Use of Body Motion to Enhance Traditional Musical Instruments

A Multimodal Embodied Approach to Gesture Mapping, Composition and Performance

Federico Visi
Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University, UK
federico.visi@plymouth.ac.uk

Rodrigo Schramm
Institute of Informatics
Federal University of Rio Grande do Sul, Brazil
rodrigos@caef.ufrgs.br

Eduardo Miranda
Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University, UK
eduardo.miranda@plymouth.ac.uk

ABSTRACT

This work describes a new approach to gesture mapping in a performance with a traditional musical instrument and live electronics inspired by theories of embodied music cognition (EMC) and musical gestures. Considerations on EMC and how gestures affect the experience of music inform different mapping strategies. Our intent is to enhance the expressiveness and the liveness of performance by tracking gestures via a multimodal motion capture system and to use motion data to control several features of the music. We then describe an application of such approach to a performance with electric guitar and live electronics, focusing both on aspects of meaning formation and motion capturing.

Keywords

NIME, gesture, multimodal, mapping, embodied music cognition, gesture recognition, liveness, expressiveness, guitar.

1. INTRODUCTION

In the last two decades, the study of music and gesture has been subject to increasing cross-disciplinary interest, moving away from a primarily musicological context and engaging multiple disciplinary fields. Extensive interdisciplinary research has recently been carried out [9, 10, 11] giving rise to new paradigms for the understanding of gesture in music. In particular, insights from research on embodied music cognition [13] inspired new viewpoints that required a rethink of the foundations of musical gestures. Within this theoretic framework, music perception is *embodied* (i.e. closely linked with bodily experience) and *multimodal*, in the sense that music is perceived not only through sound but additionally with the help of both visual cues and sensations of motion, effort and dynamics [8]. Hence gestures become a core notion as they act as a bridge between bodily movement and meaning formation.

Concurrently, motion sensing technologies have considerably developed and become pervasive as an increasing number of consumer-grade electronics employ motion data for accomplishing different functions. Such technologies are easily accessible and are being employed in both academic research and artistic practice. In musical contexts, there is a long and prolific tradition of electronic interfaces that exploit gestures and motion as means of control of musical parameters [19]. As Miranda and Wanderley also note, adopting an effective mapping strategy is crucial for the expressiveness of the interface, being the relationship between gestural variables and musical parameters often far from obvious. Mapping has therefore become a key topic in digital musical instrument research and several approaches have been documented in the proceedings of NIME conferences through the years.

Within this context – where a considerable cross-disciplinary theoretical apparatus has come of age while continuously developing and motion sensing technologies have become ubiquitous – this work aims at giving an account of approaching gesture mapping in a performance with a traditional musical instrument and live electronics, taking into consideration recent theories of embodied cognition and musical gesture and employing multiple motion-sensing technologies.

2. GESTURES AND EMBODIED MUSIC COGNITION

The concept of *gesture* is used across multiple disciplines and contexts and its definition is quite broad and sometimes vague. Jensenius et al. [12] give a comprehensive look at the term and its uses in music research in order to present a clearer overview. While initially problematic to pinpoint, the notion of gesture has considerable potential in modern music research as it works as a bridge between movement and meaning and, consequently, bypasses the boundary between physical world and mental experiences. This is at the core of the embodied music cognition paradigm [13]. In this context, "musical gestures can be described in an objective way as movement of body parts, but they have an important experiential component that is related to intentions, goals, and expressions" [15]. The body is thereby understood as a mediator between the physical environment (e.g. music as sound waves moving in the air) and the subjective musical experience (e.g. one's feelings in response to that music). By acting as a mediator, the body will build up a repertoire of gestures and gesture/action consequences, or what Leman calls a *gesture/action-oriented ontology*[15]. This repertoire can be considered as a collection of movements made to achieve a particular goal (actions) linked with the experiences and sensations resulting from such actions. Musical gestures are at the core of this repertoire.

