\_glove.html](http://www.soundidea.org/rovan/research_glove.html) (accessed on 27 February 2013).
- Rovan, J.B., Wanderley, M.M., Dubnov, S. and Depalle, P. 1997. Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. *Proceedings of Kansei: The Technology of Emotion, AIMI International Workshop*. Genova. [http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/kansei\\_final.pdf](http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/kansei_final.pdf) (accessed on 20 December 2012).
- Sapir, S. 2000. Interactive Digital Audio Environments: Gesture as a Musical Parameter. *Proceedings COST-G6 Conference on Digital Audio Effects (DAFx'00)*: 25–30. <http://profs.sci.univr.it/~dafx/Final-Papers/pdf/Sapir.pdf> (accessed on 13 December 2012).
- Singer, E. 2010. *Cyclops*. <http://cycling74.com/products/cyclops/> (accessed on 18 December 2012).
- Skogstad, S., de Quay, Y. and Jensenius, AR. 2011. OSC Implementation and Evaluation of the Xsens MVN suit. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Oslo, 300–3. <http://www.nime2011.org/proceedings/papers/G23-Skogstad.pdf> (accessed on 24 December 2012). See also: [http://www.youtube.com/watch?v=KI6BF\\_RRQs](http://www.youtube.com/watch?v=KI6BF_RRQs) (accessed on 19 December 2012).
- Smalley, D. 1986. Spectro-Morphology and Structuring Processes. In S. Emmerson (ed.) *The Language of Electroacoustic Music*. Basingstoke: Macmillan, 61–93.
- Smalley, D. 1997. Spectromorphology: Explaining Sound-Shapes. *Organised Sound* **2**(2): 107–20.
- Sonami, L. 2010. *Why \_\_ dreams like a loose engine (autoportrait) (2000–8)*. <http://vimeo.com/11316136> (accessed on 27 February 2013).
- Sonami, L. 2013. Home Page. <http://www.sonami.net> (accessed on 27 February 2013).
- Spasov, M. 2009–10. *ENACTIV*. <http://www.keele.ac.uk/music/people/miroslavspasov/> (accessed on 17 December 2012).
- Spasov, M. 2011. Music Composition as an Act of Cognition: ENACTIV – interactive multi-modal composing system. *Organised Sound* **16**(1): 69–86.
- Steiner, M. 2006. *Data Glove Concert*. <http://www.youtube.com/watch?v=uIeNZZXNACJM> (accessed on 25 February 2013).
- Torre, G. 2012a. *Molitva (2009)*. <http://vimeo.com/44650248> (accessed on 27 February 2013).
- Torre, G. 2012b. *Mani (2010)*. <http://vimeo.com/44762821> (accessed on 27 February 2013).
- Torre, G. 2012c. *Agorá (2012)*. <http://vimeo.com/44650248> (accessed on 27 February 2013).
- Torre, G. 2013. *The Design of a New Musical Glove: A Live Performance Approach*. Unpublished PhD thesis. University of Limerick.
- Torre, G., Torres, J. and Fernstrom, M. 2008. The Development of Motion Tracking Algorithms for Low Cost Inertial Measurement Units. *Proceedings of the Conference on New Interfaces for Musical Expression, 2008*. Genova. <http://nime2008.casapaganini.org/documents/Proceedings/Papers/192.pdf> (accessed on 14 January 2013).
- Vertegaal, R., Ungvary, T. and Kieslinger, M. 1996. Towards a Musician's Cockpit: Transducer, Feedback

- and Musical Function. *Proceedings of the 1996 International Computer Music Conference*. Honk Kong/San Francisco: ICMA, 308–11.
- Vigliensoni, G. and Wanderley, M.M. 2009. *Soundcatcher: Explorations in Audio-Looping and Time-Freezing Using an Open-Air Gestural Controller* [http://www.music.mcgill.ca/~gabriel/courses/mumt620/Final\\_Project/Sound\\_Catcher\\_Final\\_Report.pdf](http://www.music.mcgill.ca/~gabriel/courses/mumt620/Final_Project/Sound_Catcher_Final_Report.pdf) (accessed on 2 June 2013). See also <http://vigliensoni.com/blog/soundcatcher> (accessed on 6 February 2013).
- Vigliensoni Martin, G. 2010. The Enlightened Hands: Navigating through a Bi-Dimensional Feature Space Using Wide and Open-Air Hand Gestures. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2011*. Sydney. [http://www.music.mcgill.ca/~gabriel/courses/mumt619/The\\_Enlightened\\_Hands/The\\_Enlightened\\_Hands\\_FINAL\\_REPORT.pdf](http://www.music.mcgill.ca/~gabriel/courses/mumt619/The_Enlightened_Hands/The_Enlightened_Hands_FINAL_REPORT.pdf) (accessed on 6 February 2013). See also <http://vimeo.com/10721519> (accessed on 6 February 2013).
- Waisvisz, M. 2003. *Michael Waisvisz: The Hands*. <http://www.youtube.com/watch?v=SifumZa2TKY&feature=related> (accessed on 27 February 2013).
- Wanderley, M.M. 2001. Gestural Control of Music. *Proceedings of the International Workshop Human Supervision and Control in Engineering and Music*. Kassel. <http://www.engineeringandmusic.de/individu/Wanderley/Wanderley.pdf> (accessed on 13 December 2012).
- Waters, S. 2000. The Musical Process in the Age of Digital Intervention. *Ariada 1*. [http://www.ariada.uea.ac.uk/ariadatexts/ariada1/content/Musical\\_Process.pdf](http://www.ariada.uea.ac.uk/ariadatexts/ariada1/content/Musical_Process.pdf) (accessed on 24 December 2012).
- Wessel, D., Wright, D. and Schott, D. 2002. Intimate Musical Control of Computers with a Variety of Controllers and Gesture Mapping Metaphors. *Proceedings of the International Conference on New Interfaces for Musical Expression, 2002*. Dublin. <http://opensoundcontrol.org/files/NIME02WesselWrightSchottDmo.pdf> (accessed on 14 December 2012).