HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 13(1), May 2017, 58–81

MUSICAL INSTRUMENTS, BODY MOVEMENT, SPACE, AND MOTION DATA: MUSIC AS AN EMERGENT MULTIMODAL CHOREOGRAPHY

Federico Visi
Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University
UK
and
Institute for Systematic Musicology
University of Hamburg
Germany

Esther Coorevits
Institute for Psychoacoustics and Electronic Music (IPEM)
Ghent University
Belgium

Rodrigo Schramm
Department of Music
Federal University of Rio Grande do Sul
Brazil

Eduardo Reck Miranda
Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University
UK

Abstract: *Music is a complex multimodal medium experienced not only via sounds but also through body movement. Musical instruments can be seen as technological objects coupled with a repertoire of gestures. We present technical and conceptual issues related to the digital representation and mediation of body movement in musical performance. The paper reports on a case study of a musical performance where motion sensor technologies tracked the movements of the musicians while they played their instruments. Motion data were used to control the electronic elements of the piece in real time. It is suggested that computable motion descriptors and machine learning techniques are useful tools for interpreting motion data in a meaningful manner. However, qualitative insights regarding how human body movement is understood and experienced are necessary to inform further development of motion-capture technologies for expressive purposes. Thus, musical performances provide an effective test bed for new modalities of human–computer interaction.*

Keywords: *music, movement, performance, musical instrument, motion sensor, score.*

DOI: http://dx.doi.org/10.17011/ht/urn.201705272518

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION AND BACKGROUND SCENARIO

In the interdisciplinary field of musicological research, the idea that music is a multimodal phenomenon that engages body movement has given rise to a wide range of methodologies for studying musical gestures. Musical gestures are "human body movement that goes along with sounding music … the gestures of those that produce the sounds (the musicians), and the gestures of those that perceive the sounds (the listeners or dancers)" (Jensenius, Wanderley, Godøy, & Leman, 2010, p. 13). On the one hand, the advent of new technologies, such as infrared motion capture, has allowed researchers to observe human movement in detail, extracting precise three-dimensional data and kinematic features of bodily movement. This brought about a corpus of studies where motion analysis is based on the computation of several low-level descriptors—or movement features—that could be linked with musical expression (Godøy & Leman, 2010). For example, acceleration and velocity profiles have been shown to be useful in the study of musical timing (Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2014; Dahl, 2014; Glowinski et al., 2013; Goebl & Palmer, 2009; Luck & Sloboda, 2009). Quantity of motion (QoM) has been related to expressiveness (Thompson, 2012) and has been used to study the dynamic effects of the bass drum on a dancing audience (Van Dyck et al., 2013), while contraction/expansion of the body can be used to estimate expressivity and emotional states (Camurri, Lagerlöf, & Volpe, 2003). More advanced statistical methods, such as functional principal component analysis and physical modeling, have led to midlevel descriptors, including topological gesture analysis (Naveda & Leman, 2010), curvature and shape (Desmet et al., 2012; Maes & Leman, 2013), and commonalities and individualities in performance (Amelynck, Maes, Martens, & Leman, 2014).

On the other hand, gestures in musical performance can be accessed by means of high-level descriptors. Verbal descriptions, subjective experiences, and the musician's intentions play an important role in daily interaction with music. This is the way performers and audiences naturally communicate about music. Leman (2008b) referred to these descriptions as first-person perspectives on music experience, resulting in intention-based symbolic/linguistic expressions. In the analysis of musical performance, this qualitative approach has been explored profoundly in the studies of, among others, Davidson (Davidson, 2007, 2012; Williamon & Davidson, 2002) and King (2006). In studies such as these, musical gestures are accessed by means of verbal descriptors or directly perceivable movements that appear to be expressive. In that sense, the concept of musical gesture can be useful to bridge the gap between mental and subjective experiences of the performer/listener and the directly observable physical world.

Recent studies have made an attempt to close the gap between these two perspectives in musical performance research by applying a performer-informed analysis (Coorevits, Moelants, Östersjö, & Gorton, 2016; Desmet et al., 2012). In trying to understand the relationship between the physical aspects of movement in space and expressive qualities, the study of musical gestures has resulted in new understandings of the relationship between musician and musical instrument as well. Here, the instrument becomes a natural extension of the musician (Nijs, Lesaffre, & Leman, 2013), being part of the body and hence integrated into the mediation process of communicating musical meaning. In the context of musical practice, an instrumentalist's gestures bear great expressive potential. Many new interfaces for musical expression (NIMEs) have been developed in the past years and take advantage of this (Miranda & Wanderley, 2006). The development of human–computer interfaces also exploits the expressive potential of musical gestures to enhance the interaction between the digital and the human environments and to create meaningful applications for musical practice (Camurri & Volpe, 2003). Recently, artistic practice has been increasingly adopted as a complementary research method for the arts and humanities, leading to mixed, interdisciplinary methodologies (Smith & Dean, 2009).

In this article, we aim to explore issues related to the representation and mediation of body movement in musical performance through digital data. We do so by adopting an interdisciplinary approach, which involves the theoretical analysis of implications brought about by (a) representing movement through data, (b) defining computable motion descriptors, and (c) employing a case study of a musical composition for viola, guitar, and motion sensors.

The structure of the paper is as follows. We first review and comment on several technological approaches to analysis, computation, and interpretation of movement data obtained from different devices. We then propose some techniques useful for extracting meaningful information from data collected via wearable motion sensors. Successively, we describe a musical performance where motion sensors have been employed alongside musical instruments. The analysis of this case study then leads to a discussion on the interpretation of human expressive movement through musical experience and digitized motion data. Subsequently, we suggest that musical practice can aid the development of new techniques for the interpretation and understanding of movement data for expressive purposes. This can result in valid contributions to the development of motion-based technologies for human-computer interaction beyond music applications.

CAPTURING, STORING, AND MEDIATED MOVEMENT

Human movement can be digitally captured and stored via different means for purposes of analysis, description, and notation. In the context of musicological studies, movement has been recorded throughout the years using visual media, such as photography (Ortmann, 1929) and video (Davidson, 1993). More recently, motion capture has become widely adopted as the medium of choice for quantitative studies of human motion. Even though new technologies are emerging, marker-based optical motion capture is still regarded as the most reliable solution for precise, high-speed tracking. Data obtained from these systems is usually in the form of three-dimensional vectors referring to a global coordinate system. Each sample in the data returns three-dimensional information regarding the position of a point (marker) in space in relation to the origin of the Cartesian axes. The origin is defined during the calibration procedure and is usually set in an arbitrary place on the floor within the capture area.

As Salazar Sutil pointed out, the term motion capture (sometimes shortened to MoCap) indicates not only a technological setup but also a "technologized language of movement, involving the formalized description of movement coordinates and movement data for its subsequent computational analysis and ... processing" (2015, p. 198). Compared to photography and film, MoCap is definitely a younger medium. This has obvious technological implications, as MoCap technologies are still being developed and only recently have become more widely accessible to researchers and practitioners. However, the nature of body movement itself makes its mediation somehow still conceptually challenging. Salazar Sutil noted that the conceptualization of corporeal movement often is optically biased, whereas sensations that are unrelated to sight are often neglected. The ubiquity of visual record is certainly a factor in this process. Still, movement cannot be entirely represented and therefore fully understood exclusively by means of visual media. In fact, interpreting human movement objectively as a displacement of body parts in a three-dimensional space would result in a limited interpretation. Merleau-Ponty famously points this out by giving the example of typing:

> The subject knows where the letters are on the typewriter as we know where one of our limbs is, through a knowledge bred of familiarity which does not give us a position in objective space. The movement of her fingers is not presented to the typist as a path through space which can be described, but merely as a certain adjustment of motility, physiognomically distinguishable from any other. (Merleau-Ponty, 1945/2002, p. 166)

This is possibly one of the reasons why the use of absolute position in a Cartesian coordinate system imposes some constraints and challenges to high-level analysis of motion data and its use for expressive applications.

In previous works, we used MoCap to carry out experiments aimed at analyzing relationships between body movements and other musical features in instrumental musical performance (Visi, Coorevits, Miranda, & Leman, 2014; Visi, Coorevits, Schramm, & Miranda, 2016). For real-time musical applications, we preferred the use of various wearable sensors because they are easier to transport and use in performance situations. Optical MoCap, on the other hand, is definitely more demanding in terms of portability and setup time. As we will show more in detail, the raw data returned by wearable sensors are intrinsically different from that of MoCap, and this presents some implications for how the data are eventually interpreted and applied. Understanding how to extract meaningful descriptors from such sensors is useful beyond the domain of musical practice because similar technologies are becoming ubiquitous, employed in everyday objects such as mobile devices and game controllers.

