Body and Space: Combining Modalities for Musical Expression

Marco Donnarumma
Department of Computing
Goldsmiths, University of London
md@goldsmithsdigital.com

Baptiste Caramiaux
Department of Computing
Goldsmiths, University of London
bc@goldsmithsdigital.com

Atau Tanaka
Department of Computing
Goldsmiths, University of London
atau@goldsmithsdigital.com

TEI 2013, February 10-13, 2013, Barcelona, Spain
ACM

Abstract
This paper presents work in progress on applying a Multimodal interaction (MMI) approach to studying interactive music performance. We report on a study where an existing musical work was used to provide a gesture vocabulary. The biophysical sensing already used in the work was used as input modality, and augmented with several other input sensing modalities not in the original piece. The bioacoustics-based sensor, accelerometer sensors, and full-body motion capture system generated data recorded into a multimodal database. We plotted the data from the different modalities and offer observations based on visual analysis of the collected data. Our preliminary results show that there is complementarity of different forms in the information. We noted three types of complementarity: synchronicity, coupling, and correlation.

Author Keywords
multimodal interaction; NIME; interactive music performance; gesture; expressivity; biosignals

ACM Classification Keywords
H.5.5 [Sound and Music Computing]: Methodologies and techniques.

Introduction

Multimodal interaction (MMI) is the integration of multiple modalities (or input channels) to increase information and bandwidth of interaction. MMI uses different modalities that offer complementary information about user input. These modalities might include, for example, voice input to complement pen-based input [9]. The combination of complementary modalities provides information to better understand aspects of the user input that cannot be deduced from a single input modality. The analysis of combined information could be useful in the contexts involving forms of user expressivity in input.

Research in game control, sign language recognition, and prosthesis control looks at complementary use of spatial aspects of hand gesture, and physiological biosignals such as muscle tension (electromyogram, or EMG). Zhang et al. combined EMG and accelerometer (ACC) data in a gesture recognition system for the manipulation of virtual objects [11]. Li et al. [7] demonstrate an automatic Chinese sign language recognition system based on EMG and ACC. Fougner et al. show that the multimodal use of EMG and ACC was effective in reducing the number of EMG channels needed compared to a biosignal-only prosthetic interface [6].

Audio, video, and motion capture modalities were used alongside EMG, heart rate or EKG, electroencephalography or EEG by [8] in a study focusing on social interaction in traditional musical ensemble performance.

In the field of New Interfaces for Musical Expression (NIME), sensor-based systems capture gesture in live musical performance. In contrast with studio-based music composition, NIME (which began as a workshop at CHI 2001) focuses on real-time performance. Early examples of interactive musical instrument performance that pre-date the NIME conference include the work of Michel Waisvisz and his instrument, The Hands, a set of augmented gloves which captures data from accelerometers, buttons, mercury orientation sensors, and ultrasound distance sensors [5]. The use of multiple sensors on one instrument points to complementary modes of interaction with an instrument [1]. However these NIME instruments have for the most part not been developed or studied explicitly from an MMI perspective.

In this work-in-progress report, we show the relevance of considering biosignal-based multi-modality for the analysis of expressive musical gesture. This follows initial work looking at the integration of biosignals with relative position sensing [10]. Here, we describe a system for capturing three input modalities from the arm gestures of a musician. We then present the data from the different modalities together to look at the relationships amongst them. We identify three types of complementarity and discuss perspectives for future work.

Method

We conducted a pilot study recording musical gestures using 3 input modalities, and performed an observation-based analysis on mutual relationships in the data. The different modalities (mechanomyogram or MMG, accelerometer, and motion capture) detect physiological, movement, and spatial position information, respectively.

Sensor apparatus

The MMG is a signal generated by subcutaneous mechanical vibrations resulting from muscle contraction. For this we used the Xth Sense (XS), a biophysical NIME

instrument[1]. The XS consists of an arm band containing an electret condenser microphone (Kingstate KECG2742PBL-A) where acoustic perturbations from muscle contraction are digitised as sound at a sampling rate of 48000 Hz. Two channels of MMG were recorded, from the right and left arms over the wrist flexors, a muscle group close to the elbow joint that controls finger movement. The right sensor was located on the ring finger flexor, and the left one was slightly offset towards the little finger flexor.