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

The coupling of actions and perceived sensations forms an engine that guides our understanding of music. Through this mechanism, the listener is able to relate physical aspects of movement in space to expressive qualities, intentions and inner feelings. Conversely, perceived patterns of musical expression recall previously learned knowledge of the corresponding body movement. This continuous two-way mirroring process allows the listener not only to attribute intentions and feelings to music but also to predict the outcomes of actions and project them onto the music [17]. This is what Leman calls *action-perception coupling system*; it forms the basis of musical intentional communication and expressiveness, which then elicits several social phenomena such as empathy and social bonding [13, 14, 15].

Overall, being at the core of embodied music cognition, we could say that *gestures are a vehicle for the construction of musical meaning.* This means not only that music perception is embodied, but that it is also *multimodal* in the sense that we perceive it using multiple senses; through sound but also with the help of visual images and feelings of motion, such as kinaesthetic sensations and kinematic images [8].

From the perspective of embodiment, it is clear that gestures play an important role in how we perceive and attribute meaning to music and how we express ourselves through music. To better understand the functions of musical gestures in performance, Jensenius et al. [12] identify four functional categories; *sound-producing* gestures, *sound-facilitating* gestures, *sound-accompanying* gestures and *communicative* gestures. While their boundaries are usually blurry – mostly because in actual musical performances gestures naturally tend to have multiple functions – functional categories are a useful starting point for making sense of the role a gesture may have in a performance. Further practice-led research may bring more detailed insights on the functions gestures may have in performances with a particular instrument or within different musical contexts.

3. TRADITIONAL MUSICAL INSTRUMENTS

In this context, the term *traditional musical instrument* (henceforth TMI), is used to define instruments that have a rich idiomatic repertoire that ranges across both popular music and classical music. Examples include electric guitar, violin, electronic keyboards, etc. Such a term is preferred to acoustic instrument because it refers to the use and repertoire rather than to technological aspects. The choice of focusing on TMI performance is motivated by the vast knowledge that listeners have of the gestural and sound aspects of each instrument, which is learnt through experience. This can be summarised as the *ecological knowledge* [8] of an instrument; listeners have, and in some cases share, *a repertoire of sound-producing gestures.* TMIs also have fairly explicit and known *affordances* [6] that can delineate action relationships between the instrument and the musician, inform expectations in the listeners and be used, along with sound affordances, to devise mapping strategies for controlling electronic aspects of the performance. Using embodied music cognition terminology, instruments have a rich action/gesture repertoire that the listeners can recognise during the performance. With this in mind, using the instrumentalist's gestures may have a considerable expressive potential in performance as well as in composition as composers would be able to draw from a *gestural palette* of the instrument when writing a piece.

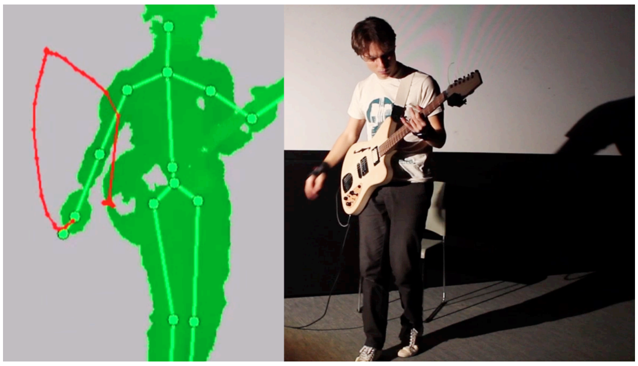

Figure 1: Multimodal motion capture setup featuring optical motion capture and wearable sensors.

4. APPLYING THE THEORY TO A CASE STUDY: THE ELECTRIC GUITAR

Given this background scenario, it is clear that gestures have a significant influence on how music is experienced and traditional musical instruments are a rich repository of shared gestural information. In recent years, there have been applications of EMC within interactive multimedia environments [2] and singing performance [16].