Previous research (e.g., Freedman & Grand, 1977; McNeill, 1996) has pointed out that upper body movements are of particular interest when observing expressive behavior. In instrumental musical performance, the upper limbs have a central role; they typically are involved in the main sound-producing gestures. Moreover, in most cases, hands and arms are the primary points of contact between the body of the performer and the instrument. Therefore, in the studies described in this article, we placed the sensor bands on the forearms of the performers. However, as we will show in the following sections, processing data from the inertial measurement units (IMUs) using motion descriptors and machine learning models allowed us to obtain information related to full body movements, which can be used to extract expressive movement features.

In earlier tests (Visi, Schramm, & Miranda, 2014a, 2014b), we made use of fingerless gloves for decreased interference during the musical instrument manipulation. We progressively moved away from gloves in order to obtain an even less obtrusive configuration. We first placed the sensor on the wrists and eventually moved further away from the hands of the performer, onto the upper forearm. Doing so did not reduce the amount of information about hand movements we were able to retrieve. On the contrary, by using multimodal sensing and exploiting the constraints posed by the structure of the limbs and the interdependence of its parts, we were able to estimate various measures describing the movement of both hands.

Initially, we mostly employed IMUs; later we sought to include a form of muscle sensing. This was done in order to address and estimate body movement components beyond those strictly related to displacement in space, such as proprioception and effort qualities.

INERTIAL MEASUREMENT UNITS (IMUs)

An IMU is a device that incorporates accelerometers and gyroscopes. When these devices are paired with magnetometers, the resulting arrays are known as magnetic, angular rate, and gravity (MARG) sensors as well. These sensor arrays allow for the tracking of acceleration, rotational velocity, and orientation of the object they are attached to relative to the earth's magnetic field. They are used extensively in aviation, robotics, and human–computer interaction (HCI), and their increasing affordability and small size have made them a very common feature of mobile and wearable devices and other consumer electronics (see Figure 1). Recently, 9 degrees of freedom (9DoF) sensors have become the most widely used type of IMU/MARG. Featuring three types of tri-axis sensors (hence the name), they enable estimating various motion features, including optimized three-dimensional (3D) orientation obtained by fusing together the data from the three types of sensors.

Whereas the raw data obtained using marker-based optical motion capture consist of samples of position based on a 3D Cartesian coordinate system,[1] the data returned by 9DoF sensors are usually in the form of three 3D vectors, each one expressing acceleration, rotational velocity, and orientation, respectively. The sensor band we used more recently[2] returns acceleration in units of g, rotational velocity in degrees per second, and orientation angles in radians. Orientation also is estimated using a quaternion representation, which—unlike Euler angles—is not subject to problematic singularities such as gimbal lock (Brunner, Lauffenburger, Changey, & Basset, 2015). In addition, the sensor band returns 8-channel electromyogram (EMG) data, which we used to compute descriptors of muscular effort and to estimate the movements of wrists and fingers.

Calculating absolute position from IMU data in real time is technically very difficult if not outright unfeasible, as the operation would require double integration of acceleration data. This would result in a considerable level of residual error because drift would accumulate quadratically.

Figure 1. Some of the wearable devices used in prior research. Clockwise from the bottom-left corner: Myo armbands, Sense/Stage controllers with wristbands, Axivity WAX9 with silicone wristband, Adafruit 9-DOF IMU Breakout, FreeIMU v0.4.

Moreover, it would also be relatively expensive in terms of computation. Madgwick et al. designed computationally efficient algorithms for compensating the residual error (Madgwick, Harrison, & Vaidyanathan, 2011). These can be used for estimating position from IMU data recorded in situations where specific constraints could be exploited, such as gait analysis and cyclic motion.

The data obtained from IMUs are therefore morphologically very different from positional data returned by optical MoCap. The differences in the way movement is tracked and represented by the two technologies has implications on how movement data are eventually interpreted and used, particularly in the context of expressive movement tracking. High-level movement descriptors often are used to extract features from the raw motion data that can help describing the *meaning* that the movements of the subject convey. This is no trivial task, and various interdisciplinary approaches have been adopted over the past 2 decades. In the following section we look at several motion descriptors most widely used with positional data and discuss how they can be adapted for use with IMU data.

MOVEMENT DESCRIPTORS AND WEARABLE SENSORS: UNDERSTANDING DIGITIZED MOVEMENT QUALITIES

Computable descriptors of human motion are used across several disciplinary fields for various applications, ranging from kinesiology and gait analysis to HCI and gaming. Even though human motion data analysis has become an increasingly active field, there is still little consensus regarding which descriptors and methodologies yield meaningful representations of human body motion.

The MoCap Toolbox (Burger & Toiviainen, 2013) provides a wide range of MATLAB scripts for offline kinematic analysis and visualization. Alternatively, expressive feature extraction and real-time interaction are prominent features of the Eyesweb platform (Camurri et al., 2007).

Going beyond traditional low-level kinematic features has proven challenging, especially when dealing with expressiveness, emotions, affective states, and meaning. Larboulette and Gibet (2015) recently attempted a thorough review of computable descriptors of human motion. This was indeed a useful endeavor; however it showed continued segmentation and that many procedures are either ill-defined or redundant (i.e., similar concepts appear in other research literature under different names).

Most of the descriptors we mention below were conceived using positional data. However, the principles behind their design are nonetheless useful for describing certain movement qualities; therefore, we attempted to adapt them to the data obtained from the IMU/MARG sensors. A series of objects were implemented using Max,[3] which was chosen over other programming environments because it allowed for rapid prototyping and testing of algorithms for real-time interaction and for easily integrating them with other music applications.

Fluidity and Jerkiness

In kinematic analysis, “jerk” is the name given to the third-order derivative of movement position, namely the variation of acceleration over time. The definition of “jerk index” as the magnitude of the jerk averaged over the entire movement (Flash & Hogan, 1985) was used by Pollick, Paterson, Bruderlin, and Sanford (2001) alongside other descriptors to correlate arm movement to basic affective states. This relates to the fluidity or smoothness of a movement—as fluid movement tends to have an even level of velocity and, therefore, low values of higher-order derivatives—that can be used to detect emotionally relevant information in movement data (Glowinski et al., 2011). In fact, roughly speaking, jerkiness could be seen as the inverse of fluidity. Piana, Staglianò, Odone, and Camurri (2016) defined the fluidity index as a local kinematic feature equal to $\frac{1}{\int (j_i+1)dt}$, where j_i is the jerk of the joint[4] i. This means that higher values of jerk correspond to lower fluidity.

To estimate jerkiness using 9DoF sensor data instead of positional data, we averaged the derivatives of longitudinal acceleration returned by the accelerometer $(\dot{ax}, \dot{ay}, \dot{az})$ into a single jerk index. The resulting value was then combined with the averaged second-order derivatives of the angular velocity returned by the gyroscope $(\ddot{gx}, \ddot{gy}, \ddot{gz})$ and then summed over a time window of length N samples:

$$IMUJerkiness(t) = \sum_{k=0}^{N-1} \alpha_1 \frac{|\dot{ax}_{t-k}| + |\dot{ay}_{t-k}| + |\dot{az}_{t-k}|}{3} + \alpha_2 \frac{|\ddot{gx}_{t-k}| + |\ddot{gy}_{t-k}| + |\ddot{gz}_{t-k}|}{3}.$$

Coefficients $\alpha_1$and α_2 are weights that balance the data magnitudes obtained from the accelerometer and gyroscope sensors. It is worth mentioning that, in real-world implementations, derivatives are very sensitive to signal noise. Therefore, sensor data may require low-pass filtering before jerkiness can be computed.

From the conceptual framework of Laban effort elements/qualities (Laban & Lawrence, 1947), jerkiness (and its counterpart fluidity) is a useful descriptor that can aid the computational analysis of expressive movements. Laban (as cited in Hackney, 2002) defined four basic effort factors (flow, weight, time, and space); each factor is a continuum between polarities described by effort element/qualities. In particular, flow is related to the continuity and control of the movement. Its polar qualities (free flow and bound flow) have been previously associated with aspects of fluidity in movement. A movement reflecting the free flow effort quality is characterized as fluid, liquid, and outpouring. On the other hand, the bound flow quality indicates containment, restraint, and control. In addition to flow, jerkiness can be related also to the time effort elements. A movement characterized by the sustained effort qualities is expected to have a low level of jerkiness. On the other hand, a movement with the sudden effort qualities (i.e., urgent, quick, staccato) will have most likely a higher rate of change in acceleration and therefore higher levels of jerkiness.

