Gestural control in electronic music performance: sound design based on the 'striking' and 'bowing' movement metaphors

Frederic Robinson
Elektronisches Studio Basel
Hochschule für Musik Basel
Basel, Switzerland
fredericanthonyrobinson
@gmail.com

Cedric Spindler
Elektronisches Studio Basel
Hochschule für Musik Basel
Basel, Switzerland
cedric.spindler@gmail.com

Volker Böhm
Elektronisches Studio Basel
Hochschule für Musik Basel
Basel, Switzerland
volker.boehm@fhnw.ch

Erik Oña
Elektronisches Studio Basel
Hochschule für Musik Basel
Basel, Switzerland
erik.ona@fhnw.ch

ABSTRACT

Following a call for clear movement-sound relationships in motion-controlled digital musical instruments (DMIs), we developed a sound design concept and a DMI implementation with a focus on transparency through intuitive control metaphors. In order to benefit from the listener's and performer's natural understanding of physical processes around them, we use gestures with strong physical associations as control metaphors, which are then mapped to sound modules specifically designed to represent these associations sonically. The required motion data can be captured by any low-latency sensor device worn on the hand or wrist, that has an inertial measurement unit with six degrees of freedom. A dimension space analysis was applied on the current implementation in order to compare it to existing DMIs and illustrate its characteristics. In conclusion, our approach resulted in a DMI with strong results in transparency, intuitive control metaphors, and a coherent audio-visual link.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; J.5 [**Arts and Humanities**]: Performing arts

General Terms

Algorithms, Design, Performance

Keywords

Gestural control, musical interface, DMI

AM15, October 07-09, 2015, Thessaloniki, Greece
© 2015 ACM. ISBN 978-1-4503-3896-7/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2814895.2814901

1. INTRODUCTION

With the growing availability of motion tracking technology, many digital musical instruments (DMIs) with gestural input have emerged and the connection between performer movement and sound has become a broad and active field of research. The information conveyed by a performer's bodily involvement is a powerful element in musical performance, as visual stimuli strongly influence the audience's perception of auditory events [13, 19, 9, 17, 21]. Most electronic music performances with industry-standard control devices contain little movement that can be described as "effective gestures," which Delalande defines as gestures, that are required in order to produce a sonic result [4]. Consequently, a common risk in these performances is an apparent disconnection from the music. In our view, designing DMI interaction with a focus on effective performer gesture encourages bodily involvement and addresses this problem.

While available sensors, controller design, and parameter mappings are the object of thorough discussion, the idiosyncrasies of different sound design concepts and their role in the DMI design process have received less attention. In our view, the choice of sound source, whether synthesis algorithms or recorded material, is not just an expression of the designer's personal aesthetic preference, but an important element of the design process, which can provide solutions to problems that are usually only addressed with mapping strategies. In this paper, we present a sound design concept and a DMI implementation[1], which puts its focus on the *physicality of sound.* Auditory events are designed to have clear physical associations which correspond to those of the performer's gestures. This provides a clear audio-visual link even before complex mapping strategies are involved, as the function of the mapping layer is simplified to linking coherent visual and auditory events or processes rather than imposing gestural information on the latter.

[1] A demonstration video can be found at https://vimeo.com/geps/excerpts

In the context of this paper, the term "gesture" refers to a dynamic movement carried out by the performer. We subscribe to Kurtenbach and Hulteen's definition of "gesture" as motion that contains information, which excludes operating control devices such as keyboards, joysticks, etc. [10]. In our definition of "physical" and "physicality", as applied to sound, the terms describe the apparent energy content and movement perceived in the sonic material, rather than the methods of sound creation (as for example in acoustical instruments or physical modeling).

2. BACKGROUND

2.1 The Audio-Visual Link

While many DMI design approaches aim for "control subtlety," "intuitiveness," "expressivity," and the potential for "virtuosity," there has also been a call for clear and transparent connections between movement and sound, as clarity in a DMI's cause/effect mechanisms is beneficial for both performers and audiences [20, 12, 5, 6, 7]. Various works have linked audience perception to other elements of performance. According to Fyans et al., it has to be included in the objective evaluation of a performer's expression and skill, as they are not measurable quantities inherent in and dictated by the interface [6]. In a subsequent study, he shows that an audience's understanding of the performer's interaction and intention raises their understanding of error in performance, an important factor in the overall judgement of the performance [7]. Wessel et al. link a responsive, low-latency interaction with compelling control metaphors to what they call "control intimacy". Compelling control metaphors are control gestures intuitively understood by performer and audience [24]. Fels et al. even go so far as to state that the expressivity of a DMI can be predicted by assessing the performer's and the audience's understanding of the instrument. By their definition, DMIs fully understood by both parties are the most expressive [5]. While we believe that it is questionable to define transparency as the only prerequisite for expressivity, the role it certainly plays should not be underestimated.