The accelerometer sensor was a 3-axis DUL Radio (Analog Devices ADXL345) sending wireless data through a Nordic transceiver with a bandwidth of 2.4GHz, and sampling rate of 100Hz. The sensor was located on the back of the forearm, close to the wrist.

The motion capture system was an Optitrack Arena. This consisted of a full body suit, with 34 markers, and 11 LED/Infrared cameras (V100:R2) with a sampling rate of 60 FPS. We recorded 24 rigid bodies, that is, groups of markers representing limb part positions.

Data acquisition
The data collected was comprised of:

- 2 MMG audio signals, along with 6 amplitude/time domain sub-features extracted from analysis of the MMG audio stream;
- one 3D vector from the accelerometer;
- 24 rigid bodies, consisting of 7D signal for each rigid body: 3D position and 4D quaternions

The data was synchronised and captured by custom software developed in the Max/MSP graphical programming environment[2]. The data was time tagged, and stored in text format for offline analysis.

Gesture vocabulary
We used an existing piece of interactive music entitled *Music for Flesh II*, a work for XS that has had repeated performances by author 1 [4]. In the piece, the XS sends MMG from arm gesture to a computer programme to articulate sound and further process sound with the same muscle data. The composition is based on a vocabulary of 12 arm gestures. These gestures comprised the gesture vocabulary for our experiment.

The gestures, described in metaphorical, musical terms are:

1. Shaking bells
2. Scattering sound grains 1
3. Stretching sound 1
4. Stretching sound 2
5. Dropping something small
6. Rotating bells
7. Grasping the void
8. Shaping a wave
9. Throwing a sound wave
10. Holding a growing force
11. Scattering sound grains 2
12. Rotating wheel

[1] Developed by the first author.
http://res.marcodonnarumma.com/projects/xth-sense/

[2] http://www.cycling74.com

Figure 1: Example gesture *Stretching sound 2*

Each gesture was repeated 10 times with different expressive articulation, that is, changing speed and intensity in each iteration. Fig. reffig:gesture illustrates a trial of gesture 4 (*Stretching sound 2*). This gesture consists of fast wrist rotation, and faster flexion of the distal phalanges.

A detailed description of each gesture as well as the complete database can be seen on-line[3].

Results

In this section we report results focusing on the mechano-myogram and accelerometer data. In the following graphs the accelerometer x, y, and z axes are the green, red, and blue traces, respectively. The MMG sound wave is the black trace.

[3] http://marcodonnarumma.com/submissions/TEI2013/

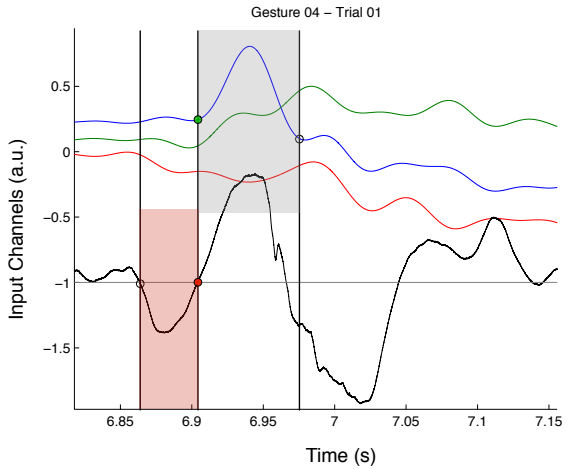

Figure 2: Detail of the MMG and ACC attack envelopes.

Fig. 2 shows the MMG (black) and ACC (blue, green, red) data recording from Gesture 4 (fig. 1). The red zone highlights the onset of muscle activity. The grey zone highlights the onset of accelerometer data. The initial activity in the muscle anticipates accelerometer data, with a second peak in MMG coinciding with the first peak in ACC. This shows the preparatory activity in the muscle at the beginning of a gesture that is not reported by the accelerometer sensor.