Here, we are proposing an initial approach to the development of a performance with electric guitar and live electronics involving multiple motion capture technologies, namely flex sensors and accelerometers located on hands and wrists of the guitarist and a Microsoft Kinect to monitor full-body movements. The choice of placing sensors on the performer's hands is motivated by the need to have more accurate and detailed data of hand movements since sound-producing gestures of guitar playing – as with many other instruments – typically involve hands and arms and such gestures are the most readily noticed by an observer [3]. This was done using flex sensors and 3-axis accelerometers mounted on custom wristbands together with a custom Arduino-based board equipped with an XBee wireless chip[1]. Conjointly, the Kinect allows the extraction of full body movements in different lighting situations, which is desirable in a staged context. On the other hand, a drawback of such technology, is that the quality of the skeleton tracking depends on factors like camera position and occlusion of body parts. Indeed, the presence of a musical instrument normally affects the accuracy of the skeleton tracking algorithm. Another limit of this sensor is its low frame rate (30 fps), which may be a problem for real time audio processing. In this case, employing a multimodal motion capture system allows to use the more stable signal of the wearable sensors to capture the subtle movements of hands and wrists while the Kinect data, even at low frame rates or with jittery joint tracking, allows for the interpretation of more complex gestures from distinct parts of the body, enabling the recognition of high-level expressive movements.

Parameter mappings are arranged in a layered conceptual framework (low-level and high-level), reflecting the layered approach to musical meaning formation as described by Leman [14].

5. MAPPINGS: LOW-LEVEL LAYER

These are low level mappings related to low-level characteristics of the movement and they are tracked using single skeleton points captured by the Kinect and the output values of the accelerometers and flex sensors. This layer

[1] `http://www.sensestage.eu`

comprises the use of one-to-one and one-to-many mapping strategies [19] relating a single gesture to sonic events in a transparent fashion. The motion data goes through simple signal processing – including rescaling, smoothing and compression – using OSC and MIDI in a data routing application[2]. The processed data is then used to control various DSP patches. These mappings are devised using functional categories of gestures in order to reinforce and play around the meaning and expectations associated to each gesture. For example, turning the palm of the strumming hand upwards and therefore interrupting one of the most evident *sound-producing gestures* (string plucking) triggers a granular sampler controlled via the wearable sensors that freezes and sustains the sound of the guitar, reflecting the suspension in the performance. *Sound-facilitating phrasing gestures*, like the movement of the head to express entrainment and follow features of the music, are instead captured via the Kinect and mapped onto parameters of a ring modulator which modify the overall timbre of the instrument.

6. MAPPINGS: HIGH-LEVEL LAYER

This layer contains mapping strategies relating more complex gestures to structural changes in the arrangement of the electronic parts of the piece and in the global sound. Motion data is therefore used to recognise gesture trajectories and semantic and cultural aspects of the performance are also taken in consideration.

6.1 Gesture recognition

Specific gesture trajectories are recognised and are useful for controlling structural features of the performance, such as the introduction of loops and samples and other more complex parts of the piece. In addition, gesture trajectories are employed to assign low-level gestures to other musical parameters, allowing the performer to control different features later in the piece. In this case, the skeleton joints tracked with Kinect are used to define a set of N-Dimensional temporal features, which are then evaluated in real time to map gestures. To realise this task, temporal classification algorithms like the Dynamic Time Warping - DTW [20] and Hidden Markov Models-HMM [5, 1] were extensively studied and experimented.

An implementation based on the N-Dimensional DTW technique described by Gillian et al. [7] was used in this work. As a result, the proposed system can learn specific gestures through a machine learning system and recognise them during the performance. Figure 2 illustrates an example where two different types of gestures are recognised.

6.2 Context (social, semiotic)

As noted by Leman [14], social context contributes indirectly to musical meaning formation. In live guitar performance, playing seated is a typical feature of acoustic performances while playing standing is usually associated with jazz and rock contexts where a more engaged performance body is expected [21]. Playing standing or seated has a meaning for the audience that relies upon previously learned social and aesthetic aspects. This mapping underlines these aspects by varying the overall timbre of both the guitar and the electronic parts according to the posture of the performer.