This all provides compelling reasons for a design focused on transparency through intuitive control metaphors. Wessel et al. propose metaphors such as "drag and drop" or "scrubbing" [24]. Different metaphors can be taken from a vocabulary proposed by Lewis and Pestova, whose gestural typology for mixed electronic music, including live performance and acousmatic music, includes "striking," "pushing," "agitating," etc [11]. In our view, a vocabulary like this is helpful for creating intuitive gesture-sound relations, as both performer and audience can intuitively relate sonic events to observed actions through our natural understanding of physical processes around us. Our approach detects movement and controls sounds which can be described based on this vocabulary.

2.2 Excitation Gesture

We see a strong relation between transparency in performer interaction and intention, and the type of gestural input a DMI is designed for. In our view, the most suitable input is what Cadoz calls "excitation gesture" [2, 3], which refers to gestures performed on acoustical instruments that result in energy being sent into a vibrating structure. Excitation gestures and their associations are well known to the listener from the realm of acoustic instrumental performance.

A range of DMIs have included the concept of excitation to some extent. Next to sensing discrete data such as button presses, the *Quarterstaff* detects movement peaks in a performer's swinging of the device [15]. The *Twister* was developed out of a movement-based design process and includes mapping the speed of a performer's action (rotation in this case) onto the speed of sound generators [23]. In a study examining the success of mappings of various complexity, the most complex and successful mapping employed the speed of mouse movement as a volume control [8]. Sound is thereby only heard when the mouse is moved. In the piece "*Agorá*," the *Pointing-at* data glove controls volume in a similar way. Changes in orientation trigger energy input into a spring emulation, which ultimately controls the volume of sound files [20]. The decay of the springs results in silence when the performer is motionless. We see potential in a prioritization of excitation gestures in performer interaction, and our approach almost exclusively focuses on this type of movement. However, we additionally extend the concept of excitation and physicality to the design of the sound modules in order to create the most natural sonic equivalent to the player's gestures.

3. GESTURE METAPHORS AND SOUND DESIGN

In this paper, we present two gesture metaphors with physical connotations and their sonic representation: *striking*, and *bowing*[2]. Both of these can be found in the vocabulary suggested by Lewis and Pestova [11]. In the context of this implementation, we refer to "performer movement" as movement of the arm or hand. The sound modules have been created with the intention of producing sounds readily associable with physical movement. The sound aesthetics are oriented towards the sound vocabulary of acousmatic music, because, in our view, it lends itself well to association with physical forces. The main focus is on varyingly energetic impacts, an organic flow of energy in the sound material, and a realistic representation of excitation and its subsequent resonance. To achieve this, we suggest the use of artificial resonators (feedback networks) as well as various sampling techniques on recorded material.

3.1 Striking

A *striking* gesture is a sudden change in the performer's movement, which can range from flicking a finger while the hand is motionless to suddenly interrupting a swing of the arm. These gestures are clear visual cues, and their precise time of occurrence can be determined in the motion data stream independently from the overall movement of the performer. More energetic *striking* requires the performer to build up momentum beforehand, in which case the motion before a *strike* can be used to determine the gesture's energy content. Subsequently, we know when a *strike* took place and how much energy was used to perform it.