Jerkiness and fluidity, then, can contribute to the analysis and recognition of expressive movement qualities, particularly in multimodal frameworks that involve multiple descriptors and sensing modalities (Camurri & Volpe, 2011; Caramiaux, Donnarumma, & Tanaka, 2015). However, it is important to emphasize that Laban effort elements are qualitative “inner attitudes” of a person moving towards the effort factor. Using computable descriptors should not be seen as an attempt to quantitatively measure the effort qualities but rather as a means to aid in the design of computational models capable of discerning and recognizing different expressive movement behaviors.

Quantity of Motion and Overall Motion Energy

Fenza et al. defined QoM as the sum of the Euclidean distances between successive points in a time window (Fenza, Mion, Canazza, & Rodà, 2005) and Glowinski et al. (2011) included a similar measure in their expressive feature set, denoted as overall motion energy. To compute an analogous feature using 9DoF sensor data, we aggregated the magnitude of the variations of the norm of the orientation quaternion ($\parallel q \parallel$) and of the average acceleration over the three axes (a). The values for each frame were once again summed over a time window of length N samples:

$$IMUQoM(t) = \sum_{k=0}^{N-1} \beta_1 \mid \parallel q_{t-k} \parallel - \parallel q_{t-k-1} \parallel \mid + \beta_2 |a_{t-k} - a_{t-k-1}|.$$

Similarly to the previous equation, β_1 and β_2 were weights to balance individual contributions from distinct sensors.

Contraction/Expansion and Symmetry

Contraction and expansion of the body can be computed in different ways. They can be achieved, for example, by calculating the area of bounding shapes (Glowinski et al., 2011), using the contraction index (Fenza et al., 2005) or measuring the volume of a convex hull that encloses the body (Hachimura, Takashina, & Yoshimura, 2005).

When wearing two 9DoF sensors on the forearm, it is possible to project the orientation values over a hypothetical 2D plane in front of the subject and thus obtain approximate coordinates of the points in the plane the arms are pointing to. First, the yaw values for both arms have to be centered while the subject is pointing both arms forward. Then, given θyaw and θpitch as the yaw and pitch angles, respectively (expressed in radians), the coordinates for each point in the plane can be calculated as follows:

$$(x, y) = \left(\frac{x_{max}}{2} + \frac{x_{max}\ \theta_{yaw}}{2\pi}\ ,\ \frac{y_{max}}{2} + \frac{y_{max}\ 2\theta_{pitch}}{2\pi}\right).$$

By calculating the Euclidean distance between the two points, it is possible to estimate whether the arms are pointing in opposite directions. When arms are spread wide, the distance between the two points will be at its maximum. In this way, it is possible to have a value that depends on whether arms are wide open or are resting close to the body. This is can be used as an expressive feature, even though it is not as precise as the contraction indexes obtained using an optical motion capture because the values are based on the orientation of the arms and not on their actual position. By comparing the coordinates, we also can see if there is horizontal or vertical symmetry between the arms, which is another useful postural feature that has been previously used for the analysis of expressive movements (Camurri & Volpe, 2011).

Periodicity and Rhythmic Qualities: Periodic Quantity of Motion

Previous techniques that focused on expressive motion analysis have intensively used the QoM estimation, especially to segment gestures by detecting the resulting motion bell curves generated by a subsequent low-pass filtering process (Fenza et al., 2005). QoM is also useful for detecting the presence of body movements and other related activities during a live performance. It is easy and quick to implement, and it is frequently used as an input for more complex algorithms, such as gesture classifiers based on machine learning.

While the techniques mentioned above could be described as spatial features of the movement, or spatiotemporal in the case of QoM, periodicity is a purely temporal quality. Despite the good estimation of body movement given by QoM, it is not suitable for describing multiple periodic gestures, which are usually associated with the musical rhythmic content. In fact, there is often sensorimotor synchronization between the rhythmic structure of the piece and the periodic motion of the body (Repp & Su, 2013). Periodic quantity of motion (PQoM; Visi, Schramm, & Miranda, 2014a) was proposed as a complementary way to measure periodicity in the movement in relation to the musical rhythm.

The first PQoM implementation was designed to extract periodic motion from the data tracked by optical motion capture systems. The PQoM estimates are obtained from the decomposition of the signal from the motion capture tracker into frequency components by using filter banks (Müller, 2007). The PQoM function uses three main parameters as input: window size, beats per minute (bpm), and the center frequency of each band-pass filter. These parameters are intrinsically associated with the musical content. The window size parameter is used to define the amount of time that should be integrated to compute the final QoM. Once the music tempo is defined in bpm, each center frequency can be expressed by note values (i.e., half note, quarter note, eighth note, etc.) reflecting the rhythmic structure of the music. The PQoM function converts these note values into the center frequencies of each band-pass filter. Each center frequency fc is used to compute the transfer function coeffcients of a respective 4th-order band-pass digital Butterworth filter. Then, the input signal is filtered by a zero-phase forward and backward digital IIR filtering algorithm (non-real-time implementation). Finally, similar to QoM, the filtered values for each sampled frame are summed over a time window of length N (window size) samples. For optical motion capture data, each tracked point is normalized using the origin [0, 0, 0] of the coordinate system as reference,[5] and a weight vector w combines distinct points from the body skeleton. It is worth noting that it is possible to ignore specific markers by setting the related coefficient in w to zero. Thus, the discrete PQoM can be expressed by

$$PQoM[t, f_{c_k}] = \sum_{n=t-N}^{t} H_{f_{c_k}} \{(w(|x[n] - x[n-1]|))\},$$

where the x vector is the Euclidean norm from the skeleton markers positions, and $H_{f_{c_k}}$ is the k^{th} band-pass filter operator. The first software implementation of the PQoM was made in MATLAB and can be downloaded as an extension for the MoCap Toolbox[6] (Burger & Toiviainen, 2013). The same approach was adopted to estimate the PQoM from data from other sensors. However, a specific w vector of coeffcients must be defined regarding each sensor type, and an appropriate norm for the x vector must be chosen. For instance, we used a similar approach as in Equation 2, in which our PQoM implementation replaced x with the weighted sum of the norm of the orientation quaternion ($\| q \|$) with the average acceleration over the three axes.

Also, for the use in real-time applications, the PQoM can be implemented using a causal filter (instead of zero-phase forward and backward digital IIR filter). Figure 2 presents the results of the PQoM measure applied to a recorded movement using a motion capture system and an armband device with a built-in IMU/MARG sensor. Figure 2a shows the locations of these sensors (passive markers and armband device). In this example, the musician performed a periodic movement with his arm, syncing it with the half note duration (from 1 to 3 s) and with the quarter note duration (from 5 to 7.5 s). Figure 2b illustrates the respective tracked x vectors: motion capture marker (green line, top), acceleration (blue line, center) and orientation (red line, bottom).

The bottom part of this figure shows the PQoM measures extracted from the tracked data. The amount of gesture periodicity over time is indicated by the yellow and light blue regions (higher values) on each matrix. The PQoM estimates from the motion capture marker (Figure 2c) are comparable with the respective PQoM measures extracted from the acceleration (Figure 2d) and orientation (Figure 2e) data. This means that PQoM is robust enough to be measured with different sensors and data types. This flexibility allows for the substitution of very precise and expensive motion capture systems with less accurate—but more affordable—devices, such as IMUs.