Neither a direct low level mapping nor the gesture recognition procedure are appropriate to interpret the posture of the performer. For a better mapping in this situation, two measures were used by Fenza et al. [4]:

[2] http://steim.org/product/junxion/

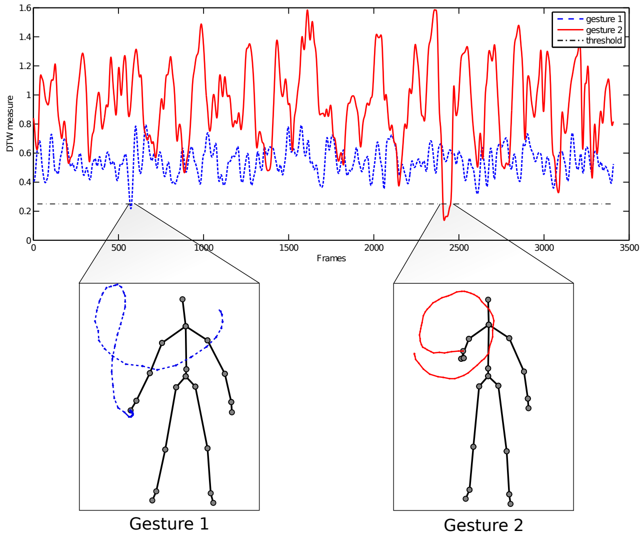

Figure 2: Gesture recognition task. The dashed and solid lines show the DTW distance from each gesture template. The dashed horizontal line shows the classification threshold. In this case, gesture 1 and gesture 2 were detected at frame 567 and at frame 2396, respectively.

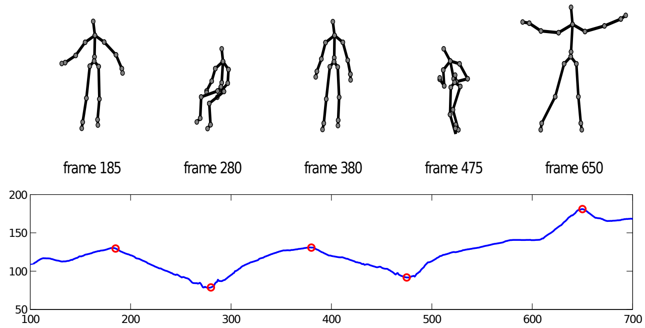

Figure 3: Contraction Index variation (solid line) of performer's body along a recorded time sequence. Skeletons (on top) illustrate the posture on each time stamp (marks).

- **Contraction Index - CI**: This measure is the sum of the distances between the skeleton barycenter and each skeleton point and provides information about the contraction and expansion of the body.
- **Quantity of Motion - QoM**: The QoM is proportional to the distance of movement, that is, it gives high values when the skeleton body is moving fast and low values when the it is more stationary.

These two values, measuring the contraction/expansion and the quantity of motion of the body, can be used to control structural features of the piece being performed. For example, intense physical engagement can trigger new sampled parts of the piece or assign the low-level gestures to a new set of musical parameters.

To illustrate this, Figure 3 shows a time-sequence during which the performer changes its posture, from standing up (frame 185) to sitting down (frame 280) and finally with the arms wide open (frame 650). The line shows the respective Contraction Index along the time.

7. CONCLUSION AND FUTURE WORK

The paradigm of embodied music cognition provides interesting insights on musical gestures and their role in our experience and understanding of music. The assumption that music is based on a tight relationship between sounds and experiences that are mediated by the body have many significant implications. From this perspective, gesture is not subordinated to sound by simply acting as a means of control, but rather it contributes to the formation of meaning, acting on the same level of sound.

If we take into account that the action/perception coupling is at the core of embodied music cognition, mapping can even more be considered as a translation layer that must be carefully designed. An effective mapping strategy is in fact pivotal to obtain good results in terms of expressiveness and liveness of performance and it can also be said that it constitutes an integral part of composition, as gestures convey meaning analogously to sounds.

Within this context, it is evident that technology plays an important role. However, different systems can be employed to reach the same expressive result working on the same gesture/sound relations, showing that the approach is to some extent hardware-agnostic.

Future work may be done to explore the gestural repertoire of other instruments. Functional categories are a good starting point for understanding the role of individual gestures. However, their boundaries are very blurry due to the multi-functional nature of gestures, therefore the development of more precise tools for semantic analysis may be helpful.

Embodied music cognition is still a relatively young research field and new contributions to the understanding of some of its inner mechanisms – such as mirroring, common coding and enaction – have been recently published [17, 18]. In addition, there are few examples of practical applications outside academic research (e.g. [16]). Gathering more empirical evidence about the formation of embodied musical meaning through academic experimentation may result in new notable insights. Nonetheless, intuitions from more practice-led, artistic environments may give considerable contributions to the research.