[2] Both gestures and their sonic representations are demonstrated at https://vimeo.com/geps/metaphors

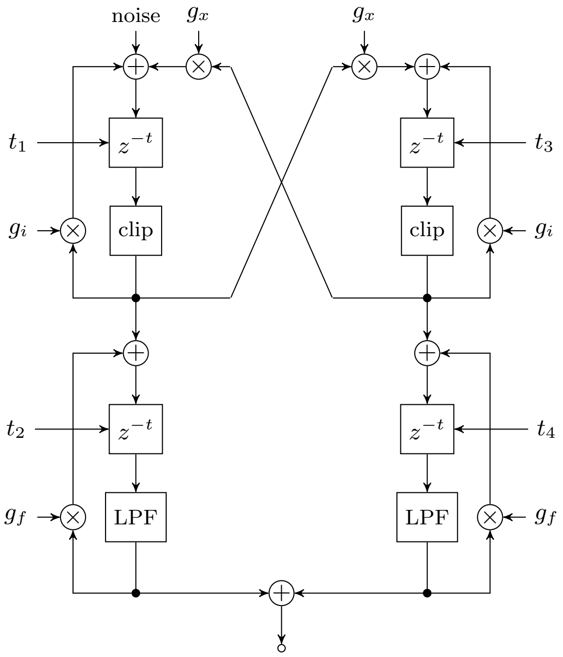

Figure 1: Digital Feedback Network

The sound material mapped to *strikes* consists of recorded impacts which have been analyzed and sorted by their energy content, providing a fixed sound set in a range of intensity which includes changes in volume, resonance and spectral features. Physical associations can vary from the perceived knocking on a piece of glass to the beating of a raindrum. Connecting the gesture's and sound material's energy content in a process commonly referred to as "one-to-one mapping" [8, 14] convincingly fuses the perceived physicality of both, provided that the sound material evenly covers a broad dynamic range.

3.2 Bowing

The *bowing* gesture is where the concept of "excitation" is most apparent. In the case of acoustic instruments, bowing continually sends energy into a vibrating structure. Stopping the excitation leaves the sounding body resonating if no damping is applied. More energetic *bowing* increases volume and spectral content. Lewis and Pestova expand the definition of *bowing* to actions like running a wet finger along the rim of a wine glass [11]. In our implementation, *bowing* is any kind of continuous movement of the arm. This controls the strength of excitation, while the orientation of the hand determines the characteristics of the sounding body. For example, circular arm motion with the palm facing down emulates the above-mentioned finger on a wine glass, while arm movement away from the body with the back of the hand facing outward resembles brushing an open string.

Several sound design approaches are used to adequately represent different types of bowing sonically. We use a digital feedback network loosely based on digital waveguide synthesis [18], but with modulated delay times (see Figure 1). The network consists of parallel and serial delay lines, filters, and nonlinear elements such as clip distortion. Constant low level noise provides the system with energy, which is then built up or decreased by the feedback coefficients $g_{i,x,f}$. Varying delay times $t_{1,2,3,4}$ in each delay line result in a thick cluster of individually moving pitches. Clip distortion saturates the signal if a certain energy level is exceeded, resulting in a richer spectrum. By applying "divergent mapping" (mapping one control parameter to several synthesis parameters [14]), we connect the strength of the bowing gesture to the feedback coefficients (i.e. perceived energy), and to the speed and amplitude of the delay time variations (i.e. perceived internal movement). A low energy gesture therefore results in a soft cloud of slowly moving low pitches with little high frequency content, while a high energy gesture results in a loud, bright cluster of broadly oscillating pitches.

In another module, the down-sampled spectrum of the above-mentioned feedback network serves as a low-resolution modulator for cross synthesis with recorded material. This allows linking the *bowing* gesture to any continuous sound with a broad and flat spectrum and little internal movement, while preserving the movement-sound relationship, which was designed using the parameters of the feedback network.

A third module (exemplified by the wine glass metaphor) uses the sound of *bowing* (and its variations) recorded at several levels of fixed intensity. The recordings are combined to form a multi-layer sample, representing the excitation of a fixed sonic material with variations in performed effort. In a one-to-one mapping, the strength of the *bowing* gesture then determines the intensity of the sounds chosen from the module's library. The choice of pre-recorded sound continually adapts to the performer's energy input, attempting to emulate the movement-sound relationship present in the actual recording process.

4. HARDWARE DESIGN AND DATA INTERPRETATION

4.1 Hardware

The motion data required for the sound mappings presented in this paper can be provided by any small sensor device located on the hand or wrist that sends accelerometer and gyroscope data at a sampling rate above 100 Hz. This includes most data gloves. We used a custom-built unit (see Figure 2) positioned at the wrist to minimize latency and size[3].

The sound modules could also be controlled with motion-sensing objects such as cell phones, game controllers, and many DMIs. However, when the performer interacts with actual physical objects, the proposed gesture metaphors start to lose meaning as the control device becomes the mediator between performer gesture and sound.