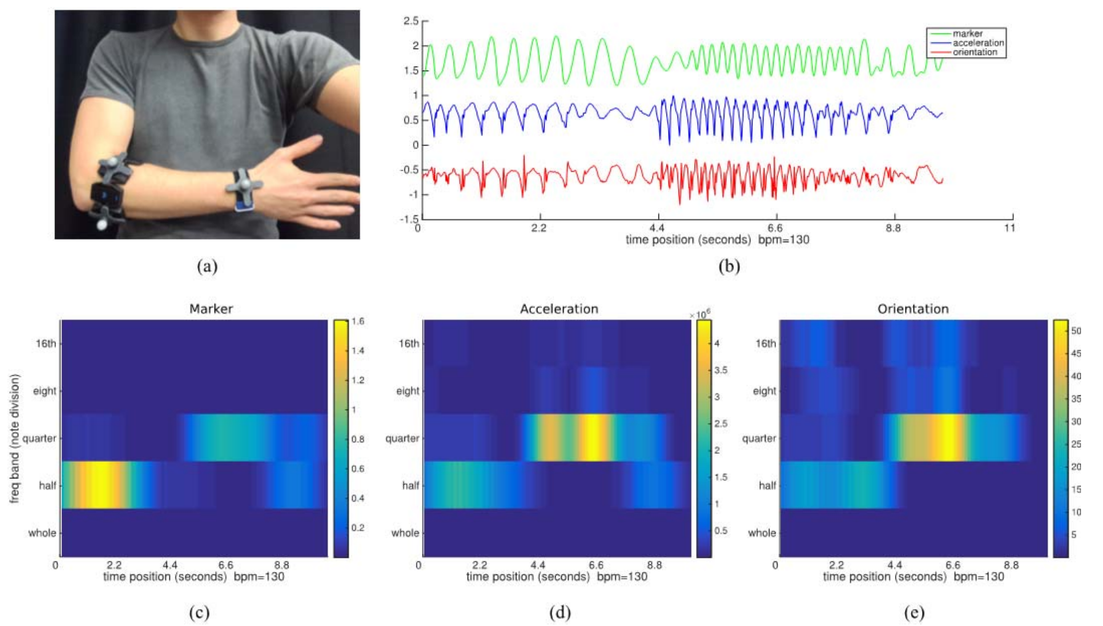

Figure 2. Periodic Quantity of Motion (PQoM) computed from motion capture and inertial measurement unit data. (a) locations of passive markers and armband device. (b) tracking data from sensors (rescaled for better visualization). (c) PQoM from motion capture marker. (d) PQoM from acceleration. (e) PQoM from orientation.

MACHINE LEARNING: MAPPING POSTURAL AND SONIC TOPOLOGIES

Motion descriptors are useful for extracting meaningful features from the raw data, as well as allow for aggregating information relative to all the axes. This helps to move away from a low-level movement representation constrained by the Cartesian coordinate system and to obtain motion data that are less dependent on it. A system based on orthogonal axes is indeed a convenient way to digitize movement. However, a meaningful conceptualization that helps in interpreting the expressivity that body movement conveys may be hindered if subordinated to a highly disciplined method of quantitative representation. In his article about topology and data, Carlsson argued that "coordinates ... are not natural in any sense,... [and] therefore we should not restrict ourselves to studying properties of the data which depend on any particular choice of coordinates" (Carlsson, 2009, p. 256). Moreover, in describing the characteristics of topological methods, he stated that, to obtain knowledge about the data, qualitative information is needed, and this must be established before proceeding with quantitative analysis. Researchers use topology to study intrinsic geometric properties of objects that do not depend on a chosen set of coordinates and this process also has been employed in the analysis of dance patterns (Naveda & Leman, 2010). This approach provides very useful notions for interpreting movement data generated by musical performance gestures. In fact, such body movements are bound to multimodal expressive features, which are inherently qualitative.

To put these concepts into practice, we used machine learning algorithms to define interaction models based on the various postures a musician may adopt during a performance. This was done by asking performers to play freely while wearing two sensor armbands. A small number of postures (4–5) were then defined. This was done by observing recurrent idiosyncrasies and peculiarities of the performance and discussing the qualities of the movements with the musicians themselves to better understand how certain movements relate to each respective instrumental techniques and with the musical features of the pieces performed.[7] Sensor data were then sampled repeatedly during each pose in order to train a support vector machine (SVM) classifier. This was implemented using the ml.lib library (Bullock & Momeni, 2015) for Max, which is itself based on the Gesture Recognition Toolkit by Gillian & Paradiso (2014). Every posture was then associated with a set of parameters of a digital sound processing engine. During the performance, the machine learning classifier compared the incoming sensor data stream with the recorded examples, returning the values for the probability (or likelihood) that the current posture of the musician matched one of the defined classes. The values of the probabilities then were used to interpolate between the parameter sets of the sound engine that was used to process the sound of the instrument in real time or to synthesize electronic sounds.

This practical approach resonated with the aforementioned notions of topology because the incoming data were not analyzed quantitatively but rather evaluated in terms of proximity/distance from the predefined postures.[8] From this perspective, the sampled postures themselves were topologies determined in relation to qualitative aspects of the movement of that particular performer, thus avoiding dependency from an abstract, artificial coordinate system. The system was instead defined by the idiosyncrasies of the performer.

This approach offers some practical advantages compared to more traditional sensor-to-sound parameter mapping approaches. First, incoming sensor data do not need to be rescaled to the range of the sound parameters they are mapped to. Moreover, the quantitative values of the sensor data can be ignored because the classifier probabilities are used to interpolate multiple sound parameters. This is another advantage because complex mappings can be defined easily and parameter modulation is independent of any coordinate system. Instead, the system quickly adapts to different users, and this is desirable considering the substantially different movements required for playing different instruments and the significant degree of idiosyncrasy that characterizes musical performance.

In the past few years, machine learning techniques increasingly have been employed for interactive computer musical performance. Notable approaches include Wekinator (Fiebrink, Trueman, & Cook, 2009), Gesture Variation Follower (Caramiaux, Bevilacqua, & Tanaka, 2013) and mapping by demonstration (Françoise, Schnell, Borghesi, & Bevilacqua, 2014).

In our early tests, orientation data and an aggregate EMG descriptor for both arms were used as inputs to train the machine learning models. Orientation was chosen because it is not an inertial measurement; therefore, it can be used to describe postures. In addition, EMG data allowed us to consider other characteristics of movement. As pointed out by Salazar Sutil (2015), the perception of body movement involves sensations that go beyond displacement in space, such as interoception and proprioception. Moreover, in their extensive work on the analysis of expressive movement, Camurri and Volpe (2011) defined gesture as a multimodal entity, citing Laban's theory of effort (Laban & Lawrence, 1947) as a central source of concepts for understanding expressive movement.

CASE STUDY: *KINESLIMINA*

Kineslimina is a musical composition for viola, electric guitar, motions sensors, and live electronics that explores the use of the musicians' instrumental movements as an expressive medium. Such gestures merge with the other musical features and become an integral part of the score. While playing their instruments, the musicians wear an armband fitted with IMUs, which tracks their movements and sends the motion data to a computer. The computer then processes the movement data and sound, responding with a wide range of dynamics: from subtle timbral alterations that follow the movements of the bow during string changes to deeper resonances when more overt gestures are performed by the musicians. The title is a portmanteau word composed of *kinespheres* and *limina*. The concept of kinesphere was defined by Laban (1966, p. 10) as "the sphere around the body whose periphery can be reached by easily extended limbs." Thus, the kinesphere is a personal space, and how an individual relates and pays attention to it contributes to its delineation. Limina is the plural form of *limen*, which is a threshold or margin. The piece aims at pushing the boundaries of the personal spaces that surround the musicians during the performance. Throughout the performance, the sound of the instruments is altered, and synthesized sounds are engaged by exceeding the usual extent of instrumental movements. The score of the piece can be considered a script through which a multimodal choreography emerges as the product of learned body schema, altered by the influence of and the reactions to an interactive system. In the ritualized context of musical performance, a nonconventional technology (the sensors) interferes with conventional ones (the instruments), reconfiguring the relationships between the score, the performers, and their tools.

Parviainen, Tuuri, Pirhonen, Turunen, & Keskinen (2013) proposed an approach to interaction design that considers choreography as the holistic, experiential continuum of human movement resulting from the interaction with artifacts. From this perspective, musical instruments, sensors, and movement/sound mappings can be seen as carriers of a set of prechoreographies. The design of these objects (whether material or not, as in the case of software) and the environment where the interaction takes place prechoreographs the performance of the piece. In other words, all the movement opportunities that these objects afford form the basis for the actual choreography that emerges as the score is enacted. Thus, a pre-existent, overarching design influences the movements within the kinesphere. Eventually, the movements that shape the performance exceed what the individual kinespheres can capture. The relations between the two musicians and between the musicians and the audience and the dynamics that arise from these connections are what Parviainen, Tuuri, Pirhonen, Turunen et al. (2013, p. 2) called "local-level movements."