The graphs in Fig. 3 show two iterations of Gesture 1 (*Shaking bells*), a gesture that consists of multiple repeated contractions of the right hand fingers. The gesture was executed with two different kinds of expression, weak (low muscle force) and strong (high muscle force). By plotting the MMG and the 3 axes of the ACC data, we are able to see if the signals are coupled - whether the different modalities parallel each other, and

coincide, or not. At a weak articulation the oscillating nature of the finger grasping is slightly visible in the MMG trace. The accelerometer axes do not seem coupled to the MMG or to each other. At a higher articulation force, the signal for all the modalities undulates at close frequency in a coupled manner.

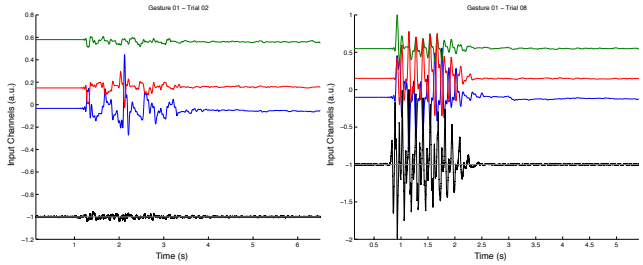

Figure 3: Gesture 1 executed at two different force levels.

Fig. 4 shows a more complex gesture, (*Scattering sound grains 2*). This consists of repeated *subito* finger clasping alongside swift wrist rotation. Here we see a change in correlation across modalities over the course of the gesture. Based on this shifting correlation, we divided the gesture into three segments. The first and third segments exhibit high correlation between MMG and ACC components, while in the second segment the subtleties of the MMG envelope are not correlated with the peaks in the ACC data.

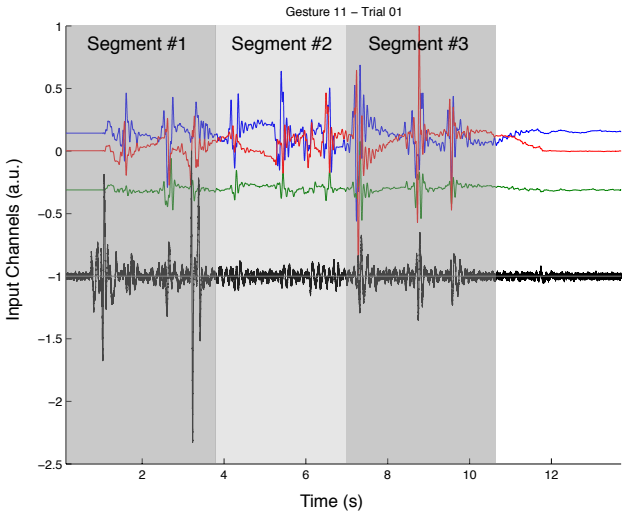

Figure 4: Correlation and independent variations of MMG and ACC channels within the same gesture.

Discussion

Synchronicity

In Gesture 4, we noted a time offset between MMG and ACC onsets. This may be due to different factors. The two modalities - muscle force and limb acceleration - may be asynchronous. One modality may have a lower latency than another. Or, the corporeal sensor may detect preparatory activity of the body that is not seen in the spatial/physical sensors. This could have useful application whereby the preparatory nature of one modality may aid in anticipating the arrival of data on another modality.

Coupling

We noted that coupling of modalities can vary with changing expressivity of the same gesture. In Gesture 1,

during weaker articulation of the gesture, the MMG and the 3 ACC axes were decoupled while at stronger expression of the same gesture, we observed tighter coupling in these modalities. This may indicate that two "regimes" can be engaged: either a low force with fairly independent signals or high force with coupled signals. This has potential to be the subject of further study.