4.2 Data Interpretation

The two player gestures presented here are extracted through explicitly defined detection algorithms. More complex methods involving machine learning can be used to further differentiate the gesture repertoire, but are not required for the basic functionality of the DMI.

[3] Building instructions for an older model can be found at http://geps.synack.ch/doc-build.html

4.2.1 Striking Detection

By computing the derivatives $g'(t)$ of the data streams from each axis of the gyroscope and consequently summing their absolute values, we disregard the information on three-dimensional rotation and observe changes in overall movement energy $m(t)$ instead.

$$m(t) = \left(\frac{|g'_x(t)| + |g'_y(t)| + |g'_z(t)|}{3}\right)^2 \tag{1}$$

Striking, the sudden start or end of a performer gesture, is then recognized by detecting peaks in this data stream.

$$s = m(t-1) < m(t) > m(t+1) \land m(t) \geq \vartheta \tag{2}$$

Treating the gyroscope data in the same way but without computing the derivatives results in a control signal that represents the momentum built up before a *striking* gesture (see Formula 3). Sampling this data stream at the point of the *strike* detection provides a reliable indicator of the energy of the performer gesture.

4.2.2 Bowing Detection

The control signal $b(t)$ for *bowing* gestures is computed by summing the absolute values of the three gyroscope axes $g(t)$ (see Formula 3) and subsequently applying an adaptive low-pass filter, whose behaviour varies depending on whether its input increases or decreases (see Formula 4). Changing the filter coefficients (α_1, α_2) then emulates different responses to energy input. For example, a fast increase and slow decrease represents a system that can be easily excited by movement energy and keeps releasing the energy slowly over time.

$$q(t) = \frac{|g_x(t)| + |g_y(t)| + |g_z(t)|}{3} \tag{3}$$

$$b(t) = \begin{cases} b(t-1) + \alpha_1\Big(q(t) - b(t-1)\Big), & \text{if } q(t-1) < q(t) \\ b(t-1) + \alpha_2\Big(q(t) - b(t-1)\Big), & \text{if } q(t-1) \geq q(t) \end{cases} \tag{4}$$

Bowing gestures are differentiated by orientation of the hand, which is detected by examining all three accelerometer axes and defining the required hand orientation with a combination of logical operators and temporal thresholds.

5. ANALYSIS

We evaluated our implementation by applying the *Dimension Space Analysis* proposed by Birnbaum et al. in order to illustrate its characteristics and to compare it to existing instruments [1] (see Figure 3). The characteristics of the two underlaid DMIs, Waisvisz' "The Hands" [22] and the Theremin are taken from the abovementioned article. Our implementation's evaluation is based on a subjective assessment carried out by the authors, who regularly use the DMI in a performance context. As the characteristics of all three instruments have not been verified through user tests, the analysis provides a framework for discussing design choices rather than informing on user perception.

As observed by Birnbaum et al., the arrangement of axes causes the plots of this analysis to group installations on the left and instruments on the right of the dimension space [1]. In our case the result turned out to be a classification in the instrument domain.