As noted by Wilson, a traditional musical instrument is not merely a piece of technology: experiential relationships to it are shaped by "the way it is 'meant' to be played, the canonic tradition that stands behind it as repertoire, and the normative expressive gestures that are 'input' by the player and 'output' sonically by the instrument" (Wilson, 2013, p. 426). Bodily relationships with these cultural artifacts are mediated historically and become part of a shared knowledge. Introducing motion sensor technology into this picture adds another layer of complexity, tightly woven to the already established gesture–sound relationships. Figure 3 shows an example of how these different aspects of the piece are interrelated. Towards the end of the piece, the score requires the viola player to repeat an arpeggiated pattern with increasing dynamics, until a chord played by a synthesizer is heard. The part entails repeated bow strokes, and the movement pattern is captured by the sensors placed on the right wrist (Figure 3b). The peaks in the acceleration data control a granular synthesis engine that samples and alters the timbre of the instrument at each peak. At the same time, QoM is computed (red line, Figure 3c)

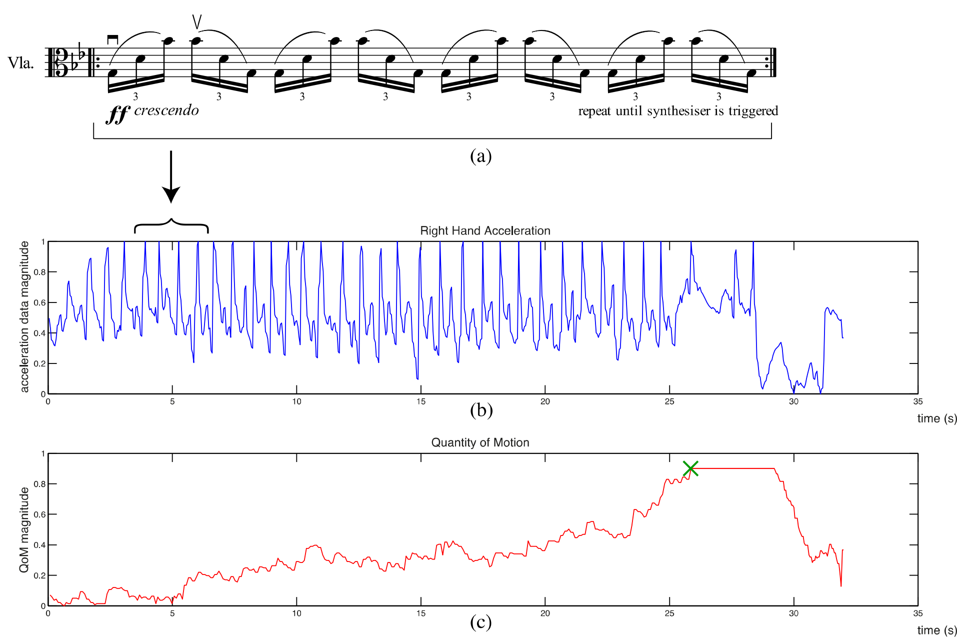

Figure 3. Excerpt of the viola part of *Kineslimina* and corresponding kinematic features of the right hand of the viola player: (a) scored pattern, (b) right hand acceleration of the viola player, and (c) the Quantity of Motion. The green cross indicates when the QoM threshold is crossed and the synthesized part that closes the composition is played back.

and the increase of motion activity introduces other electronic parts, until the QoM data reach a predefined threshold (green cross, Figure 3c) at which point the closing synthesizer chord is triggered and the musicians can move onto the closing notes of the piece. The number of repetitions required to reach this point depends on the movements of the performer, which may vary according to how different musicians interpret the score. The score engenders the instrumental movements required to perform the piece; the movement data alter the sound of the instrument and also impact the structure of the score itself (i.e., the number of repetitions required). This closes a feedback loop in which every part mutually influences the other. The body of the performer is the medium, the locus, where this dynamic entanglement takes place.

This interdependency also affects how motion data are mapped to sound parameters. Mapping sensor parameters to sounds has long been debated in the HCI and NIME research communities (Hunt, Wanderley, & Paradis, 2003). Leman noted that freedom of mapping that characterizes digital musical interfaces "may disturb the sense of contact and of non-mediation" (Leman, 2008a, p. 164). Drawing from an established vocabulary of gesture of a traditional musical instrument and exploiting the constraints that instrumental techniques pose on the body can result in an advantage for obtaining meaningful interactions for expressive musical performance. This approach takes advantage of the ecology of musical instruments (Clarke, 2005) in order to obtain expressive transparency in gesture-sound mapping.

Kineslimina premiered at the Gala Concert of CMMR 2015 (the 11th International Symposium on Computer Music Multidisciplinary Research;[9] see Figure 4) and was performed later that same year at MuSA 2015 (the Sixth International Symposium on Music/Sonic Art,[10]) held at the Institut für Musikwissenschaft und Musikinformatik in Karlsruhe, Germany. From the perspective of the performers,[11] the piece reconfigured the relationship between the musicians and their instruments, extending expressive possibilities through their instrumental movements tracked by the sensors. However, this also required the performers to learn new skills and embed them within their existing instrumental techniques. This process became evident during the rehearsals. The performers experienced an increased awareness of the fundamental body schema of their instrument playing, as subtle movements created new sonic results via the motion sensors. This made them pay renewed attention to movements they learned in the early days of their formation as musicians, essential parts of the vocabulary of gestures of their respective instruments. While the musicians learned and became more familiar with the sensors, the system itself continually adapted and adjusted to accommodate the needs of the performers and to better follow their performance styles. As the rehearsals continued, relationships between body movements, instrumental gestures, and sensor data became renegotiated. This did not involve mere parameter adjustments and technical improvements to the sensor system: The process elicited and entailed a careful analysis of the relationship between movements, sound, and score from the privileged perspective of the performers themselves, thus resulting in a useful contribution to research from a practice-led perspective (Sullivan, 2009).

The performers then progressively got to know how the mappings of movement features to sound worked, and how they could explore this unconventional technology in a meaningful way. More than just "sonifying" the movements made by the performers, the sensor system induced the musicians to reconsider their relationships to the performance space. Such space—where the relationships among the players, instruments and audience are located—encompasses a set of conventions and cultural practices that are established and embodied in the performer. This can be compared to what Ervin Goffman (1974) referred to as "the frame," which is the

Figure 4. *Kineslimina* performance during the Gala Concert of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR), 16 June, 2015, Plymouth, UK.

perceptual mechanism that indicates the nature and purpose of a behavior and how it is to be interpreted. It is a tool for understanding the implicit agreement between performer and audience on the symbolic status of the performance. From the perspective of a musician, this frame consists of the historically established relationships that transpire between players and instruments. Performance techniques and experience with instruments and instrumental music are habitualized through historical practice, conventions, and education. These relationships are thus part of the embodied knowledge of the performer. In *Kineslimina*, performing with reconfigured instrument/body/space relationships has made the musicians more aware of other qualities of their movements and their kinespheres.

> Laban identified space, weight, time and flow as motion factors toward which performers of movement can have different attitudes depending on temperament, situation, environment and many other variables. The attitudes toward the motion factors he called... Effort. ... Choices are continuously made by all people in motion, consciously or unconsciously, to determine what combinations of these Effort elements will best serve the purposes of their intentness or modify their behavior.... Whatever the action in which the effort combinations appear, the whole biological/psychological system is involved. (Bartenieff & Lewis, 1980, p. 51)

Intentionality is a key aspect in the study of musical gestures; the fact that they are goal-directed actions is an essential quality for the understanding of their expressive qualities (Godøy & Leman, 2010). The effort qualities of a movement are very much the result of this intentionality and they play an important role in the perception and understanding of body movement. During the *Kineslimina* performances, effort qualities and intentionality appeared amplified by the presence of the motion sensors and their effect on the conventional performance gestures. Moreover, as some audience members commented after the performance, this interplay between the performers and their role in the intersubjective space was transparent through the augmentation of the performers' musical intentionality.

DISCUSSION AND FUTURE WORK

We presented techniques for interpreting motion data and discussed the implications that arise when employing motion sensors in conjunction with traditional instruments in musical practice. Within this context, it is clear that body movements go well beyond simply being activators of technological objects, whether these objects are motion sensors or musical instruments. Bodily movement is considered a key element in forming embodied musical meaning (Leman, 2010). However, its role as an important cog of the engine that engenders signification and cognition is obviously not limited to the musical context.

Technological objects have the capacity to entail gestures and store their potential meaning. As we have observed in our research, a musical instrument can be seen as a receptacle of gestures, of kinemes that—through a performance—give rise to a multimodal choreography. From a wider perspective, we could say that objects extend human cognition: they are cosubstantial, continuous, and coextensive elements of minds in action (Malafouris, 2013). Moreover, Wilson stressed the importance of the relationship between instrument technology and the instrumentalist's pedagogy: "Technology—what the instrument *is*—is inherently entangled with pedagogy, the historically established relationships found between instrument and instrumentalist" (Wilson, 2013, p. 430; italics in original). In the case study we presented, this "inherent entanglement" encompasses also the score of the piece, which elicited the instrumental movements necessary to execute its parts while, at the same time, being affected by them through the use of motion descriptors. Within this layered process of signification—situated in a cultural ecology and shaped by shared knowledge—the body is the medium where everything takes place. This perspective is akin to Merleau-Ponty's (1945/2002) phenomenological approach.