Correlation
We observed the shifting of cross-modal correlation within the span of one gesture. We visually segmented zones that displayed high and low correlation between MMG and ACC. This may point out differing levels of dependency between muscle control and the resulting movement. Interestingly, the ACC data in Gesture 11 have more or less the same amplitude throughout the gesture, but the MMG amplitude varies greatly, pointing out a difference in dynamic range in the different input modalities. This may be studied in more detail using automatic feature extraction methods [2], to examine quantitatively which features of the MMG and ACC signals are correlated.

Observations on other modalities
In gestures characterised by finger contractions (1, 2, 4, and 11), unless a marker is positioned on each finger, the tracking of phalange movement is not detected. Motion capture is, however, useful in conveying information about the movement of other parts of the body such as head and torso. Studies of musical instrument performance [3] point out the importance of such corporeal movement to musical expressivity. This auxiliary information, beyond the constraint space of multiple input modes sensing limb gesture, could be useful in multimodal musical interaction.

Conclusion

This paper reported on preliminary work applying a Multimodal Interaction approach to analyse expressive musical gesture. By using three distinct modalities to track gesture in the performance of an existing NIME-type work, we were able to make several early observations on the relationships between the kinds of information reported by the different input channels. We found that: the complementarity across different input modalities may be due to differing sensitivity of sensors to the preparation of a gesture; what seems like a single musical gesture may be comprised of different sections where the relationship amongst modalities may change over time; a range of modalities might detect the independent control different aspects of a single gesture.

The results presented are initial observations on a subset of the data we collected. One challenge in working across a diverse range of information-rich modalities is in the data management and numerical processing. Signal acquisition and synchronisation across free-standing hardware systems, different sampling rates, and network communication latency pose additional technical challenges that would need to be rigorously addressed in an in-depth study.

One element that made up part of our capture session which we did not report on in this paper is the capture of audio (of the musical output) and video (of the performer conducting the gesture). These media channels could provide useful points of reference, and might, given the appropriate configuration, even be exploited as auxiliary input modalities.

References

[1] A. Camurri, P. Coletta, G. Varni, and S. Ghisio. Developing multimodal interactive systems with EyesWeb XMI.

Proceedings of the 7th international conference on New interfaces for musical expression - NIME '07, page 305, 2007.

[2] B. Caramiaux, F. Bevilacqua, and N. Schnell. Towards a gesture-sound cross-modal analysis. In *In Embodied Communication and Human-Computer Interaction, volume 5934 of Lecture Notes in Computer Science*, pages 158—-170. Springer Verlag, 2010.

[3] J. W. Davidson. Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. *Psychology of Music*, 35(3):381–401, July 2007.

[4] M. Donnarumma. Incarnated sound in Music for Flesh II. Defining gesture in biologically informed musical performance. *Leonardo Electronic Almanac (Touch and Go)*, 18(3):164–175, 2012.

[5] E. Dykstra-Erickson and J. Arnowitz. Michel Waisvisz: the man and the hands. *Interactions*, 12(5):63–67, Sept. 2005.

[6] A. Fougner, E. Scheme, A. D. C. Chan, K. Englehart, and O. Stavdahl. A multi-modal approach for hand motion classification using surface EMG and accelerometers. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011(Grant 192546):4247–50, Jan. 2011.

[7] Y. Li, X. Chen, J. Tian, X. Zhang, K. Wang, and J. Yang. Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors. *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on - ICMI-MLMI '10*, page 1, 2010.

[8] O. Mayor, J. Llop, and E. Maestre. RepoVizz: A multimodal on-line database and browsing tool for music performance research. In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, USA, 2011.

[9] S. Oviatt, R. Coulston, S. Tomko, B. Xiao, R. Lunsford, M. Wesson, and L. Carmichael. Toward a theory of organized multimodal integration patterns during human-computer interaction. *Proceedings of the 5th international conference on Multimodal interfaces*, pages 44–51, 2003.

[10] A. Tanaka and R. B. Knapp. Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. *Proceedings of the 2002 conference on New interfaces for musical expression*, pages 1–6, 2002.

[11] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, and K.-q. Wang. Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. In *Proceedingsc of the 13th international conference on Intelligent user interfaces - IUI '09*, page 401, New York, New York, USA, 2008. ACM Press.