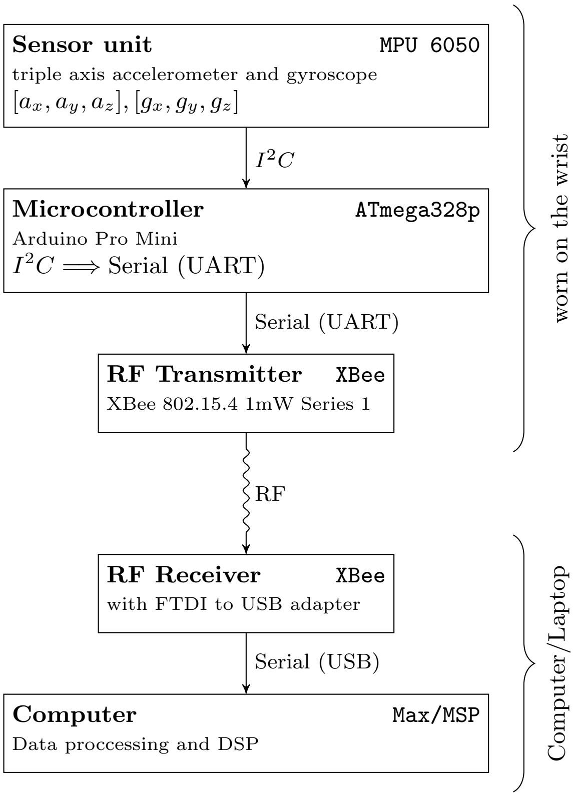

Figure 2: Hardware Components and Data Flow

Required Expertise: The performer has to explore the functionality of the instrument by playing it. In order to perform one should be familiar with the gestural repertoire and its link to the audible result, which results in a moderately high level of required expertise. The coherent cause-effect link and the limitation in the number of available gestures makes the instrument more accessible for a performer than "The Hands", with their higher number of input controls, or the Theremin, which requires precise positional input.

Musical Control: The DMI allows *timbral control* as well as control of individual sound events (*note level*) and their combination (*musical processes*). These categorizations, as proposed by Schloss, are not mutually exclusive and can be seen as control in different levels of detail [16]. As in acoustical instruments, the timbre is often linked to dynamics and other sound parameters. This is a result of the sound modules' focus on physicality. "The Hands" provide timbral control as well, while the Theremin remains on note-level with volume and pitch control.

Feedback Modalities: Much focus has been put on providing an intuitive and natural auditory feedback. As the position and size of the sensor-unit precludes the actual or apparent existence of a physical object of manipulation, the implementation relies on the performer's proprioception to link the audible result to gestures. Therefore, the system can be considered as having a kinesthetic feedback mode as well. The Theremin is not designed for a realistic sonic response to the performer's movements, which results in a

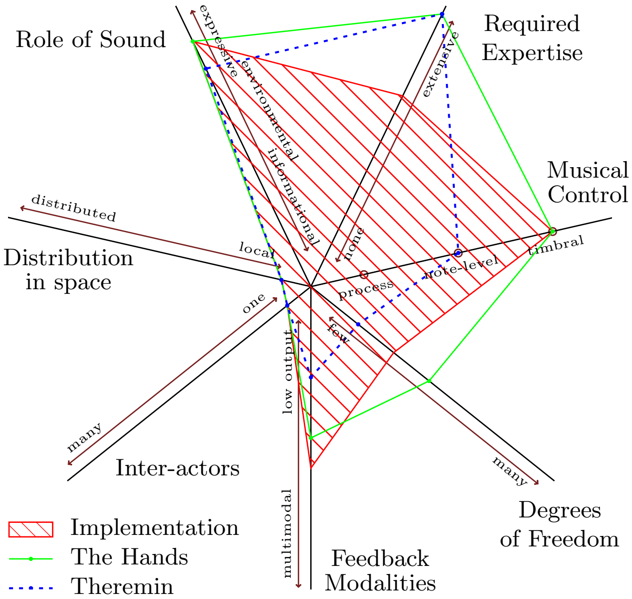

Figure 3: Dimension Space Analysis

lower score. “The Hands” provide haptic feedback through their buttons controls.

Degrees of Freedom: The specific function of the sound modules (representing a certain type of physical movement) limits the number of each module’s input controls, as the performed gesture has to be coherent with the underlying metaphor. Therefore, the overall degrees of freedom are dependent on the available gesture repertoire. The absence of meta-control of larger musical processes further reduces the score. The combination of motion data with other sensors provides “The Hands” with more degrees of freedom. The Theremin is limited to two input controls. We have therefore given it a lower score than originally proposed by Birnbaum et al..

Inter-actors and **Distribution in Space**: The DMI is designed to be played by one person alone.

Role of Sound: Our implementation is an instrument with an artistic and expressive role. Sound plays an important role which almost overshadows that of the performer, as he or she often seems to cause and shape sonic processes rather than creating and controlling them. While it has exploratory potential, exploration is part of the learning process and not of the performance itself. “The Hands” and the Theremin are both artistic and expressive as well, with the Theremin scoring slightly lower because of its focus on melody and the exclusion of timbre control.

6. CONCLUSIONS

Based on reports that emphasize the importance of transparency, intuitive control metaphors and a clear audio-visual link in DMI design, we proposed a performance system addressing these demands. We suggested the use of excitation gestures with their intuitively understandable physical associations as meaningful control metaphors and presented a sound design concept capable of representing these associations sonically. The resulting DMI provides a very clear cause/effect link for both performer and audience, and allows what we believe to be convincing gestural control.