Once recognizing the centrality of the body and its movements in the ways humans make sense of the world, it is clear that—in an increasingly pervasive digital "semiosphere"—being able to digitize movement and interpret motion data become of primary interest. However, movement seems to have properties that exceed the system used to represent it. We have discussed the limitations emerging from representing movements exclusively through visual media, and the ubiquity of visual record is certainly a factor in this process. However, solving this bias is one of the challenges that contemporary researchers and practitioners must address in making progress in the discourse on human movement. The development of different computational techniques to describe the qualities of body motion is a necessary step towards more meaningful interpretations of data generated by human movement. However, it is also vital to consider the constraints posed by rigid methods of representation and move towards approaches that allow for addressing the complex, nonlinear phenomena that characterize expressivity.

Using inexpensive and unobtrusive devices, such as 9DoF IMU/MARG sensors, also may help to move the research beyond laboratories. As researchers, we have seen how ecology plays an important role in the way people make sense of music (Clarke, 2005). Similarly, being able to study movement "in the wild" may have considerable implications, as shown in previous studies (e.g., Woolford, 2014).

IMU/MARG sensors provide data that are morphologically distinct from those obtained through optical motion capture. However, it is possible to obtain analogously meaningful information if the data are correctly interpreted. In this context, using machine learning techniques is not only a quicker method toward obtaining complex interaction models that adapt to different individuals, it also provides a way to study and reflect upon the topological qualities of human movement through applied research and practice. More sophisticated algorithms to interpret motion data can help address its complexity, reclaiming the centrality of the body over a rigid representation of data structures.

Extracting expressive movement features from 9DoF data can lead to many other applications, well beyond the field of musical interaction. The ubiquity of 9DoF sensors—which is a very common feature of recent communication and entertainment devices—brings about a vast number of potential scenarios where the techniques we describe can be implemented. Using higher level descriptors to estimate expressive qualities of body movements is a way towards implementing dynamic HCI designs that handle gestures not only as isolated objects of application but as part of longer experiential chains, with multiple layers of significance. This goes beyond the traditional use-oriented approach and is akin to the use of choreographies for interaction design proposed by Parviainen, Tuuri, Pirhonen, Turunen, et al. (2013) and Pirhonen, Parviainen, and Tuuri (2013). It also parallels the work on affective computing carried out by the researchers at InfoMus/Casa Paganini (Glowinski et al., 2011; Piana, Staglianò, Camurri, & Odone, 2013). Both approaches avoid limiting HCI design to goal-directed actions and adopt a more holistic approach that takes into consideration a wider ecology of human movement. Moreover, 9DoF sensors coupled with sound synthesis techniques have already found applications in the field of rehabilitation of stroke patients (e.g., Bevilacqua et al., 2013).

As suggested by a topological approach (Carlsson, 2009), gaining new higher level knowledge from motion data also requires qualitative insights. To access more complex, structural, and subjective qualities that are considerably difficult to model quantitatively, data need to be interpreted through qualitative methods. Intuitions arising from qualitative approaches can contribute to the understanding of how body schema and kinemes work in generating embodied meaning. This can successively inform more advanced computational models able to identify complex and meaningful qualities of human movement. Practice as research can address the need of qualitative insights in the interpretation of motion data. Particularly, music can be an effective test bed, given its multilayered complexities and rich cultural, multimodal qualities. As other projects have previously shown, music and the performing arts can be effective test beds for new modalities of expressive HCI (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004), and practice-led approaches have yielded technical and conceptual material useful for the development of motion capture technologies (Norman & Blackwell, 2010). Moreover, musicians often are early adopters of new paradigms of interaction that eventually become mainstream (Kirn, 2013). Notable examples are gestural controllers and multitouch technologies, which were adopted by musicians long before they become widespread.

Practice-led approaches are helpful also for carrying out the conceptual work required to make sense of motion data and understand the meanings it can potentially convey. Through their work, some artists seek to affirm the irreducibility of the corporeal presence while simultaneously sublimating it through digital processing (Norman, 2015). This resonates with the rationale behind *Kineslimina*, and this creative tension can lead to new insights into how movement can carry meaning across physical and digital environments. In *Kineslimina*, relating motion data to a musical score has shown how multimodal qualities of music are entangled, mutually affecting each other.

We expect that future work will adopt this mixed methodology in order to address technical and conceptual aspects related to body movement, motion data, and meaning formation. Other machine learning algorithms will be tested in order to map instrumental gestures to sounds synthesized by means of physical modeling. Sound synthesis techniques based on physical modeling allow artists to generate sounds that resemble those of certain musical instrument families. Working with synthesis parameters facilitates the ability to go beyond the timbral ranges and sonic capabilities of the physical instruments while preserving timbral resemblances to the instrument family. This poses interesting conceptual challenges, as the recognition of timbral qualities of musical instruments relies on a shared knowledge. As we discussed above, the relationships between instruments and instrumental movements is also something encoded in a shared gestural vocabulary. The ecology around musical instruments, their timbres, and their instrumental gestures offers a rich conceptual framework for developing meaningful cross-modal mappings between motion data and synthesis parameters. Cross-modal relations between performance movements and the sonic outcomes will also be inspired by the concept of Uncanny Valley, which was previously adopted in a composition that involved tension and relaxation structures in timbrally varied musical phrases generated by physical models (Bessell, 2011).

From the perspective discussed so far, instrumental music is constituted by abstract structures and performance movements in continual interaction, entangled with technological and cultural knowledge. The concept of choreography appropriately describes the process of multimodal signification that emerges from the performance of a musical score. Body schema and kineme are useful conceptual tools to gain a better understanding of how the body is the central medium in human communication. Movement is a modality of knowledge; therefore, being able to interpret it and represent it through technology offers potential in avenues not yet imagined that should certainly be further explored.

In summary, the article has addressed the challenge of extracting meaningful expressive features from motion data in the context of musical performance. The scope of this work is limited to the context of music and to the case study described. However, the interdisciplinary approach we adopted has led to the design of effective solutions that were implemented in the case study and in other works (Visi, 2017). This has highlighted that qualitative insights into how human body movement is understood and experienced are essential for informing the development of motion capture technologies for expressive purposes, as well as to broaden the discourse on music-related body motion.

IMPLICATIONS FOR EMBODIED HCI

In scenarios where computing is becoming ubiquitous, embodied, and considered a fundamental factor for designing interactions with technology (Parviainen, Tuuri, & Pirhonen, 2013), the implications of being able to extract meaningful information from motion data are manifold. Moreover, the study of music-related motion and the analysis of motion descriptors certainly has applications beyond music making. As an example, Bennett, Hinder, and Cater (2016) recently measured periodicity in data obtained from motion sensors applied to rocking chairs in care homes. This was done to help improve the quality of life of residents in dementia care by creating subtle, engaging interactions that support and stimulate memory through music and movement (Bennett et al., 2016). This is but one example of current applications of motion data analysis in conjunction with music. As computing becomes increasingly embedded in the environments people live in, interpreting the data produced by human activity in a meaningful way remains a key issue. This is central for recent trends in computer science, such as affective computing, calm technology, and human-centered machine learning.

ENDNOTES

1. Most marker-based systems also allow capture of 6DoF data (six degrees of freedom, consisting of three-dimensional position and Euler angles) by defining rigid bodies. However, these data are achieved by processing positional data of the single markers grouped into a rigid body.
2. These data were collected with the Myo Gesture Control Armband, produced by Thalmic Labs (https://www.myo.com).
3. Web page of the Max programming environment: https://cycling74.com/products/max/
4. In motion capture terminology, a joint is a point belonging to a three-dimensional representation of a body. Joints are usually defined by positional coordinates.
5. In fact, it is possible to define any arbitrary origin for the coordinate system. [0,0,0] is the default option.
6. The MoCap Toolbox and the PQoM extension can be downloaded from https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mocaptoolbox
7. To view some early tests with different musicians, please go to https://youtu.be/stWI43-EZGA
8. This idea is reinforced by how support vector machines work: "An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on" (Bullock & Momeni, 2015, p. 4).
9. The Web page of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR) is available at http://cmr.soc.plymouth.ac.uk/cmmr2015/index.html
10. The Web page of the Sixth International Symposium on Music/Sonic Art is available at http://zilmusic.com/musa2015/
11. Esther Coorevits: viola, motion sensors, live electronics; Federico Visi: electric guitar, motions sensors, live electronics.