As Godøy notes in [8] Western musical thought has not been well equipped for thinking the inclusion of musical elements within the context of a gesture. Further research into gestural and embodied aspects of music and practice of a more *gesture-aware* music might bring about unprecedented insights and perhaps lead to the emergence of new aesthetic categories.

References

[1] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana. Continuous realtime gesture following and recognition. In *Proceedings of the 8th International Conference on Gesture in Embodied Communication and Human-Computer Interaction*, GW'09, pages 73–84, Berlin, Heidelberg, 2010. Springer-Verlag.

[2] A. Camurri, G. D. Poli, M. Leman, and G. Volpe. A multi-layered conceptual framework for expressive gesture applications. In *Proceedings of MOSART: Workshop on Current Directions in Computer Music*, pages 29–34, 2001.

[3] S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana. Gesture in performance. In R. I. Godøy and M. Leman, editors, *Musical Gestures: Sound, Movement, and Meaning*. Routledge, 2010.

[4] D. Fenza, L. Mion, S. Canazza, and A. Rodà. Physical movement and musical gestures: a multilevel mapping strategy. In *Proceedings of Sound and Music Computing Conference*, Salerno, 2005.

[5] G. A. Fink. *Markov Models for Pattern Recognition: From Theory to Applications.* Springer-Verlag, Berlin Heidelberg, 2008.

[6] J. J. Gibson. The theory of affordances. In *Perceiving, Acting, and Knowing.* Erlbaum, 1977.

[7] N. Gillian, B. Knapp, and S. O'Modhrain. Recognition Of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 337–342, 2011.

[8] R. I. Godøy. Gestural affordances of musical sound. In R. I. Godøy and M. Leman, editors, *Musical gestures: Sound, movement, and meaning.* Routledge, 2010.

[9] R. I. Godøy and M. Leman, editors. *Musical Gestures: Sound, Movement and Meaning.* Routledge, 2010.

[10] A. Gritten and E. King. *Music and Gesture.* Ashgate, 2006.

[11] A. Gritten and E. King. *New Perspectives on Music and Gesture.* SEMPRE studies in the psychology of music. Ashgate, 2011.

[12] A. R. Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman. Musical gestures: Concepts and methods in research. In R. I. Godøy and M. Leman, editors, *Musical gestures: Sound, movement, and meaning*, pages 12–35. Routledge, 2010.

[13] M. Leman. *Embodied Music Cognition and Mediation Technology.* MIT Press, 2008.

[14] M. Leman. Music, gesture, and the formation of embodied meaning. In R. I. Godøy and M. Leman, editors, *Musical Gestures: Sound, Movement and Meaning*, pages 126–153. Routledge, 2010.

[15] M. Leman. Musical gestures and embodied cognition. In T. Dutoit, T. Todoroff, and N. d'Alessandro, editors, *Actes des Journées d'Informatique Musicale (JIM 2012)*, pages 5–7, Mons, Belgique, 9-11 mai 2012. UMONS/numediart.

[16] P.-J. Maes, M. Leman, K. Kochman, M. Lesaffre, and M. Demey. The "one-person choir": A multidisciplinary approach to the development of an embodied human-computer interface. *Computer Music Journal*, 35(2):22–35, 2011.

[17] P.-J. Maes, C. Palmer, M. Leman, and M. Wanderley. Action-based effects on music perception. *Frontiers in Psychology*, 4, 2014.

[18] J. R. Matyja and A. Schiavio. Enactive music cognition: Background and research themes. *Constructivist Foundations*, 8(3):351–357, 2013.

[19] E. R. Miranda and M. Wanderley. *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series).* A-R Editions, Inc., Madison, WI, USA, 2006.

[20] M. Müller. *Information Retrieval for Music and Motion.* Springer Verlag, 2007.

[21] V. Verfaille, O. Quek, and M. M. Wanderley. Sonification of musicians' ancillary gestures. In *ENACTIVE Workshop, Montreal*, May 2006.