The clearly defined set of gesture metaphors requires a specific sonic vocabulary which favors sounds with obvious physical associations. Many traditional synthesis techniques are therefore not suited for use with our DMI implementation. Because of limitations to the number of available input controls for individual sound modules, complex and refined performances demand the simultaneous use of several modules. The system is thereby better fit for solo performances than ensemble play.

Further work can be done to expand the gesture repertoire, which may require additional sensor technology. More control over individual sound elements can be achieved by adding control parameters to the sound modules which do not contradict the established physical associations.

7. REFERENCES

[1] D. Birnbaum, R. Fiebrink, J. Malloch, and M. M. Wanderley. Towards a dimension space for musical devices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 192–195, Vancouver, BC, Canada, 2005.

[2] C. Cadoz. Proceedings of the international computer music conference. In *Proceedings of International Computer Music Conference*, pages 1–12, San Francisco, 1988.

[3] C. Cadoz and M. M. Wanderley. *Trends in Gestural Control of Music*, chapter Gesture-Music, page 101. Ircam - Centre Pompidou Paris, IRCAM - Centre Georges Pompidou, 2000.

[4] F. Delalande. Le geste, outil d’analyse: quelques enseignements d’une recherche sur la gestique de glenn gould. *Analyse Musicale: Geste et Musique*, 10:43–46, 1988.

[5] S. Fels, A. Gadd, and A. Mulder. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, 7(2):109–126, aug 2002.

[6] A. C. Fyans, M. Gurevich, and P. Stapleton. Where did it all go wrong? a model of error from the spectator’s perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 171–172, Pittsburgh, PA, United States, 2009.

[7] A. C. Fyans, M. Gurevich, and P. Stapleton. Examining the spectator experience. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 451–454, Sydney, Australia, 2010.

[8] A. Hunt and M. M. Wanderley. Mapping performer parameters to synthesis engines. *Organised Sound*, 7(2):97–108, aug 2002.

[9] J. Juchniewicz. The influence of physical movement on the perception of musical performance. *Psychology of Music*, 36(4):417–427, 2008.

[10] G. Kurtenbach and E. A. Hulteen. Gestures in human-computer communication. *The art of human-computer interface design*, pages 309–317, 1990.

[11] A. Lewis and X. Pestova. The audible and the physical: a gestural typology for mixed electronic music. In *Proceedings of the Electroacoustic Music Studies Network Conference*, Stockholm, June 2012.

[12] M. Marier. The sponge: A flexible interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 356–359, Sydney, Australia, 2010.

[13] S. J. Morrison, H. E. Price, E. M. Smedley, and C. D. Meals. Conductor gestures influence evaluations of ensemble performance. *Frontiers in Psychology*, 5(806), 2014.

[14] J. B. Rovan, M. M. Wanderley, S. Dubnov, and P. Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Kansei, The Technology of Emotion. Proceedings of the AIMI International Workshop*, pages 3–4, 1997.

[15] J. C. Schacher. The Quarterstaff, a Gestural Sensor Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 535–540, Daejeon, Republic of Korea, may 2013. Graduate School of Culture Technology, KAIST.

[16] W. A. Schloss. Recent advances in the coupling of the language max with the mathews/boie radio drum. In *Proceedings of the International Computer Music Conference*, pages 398–400, Glassgow, 1990.

[17] M. Schutz and S. Lipscomb. Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36(6):888–897, 2007.

[18] J. O. Smith III. Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–91, 1992.

[19] W. F. Thompson, P. Graham, and F. A. Russo. Seeing music performance: Visual influences on perception and experience. *Semiotica*, 2005(156):203–227, 2005.

[20] G. Torre. *The design of a new musical glove: a live performance approach*. PhD thesis, University of Limerick, 2013.

[21] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin. Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113, 2006.

[22] M. Waisvisz. The hands: A set of remote midi-controllers. In *Proceedings of the International Computer Music Conference*, pages 313–318, San Francisco, 1985. Ann Arbor, MI: MPublishing, University of Michigan Library.

[23] N. Ward and G. Torre. Constraining movement as a basis for dmi design and performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 449–454, London, United Kingdom, 2014. Goldsmiths, University of London.

[24] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22, 2002.