REFERENCES

Amelynck, D., Maes, P.-J., Martens, J. P., & Leman, M. (2014). Expressive body movement responses to music are coherent, consistent, and low dimensional. *IEEE Transactions on Cybernetics*, *44*(12), 2288–2301. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25415938

Bartenieff, I., & Lewis, D. (1980). *Body movement: Coping with the environment*. London, UK: Routledge.

Bennett, P., Hinder, H., & Cater, K. (2016). Rekindling imagination in dementia care with the resonant interface rocking chair. In Proceedings of the 2016 *CHI conference, extended abstracts on human factors in computing systems* (pp. 2020–2026). New York, New York, USA: ACM Press. doi: 10.1145/2851581.2892505

Bessell, D. (2011). *Ophidian and the Uncanny Valley*. In the annual *Proceedings of the International Computer Music Conference* (ICMC; pp. 346–349). San Francisco, CA, International Computer Music Association, Inc.

Bevilacqua, F., Van Zandt-Escobar, A., Schnell, N., Boyer, E. O., Rasamimanana, N., Françoise, J., Hanneton, S., & Rasamimanana, N. (2013, September–October). *Sonification of the coordination of arm movements*.

Paper presented at the International Conference on Multisensory Motor Behavior: Impact of Sound. Hanover, Germany.

Brunner, T., Lauffenburger, J.-P., Changey, S., & Basset, M. (2015). Magnetometer-augmented IMU simulator: In-depth elaboration. *Sensors*, *15*(3), 5293–5310. doi: 10.3390/s150305293

Bullock, J., & Momeni, A. (2015). ml.lib: Robust, cross-platform, open-source machine learning for Max and Pure Data. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 265–270). Retrieved from http://www.nime.org/proceedings/2015/nime2015_201.pdf

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., & Toiviainen, P. (2014). Hunting for the beat in the body: On period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, *8*(November), 1–16. doi: 10.3389/fnhum.2014.00903

Burger, B., & Toiviainen, P. (2013). MoCap toolbox: A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the 10th Sound and Music Computing Conference* (pp. 172–178). Stockholm, Sweden: KTH Royal Institute of Technology.

Camurri, A., Coletta, P., Demurtas, M., Peri, M., Ricci, A., Sagoleo, R., Simonetti, M., Varni, Volpe, G. (2007). A platform for real-time multimodal processing. In *Proceedings International Conference Sound and Music Computing* (pp. 11–13). Retrieved from http://www.smc-conference.org/smc07/SMC07 Proceedings/SMC07 Paper 59.pdf

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, *59* (1–2), 213–225. doi: 10.1016/S1071-5819(03)00050-8

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human–computer interaction* (pp. 20–39). Berlin, Germany: Springer. doi: 10.1007/978-3-540-24598-8_3

Camurri, A., & Volpe, G. (Eds.). (2003). *Gesture-based communication in human–computer interaction*. Berlin, Germany: Springer. doi: 10.1007/b95740

Camurri, A., & Volpe, G. (2011). Multimodal analysis of expressive gesture in music performance. In J. Solis & K. Ng (Eds.), *Musical robots and interactive multimodal systems* (pp. 47–66). Berlin, Germany: Springer. doi: 10.1007/978-3-642-22291-7_4

Caramiaux, B., Bevilacqua, F., & Tanaka, A. (2013). Beyond recognition: Using gesture variation for continuous interaction. In *CHI '13 extended abstracts on human factors in computing systems* (p. 2109). New York, NY, USA: ACM Press. doi: 10.1145/2468356.2468730

Caramiaux, B., Donnarumma, M., & Tanaka, A. (2015). Understanding gesture expressivity through muscle sensing. *ACM Transactions on Computer–Human Interaction*, *21*(6), 1–26. doi: 10.1145/2687922

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, *46*(2), 255–308. doi: 10.1090/S0273-0979-09-01249-X

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York, NY, USA: Oxford University Press. doi: 10.1093/acprof:oso/9780195151947.001.0001

Coorevits, E., Moelants, D., Östersjö, S., Gorton, D., & Leman, M. (2016). Decomposing a composition: On the multi-layered analysis of expressive music performance. In R. Kronland-Martinet, M. Aramaki, & S. Ystad (Eds.), *Music, mind, and embodiment: CMMR 2015. Lecture Notes in Computer Science* (Vol. 9617; pp. 167–189). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-46282-0_11

Dahl, L. (2014). Triggering sounds from discrete air gestures : What movement feature has the best timing? In B. Caramiaux, K. Tahiroglu, R. Fiebrink, & A. Tanaka (Eds.), *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 201–206). London, UK: Goldsmiths, University of London.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*(2), 103–113. doi: 10.1177/030573569302100201

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: A case study approach. *Psychology of Music*, *35*(3), 381–401. doi: 10.1177/0305735607072652

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, *40*(5), 595–633. doi: 10.1177/0305735612449896

Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J.-P., & Leman, M. (2012). Assessing a clarinet player's performer gestures in relation to locally intended musical targets. *Journal of New Music Research*, *41*(1), 31–48. doi: 10.1080/09298215.2011.649769

Fenza, D., Mion, L., Canazza, S., & Rodà, A. (2005). Physical movement and musical gestures: A multilevel mapping strategy. In *Proceedings of Sound and Music Computing Conference, SMC 2005* (unpaginated). Salerno, Italy: SMC

Fiebrink, R., Trueman, D., & Cook, P. R. (2009). A metainstrument for interactive, on-the-fly machine learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 280–285). Pittsburgh, PA, USA: NIME

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5*(7), 1688–1703.

Françoise, J., Schnell, N., Borghesi, R., & Bevilacqua, F. (2014). Probabilistic models for designing motion and sound relationships. In B. Caramiaux, K. Tahiroglu, R. Fiebrink, & A. Tanaka (Eds.), *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 287–292). London, United Kingdom: Goldsmiths, University of London.

Freedman, N., & Grand, S. (Eds.). (1977). *Communicative structures and psychic structures*. Boston, MA, USA: Springer. doi: 10.1007/978-1-4757-0492-1

Gillian, N., & Paradiso, J. A. (2014). The gesture recognition toolkit. *Journal of Machine Learning Research, 15*, 3483–3487. Retrieved from http://jmlr.org/papers/v15/gillian14a.html

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, *2*(2), 106–118. doi: 10.1109/T-AFFC.2011.7

Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., & Doherty, C. (2013). The movements made by performers in a skilled quartet: A distinctive pattern, and the function that it serves. *Frontiers in Psychology*, *4*(NOV), 1-8. doi: 10.3389/fpsyg.2013.00841

Godøy, R. I., & Leman, M. (Eds.). (2010). *Musical gestures: Sound, movement, and meaning.* London, UK: Routledge.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception: An Interdisciplinary Journal*, *26*(5), 427–438. doi: 10.1525/mp.2009.26.5.427

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Boston, MA, USA: Harvard University Press.

Hachimura, K., Takashina, K., & Yoshimura, M. (2005). Analysis and evaluation of dancing movement based on LMA. In *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN2005* (pp. 294–299). Nashville, TN, USA: IEEE. doi: 10.1109/ROMAN.2005.1513794

Hackney, P. (2002). *Making connections: Total body integration through Bartenieff fundamentals*. London, UK: Routledge.

Hunt, A., Wanderley, M. M., & Paradis, M. (2003). The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, *32*(4), 429–440. doi: 10.1076/jnmr.32.4.429.18853

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12–35). New York, NY, USA: Routledge.

King, E. (2006). Supporting gestures: Breathing in piano performance. In A. Gritten & E. King (Eds.), *Music and gesture* (pp. 142–164). Aldershot, UK: Ashgate.

Kirn, P. (2013, May). *How music can predict the human/machine future.* Lecture presented at the 2013 re:publica conference 2013. Berlin, Germany. Retrieved from https://www.youtube.com/watch?v=Fr_9_6AyCP4

Laban, R. (1966). *Choreutics*. London, UK: Macdonald & Evans.

Laban, R., & Lawrence, F. C. (1947). *Effort* (Reprinted ed.). London, UK: Macdonald & Evans.

Larboulette, C., & Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International Workshop on Movement and Computing* (MOCO '15; pp. 21–28). New York, New York, USA: ACM Press. doi: 10.1145/2790994.2790998

Leman, M. (2008a). *Embodied music cognition and mediation technology*. Cambridge, MA, USA: MIT Press.

Leman, M. (2008b). Systematic musicology at the crossroads of modern music research. In A. Schneider (Ed.), *Systematic and comparative musicology: Concepts, methods, findings* (pp. 89–115). Frankfurt am Main: Peter Lang.

Leman, M. (2010). Music, gesture, and the formation of embodied meaning. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement and meaning* (pp. 126–153). London, UK: Routledge.

Luck, G., & Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception: An Interdisciplinary Journal*, *26*(5), 465–473. doi: 10.1525/mp.2009.26.5.465

Madgwick, S. O. H., Harrison, A. J. L., & Vaidyanathan, R. (2011). Estimation of IMU and MARG orientation using a gradient descent algorithm. In *Proceedings of 2011 IEEE International Conference on Rehabilitation Robotics* (pp. 1–7). Zurich, Switzerland: IEEE. doi: 10.1109/ICORR.2011.5975346

Maes, P.-J., & Leman, M. (2013). The influence of body movements on children's perception of music with an ambiguous expressive character. *PLoS ONE*, *8*(1), unpaginated. doi: 10.1371/journal.pone.0054682

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement.* Cambridge, MA, USA: MIT Press. doi: 10.1007/s11023-015-9356-z

McNeill, D. (1996). *Hand and mind: What gestures reveal about thought*. Chicago, IL, USA: University of Chicago Press.

Merleau-Ponty, M. (2002). *Phenomenology of perception* (C. Smith, Trans.). London, UK: Routledge. (Original work published 1945)

Miranda, E. R., & Wanderley, M. M. (2006). *New digital musical instruments: Control and interaction beyond the keyboard*. Madison, WI, USA: A-R Editions, Inc.

Müller, M. (2007). *Information retrieval for music and motion*. Berlin, Germany: Springer.

Naveda, L., & Leman, M. (2010). The spatiotemporal representation of dance and music gestures using Topological Gesture Analysis (TGA). *Music Perception*, *28*(1), 93–111. doi: 10.1525/mp.2010.28.1.93

Nijs, L., Lesaffre, M., & Leman, M. (2013). The musical instrument as a natural extension of the musician. In M. Castellegno, H. Genevois, & J.-M. Bardez (Eds.), *Music and its instruments* (pp. 132–133). Sampzon, France: Editions Delatour.

Norman, S.-J. (2015). Oskar Schlemmer's programmatic gesture research. In N. S. Sutil & S. Popat (Eds.), *Digital movement: Essays in motion technology and performance* (pp. 21–34). London, UK: Palgrave Macmillan. Retrieved from http://sro.sussex.ac.uk/56405/

Norman, S. J., & Blackwell, A. F. (2010). Gesture and embodied interaction: Capturing motion / data / value. *Leonardo Journal*, *43*(2), 198–199. doi: 10.1162/leon.2010.43.2.198

Ortmann, O. (1929). *The physiological mechanics of piano technique: An experimental study of the nature of muscular action as used in piano playing, and of the effects thereof upon the piano key and the piano tone*. Boston, MA, USA: Da Capo Press.

Parviainen, J., Tuuri, K., & Pirhonen, A. (2013). Drifting down the technologization of life: Could choreography-based interaction design support us in engaging with the world and our embodied living? *Challenges*, *4*(1), 103–115. doi: 10.3390/challe4010103

Parviainen, J., Tuuri, K., Pirhonen, A., Turunen, M., & Keskinen, T. (2013). Gestures within human-technology choreographies for interaction design In *Proceedings of the 10th International Gesture Workshop and the 3rd Gesture and Speech in Interaction Conference* (unpaginated). Tilburg, the Netherlands: Tilburg University. Retrieved from https://jyx.jyu.fi/dspace/handle/123456789/41920

Piana, S., Staglianò, A., Camurri, A., & Odone, F. (2013, May). *A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition*. Paper presented at the 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI 2013). Chania, Crete, Greece.

Piana, S., Staglianò, A., Odone, F., & Camurri, A. (2016). Adaptive body gesture representation for automatic emotion recognition. *ACM Transactions on Interactive Intelligent Systems*, *6*(1), 1–31. doi: 10.1145/2818740

Pirhonen, A., Parviainen, J., & Tuuri, K. (2013, April–May). *Choreographies at home as objects of design*. Paper presented at the CHI 2013 Workshop on Methods for Studying Technology in the Home, Paris, France. Retrieved May 26, 2017, from http://studyingthehome.wp.horizon.ac.uk/wp-content/uploads/2012/11/Choreographies-at-Home-as-Objects-of-Design.pdf

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), B51–B61. doi: 10.1016/S0010-0277(01)00147-0

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. doi: 10.3758/s13423-012-0371-2

Salazar Sutil, N. (2015). *Motion and representation: The language of human movement*. Cambridge, MA, USA: MIT Press.

Smith, H., & Dean, R. T. (Eds.). (2009). *Practice-led research, research-led practice in the creative arts: Research methods for the arts and humanities*. Edinburgh, UK: Edinburgh University Press.

Sullivan, G. (2009). Making space: The purpose and place of practice-led research. In H. Smith & R. T. Dean (Eds.), *Practice-led research, research-led practice in the creative arts: Research methods for the arts and humanities* (pp. 41–65). Edinburgh, UK: Edinburgh University Press.

Thompson, M. (2012). *The application of motion capture to embodied music cognition research*. Unpublished doctoral dissertation, University of Jyväskylä, Finland.

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The impact of the bass drum on human dance movement. *Music Perception*, *30*(4), 349–359. doi: 10.1525/mp.2013.30.4.34

Visi, F. (2017). *Methods and technologies for the analysis and interactive use of body movements in instrumental music performance*. Unpublished doctoral dissertation, Plymouth University, UK. Retrieved from: https://pearl.plymouth.ac.uk/handle/10026.1/8805

Visi, F., Coorevits, E., Miranda, E., & Leman, M. (2014). Effects of different bow stroke styles on body movements of a viola player: An exploratory study. In A. Georgaki & G. Kouroupetroglou (Eds.), *Proceedings of the ICMC|SMC|2014 Conference* (40th International Computer Music Conference and 11th Sound and Music Computing Conference; pp. 1368–1374). Retrieved from http://speech.di.uoa.gr/ICMC-SMC-2014/images/VOL_2/1368.pdf

Visi, F., Coorevits, E., Schramm, R., & Miranda, E. R. (2016). Analysis of mimed violin performance movements of neophytes. In R. Kronland-Martinet, M. Aramaki, & S. Ystad (Eds.), *Music, mind, and embodiment: CMMR 2015. Lecture Notes in Computer Science* (Vol. 9617; pp. 88–108). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-46282-0_6

Visi, F., Schramm, R., & Miranda, E. (2014a). Gesture in performance with traditional musical instruments and electronics. In *Proceedings of the 2014 International Workshop on Movement and Computing* (MOCO ’14; pp. 100–105). New York, NY, USA: ACM Press. doi: 10.1145/2617995.2618013

Visi, F., Schramm, R., & Miranda, E. (2014b). Use of body motion to enhance traditional musical instruments: A multimodal embodied approach to gesture mapping, composition and performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 601–604). London, United Kingdom: Goldsmiths University of London.

Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, *VI*(1), 53–72. doi: 10.1177/102986490200600103

Wilson, S. (2013). Building an instrument, building an instrumentalist: Helmut Lachenmann’s Serynade. *Contemporary Music Review*, *32*(5), 425–436. doi: 10.1080/07494467.2013.849871

Woolford, K. (2014). Capturing human movement in the wild. In *Proceedings of the 2014 International Workshop on Movement and Computing* (MOCO '14; pp. 19–24). New York, NY, USA: ACM Press. doi: 10.1145/2617995.2617999

Authors' Note

The first author would like to thank Andrew Telichan Phillips and Dr. Tae Hong Park at NYU for making the collaboration between the Interdisciplinary Centre for Computer Music Research (ICCMR) and the Music and Audio Research Laboratory (MARL) possible, and for their valuable contributions to the project. Many thanks also to the members of the MARL research group at NYU for their hospitality, friendliness and knowledgeable insights. Special thanks to the musicians that took part in the tests at NYU Steinhardt during summer 2015. The collaborative project between Plymouth University and NYU was generously supported by Santander Universities. This study was funded in part by the FWO-project "Foundations of expressive timing control in music" and it is part of the newly founded MAaV project of the Universidade Federal do Rio Grande do Sul.

All correspondence should be addressed to
Federico Visi
Institute for Systematic Musicology
Universität Hamburg
Alsterterrasse 1, R111
20354 Hamburg, Germany
mail@federicovisi.com

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi