Cooperative musical creation using Kinect, WiiMote, Epoc and microphones: a case study with *MinDSounDS*

Tiago Fernandes Tavares, Gabriel Rimoldi, Vânia Eger Pontes, Jônatas Manzolli
Interdisciplinary Nucleus of Sound Communication
University of Campinas - Brazil
tiago@nics.unicamp.br

ABSTRACT

We describe the composition and performance process of the multimodal piece *MinDSounDS*, highlighting the design decisions regarding the application of diverse sensors, namely the Kinect (motion sensor), real-time audio analysis with Music Information Retrieval (MIR) techniques, WiiMote (accelerometer) and Epoc (Brain-Computer Interface, BCI). These decisions were taken as part of an collaborative creative process, in which the technical restrictions imposed by each sensor were combined with the artistic intentions of the group members. Our mapping schema takes in account the technical limitations of the sensors and, at the same time, respects the performers' previous repertoire. A deep analysis of the composition process, particularly due to the collaborative aspect, highlights advantages and issues, which can be used as guidelines for future work in a similar condition.

1. INTRODUCTION

MinDSounDS is an multimodal piece for computer, movement, WiiMote, flute, Brain-Computer Interface (BCI) and images, world premiered at the Generative Arts 2014 conference (December 2014, Rome). It was composed to be controlled live by a group of performers by means of a network of consumer sensors. The work is based on previous piece, namely Re(PER)Curso [1], and illustrates how the aesthetic experience can be related to an organization that emerges from the interaction between the performers and a virtual environment.

MinDSounDS narrates the story of a virtual avatar – a humanoid projection on screen – that learns the movements of a human dancer and builds its own movements. This process is mediated by human performers, which interact among themselves and with the virtual environment. As the avatar builds its own movements, it also interacts with humans, thus actively joining the performance group.

We defined that the piece would be composed by the whole group, without a prior agreement on its content or its language. Each of the involved musicians, which are the authors of this paper, had their own set of skills and their own artistic intentions towards what *MinDSounDS* should become. Communication in these conditions has proven essential, and, at the same time, not trivial, as it is easy to find misunderstandings of several natures.

We conducted a collaborative composition process for related to each one of these instruments, which gave rise to specific problems and advantages related to group work. Prior work by Cornacchio [2] has discussed issues related to group musical composition in music classrooms, and we have noticed some similarities to our process. However, our process was not bounded to a clear goal or musical language, which gave rise to specific difficulties and discussions.

Through this process, we developed the piece as an expression of the group's multidisciplinarity, which reflected in the sensor network multimodality. Because of the group's cooperation, we were able to build interesting mappings between the sensors inputs and their sonic and visual representations. The use of different sensors was a natural result of the process, as each of them had an important artistic contribution to the piece.

The group's composition proposal allowed the development of an iteractive method for composing mappings between gestures and media, which was especially important in the case of the Kinect. Prior art mainly focused on mappings defined by the composer and delivered as instructions for the performer [3, 4] or in processes in which the composer and the performer are the same person [5–7]. In *MinDSounDS*, the composition process considered a dance movement repertoire as part of the performance, thus composing a virtual environment that enhanced the movement possibilities of the performer.

The result of our process also presents sensible differences from prior art. We do not design a virtual environment that emulates real interactions [6, 8], and, at the same time, we do not design an arbitrary virtual instrument [3, 4] or interactive control of sound effects [5, 7]. Instead, we use motion data to augment the expressive possibilities of the dancer, respecting their original repertoire and progresively exploring new expressive aspects.

Our approach towards the Epoc was also significantly different from related work using BCI. We have found that previous work has largely focused on the sonification of brain waves [9–11], which means that sound is generated using voltages measured in the scalp as raw material. In this approach, the musical intentions of the user are disregarded during the composition process, even if they can be

indirectly controlled by training.

In other approaches, BCI was used to trigger events, attempting to mimic actions that could be performed using the body [12, 13]. However, state-of-the-art BCI systems yield several false negatives and false positives in intentional triggers. Therefore, previous work has used post-filtering techniques like offline usage [14], beat synchronization [12] and low-pass filtering [13] to overcome these difficulties.

We overcame this problem by incorporating the BCI concept into the piece construction. The BCI device was responsible to mediate a high-level process whose fine details were controlled by the dance performer and a timer. Therefore, we incorporated the BCI in a context in which false positives and false negatives would not cause drastic consequences to the performance.

The remainder of this paper is organized as follows. Section 2 describes the implementation of the sensor network. Section 3 discusses the advantages and drawbacks found in the composition and performance processes. Last, Section 4 brings conclusive remarks.

2. SENSORS

MinDSounDS relies on the interaction between performers and a virtual environment by means of sensors. This interaction took place by means of mapping between sensor data and sound and visual representations. The process of building these mappings was an important part of the construction of the virtual environment.

An important aspect of *MinDSounDS* is that it aims at creating specific causalities between inputs and outputs, that is, avoids generative processes that are not controlled – or, at least, controllable – by the performers. This comes from the group's perception that the audience should be able to understand the relation between the perforers' movements and the audio and video responses. Thus, our composition process greatly acounted for consistency between actions and their mappings.

During the composition process, we used different kinds of sensors to provide musicians with diverse expressive possibilites. As depicted in Figure 1, each sensor data is used in a different context, and interferes with other sensors, jointly controlling synthesis processess. Below, we present a throrough explanation of the interaction related to each sensor.

We used a motion sensor to capture dance movements of a performer. Different movements should lead to different sonic responses, but it is initially unclear how to make these responses meaningful to the piece's intention and the performer's repertoire. In Section 2.1, we describe how the process of building this mapping was conducted to mediate between these aspects.

A game controller with an accelerometer emulate virtual bells. Due to the computational nature of these bells, we found issues concerning the sensor's sensitivity. Also, as we describe in Section 2.2, the accelerometer was added with dynamic filtering capabilities, increasing the expressive potential of the controller.

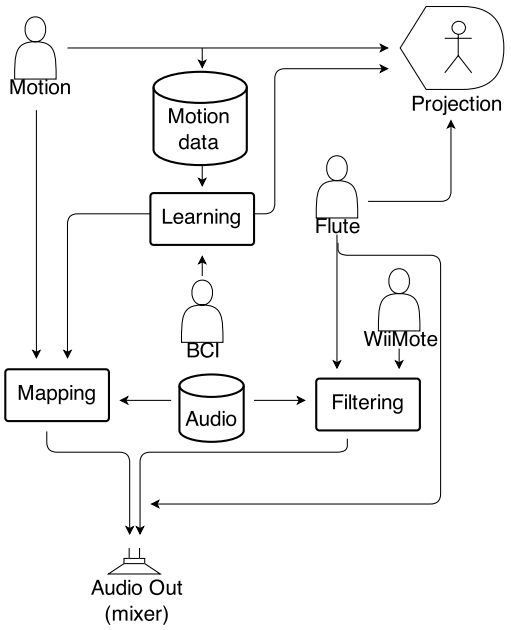

Figure 1. MindSounds interaction diagram, depicting how each sensor data interacts with the others.

We also explored the Brain-Computer Interface (BCI), which translates voltages between key points in the user's scalp to triggers thay may be used as game controllers. The BCI has been increasingly used in musical contexts for different purposes. We have developed a particular musical language, suitable for both the purposes of the piece and the characteristics of the BCI, which is described in Section 2.3.

Last, Music Information Retrieval (MIR) techniques allowed using an acoustic flute as a controller. The information derived from these techniques was bounded to the control of characteristic of a video projection, thus the composition process raised new possibilities, as well as specific restrictions. We describe this process, and its results, in Section 2.4.

2.1 Kinect

The Kinect is a motion capture sensor developed for gaming purposes. Using specialized software, it is possible to obtain a tri-dimensional position (using $p = (x, y, z)$ triples) for each of the body's limbs (elbows, knees, hands, etc.) at a frame rate of 30 Hz. The positions of limbs were interpreted as related to the performer's torso, namely the kinesphere.

The kinesphere was used due to the performer's dance repertoire, which comprises mostly arm and leg positionings as a form of expression. The kinesphere allowed a more precise acquisition of these movements, while at the same time disregarding jumps and dislocations through the stage. From a purely technical view, this also added the advantage of reducing the time required to calibrate the sensor to different venues.

There is no theoretically best mapping between limb movements and controls, as this depends on the perform's movement repertoire, sound designer's technique repertoire and

the piece's intention. Since the piece's creative intention was unclear at early stages of the composition process, the mapping's construction comprised several iteractions between the performer and the sound designer, assisted by the remaining of the group. In this process, mapping proposals were presented and discussed, leading to a final decision.

The mappings we have found more interesting for the piece are shown in Table 1, but it is very likely that they will be re-built in other future work. This will happen not because they are not good in any sense, but because they are the result of a composition process, which will, inevitably, happen again. However, we have developed useful strategies for finding this mapping, which may be employed again in the future.

Movement	Control
Hands around kinesphere	Spatialization (panning) Sample selection Video control
Distance between hands	Pitch Sample selection
Feet velocity	Sound intensity
Relative feet position	Granulation control

Table 1. Mapping of gestures to controls using the kinect

We have found that it may be useful not to map all movements to audiovisual representations. This gives the performer a greater freedom to develop a more natural dance sequence, including movements whose contribution to the piece is solely visual. This means that, while the motion sensor enables live control of computer-based sound and video, it may also constrain dance movements, potentially harming the performance.

The same hold for another decision, regarding the nature of the movements that will be mapped. Nowadays, there exists technology that allows mapping specific dance moves (for example, a spin) to an event trigger. We did not want to use this because we wanted to allow an exploration and improvisation process to be part of the dance performance.

Therefore, we opted to use more general movement parameters as controls. An example that worked was the panning control, done by the position of the hands around the kinesphere. This mapping allows a great variation on the movement, for example, regarding the performer's elbows and shoulders, while resulting in the same controls.

We have also noted that discrete controls that trigger to specific movements should be used carefully. Triggers are efficient for some purposes, like selecting sound samples, but they may restrict the performer's movements in order to avoid false positives or false negatives. Thus, their extensive use may inhibit the performer's fluency.

Continuous mappings, on the other hand, are unable to trigger discrete events. In our composition process, they were easier to incorporate into the dance performance, because they were felt more as movement suggestions than as coreographed steps. Thus, we were careful to maintain balance between discrete and continuous movement mappings.

By using movement velocity as a sound intensity control, we were able to map a perceived visual effort to a perceived auditory effort. This helped on our goal of allowing the audience to understand the mapping process, as it emulates the behavior of acoustic instruments. In these instruments, a stronger effort usually reflects on a stronger sound, allowind the control of event dynamics, which are important for expressive performances.

It is also important to note that hese mappings were not all used at the same time, but scattered on particular movements of the piece. Each of them induced a different exploration of the sonic space by the performer, leading to the use of a different repertoire of gestures, sounds and visuals, as highlighted in Figure 2. Thus, although we aimed at not creating an invasive and restrictive virtual environment, the interaction possibilities inevitably favoured particular movements over others.

The process of finding mappings, gestures and sounds that would fit the purposes of the piece demanded a great amount of interaction between all members of the group, especially the sound designer and the dancer. During this process, one of the greatest problems we faced was due to the abscence of a language that could consistently and efficiently convey sonification ideas, which lead to many misunderstandings. Another problem is that the implementation of a new mapping proposal was very time-consuming, as it demanded understanding the movement and translating it into code.

We used a similar interactive approach to develop mappings and sonifications for the WiiMote. The nature of the controller lead to the development of different algorithms. The process regarding the WiiMote is described in the next section.

2.2 WiiMote

The WiiMote is a handheld console that contains nine buttons and a three-axis accelerometer, which were mapped according to Table 2. Using third-party software, it is possible to acquire the accelerometer data at 100 Hz, as well as triggers related to pressing the buttons. In comparison to the Kinect, it has a faster response, but also yields significantly more noise.

Input	Control
Button A	Enable percussion
Slap gesture	Use percussion
Directional buttons	Record data for adaptive filter
Accelerometer	Control filter interpolation
Button B	Use filter

Table 2. Mapping of inputs to controls using the Wiimote. They are further explained along this section.

The device was used to control a virtual percussion device. This functionality could be enabled or disabled through one of the buttons, and, if enabled, triggered by using the device as a drumstick in the air, in a *slap* gesture.

Detecting a slap gesture was done detecting acceleration values above a pre-defined threshold in any axis. The pitch

Figure 2. Examples of the interaction in two different movements of the piece. Different movements were used to control different visual representations.

and roll parameters during the *slap* gesture controlled filters that would modify the percusive sounds. Thus, different angles of attack resulted in sounds with diverse spectral content.

The controller was also linked to an interpolated filter derived from ambient sound. This application was based on recording sound samples captured from microphone and interpolating them, using the result as the impulse response of a FIR filter. In our piece, the we acquired sound samples from the acoustic flute, and applied the resulting filter to pre-recorded vocal samples that controlled the soundscape.

To control this functionality, four buttons were used to trigger recording in four different audio buffers. The resulting impulse response would correspond to their weighted sum, in which the weights were controlled by the pitch and roll of the WiiMote. While a fifth button was pressed, the system would apply the filter to the audio output.

Hence, a variable, interpolated filter was developed. Its control using the accelerometer quickly became intuitive, with the advantage of preserving the presencial action of the performance because of the live movements of the performer. The hardware has show to be reliable and fast for low-level audio control, which was not the case for all sensors, as will be seen.

2.3 Epoc

The Epoc is a consumer device that provides a Brain Computer Interface (BCI). It consists of several electrodes and an accelerometer, which provide readings of the Electroencephalogram (EEG). Its software suite works under the assumption that similar thoughts correlate to similar EEG signals, hence allowing memorizing mental states and ultimately providing the ability to use thoughts to control software.

We have found two main problems with the use of the device, which are shared by many BCI approaches. The first problem is the instability of training – the system has to be calibrated each time it is used, and the user must keep a clear mind during the use. The second is the high amount of false positives and false negatives.

These limitations were avoided by using the device in a context that allows for errors without drastic consequences to the performance. This means that we developed a musical paradigm in which these false negatives and false positives would be part of the musical discourse, instead of undesireable artifacts. For this reason, we used the BCI device for the control of high-level parameters.

In the musical context, the BCI was used in a piece movement in which the avatar is learning the movements of the dancer. These movements are recorded directly from the dancer, in previous movements. The learning process is represented by the application of a recombination algorithm.

The recombination algorithm takes as input the recordings of the dancer's limb position. Then, it applies a random time-shift in each stream, thus creating combinations of limb positions that are impossible to be performed by a human being, but are rendered on the screen creating a perceptually weird form. Through the learning process, the amount of time-shift allowed in each stream is reduced, which makes the rendered form gradually assume humanoid appearance, leading to the perception that the avatar is slowly immitating the performer's movements.

In this context, the BCI device was used to trigger a next step in the learning process, corresponding to a new maximum value for the random time-shift. The next maximum time-shift is defined as the previous value minus a fraction of the elapsed time since the start of the previous step. The beggining of each step is also marked by the sound of a bell.

As a result, it was possible to estimate the duration of the movement and its possible outcomes, which was useful for planning the interaction with the other musicians. The detail-level of the piece, that is, the time when each learn-

ing step would be triggered, could be actively controlled by the musician. This way, we were able to overcome the limitations of the device while still using it in a meaningful way.

2.4 Computational Ear

Using MIR techniques, we were able to use an acoustic flute as a musical controller. Audio was acquired from the instrument using a microphone, and processed yielding spectral and temporal features of sound. Later, these paramenters were used to modify the visual part of the piece.

We chose to use two audio features to control continuous values in video processing. The Chroma feature determines a range of hue while the Loudness determines the luminosity of rendered textures on video. This allowed mapping note classes to projection colors, which was done arbitrarily.

However, the decision of using audio for this purpose implied in other artistic decisions. The chosen features (Loudness and Chroma) only make sense in the context of sound with defined pitch. This means that, while this controller was used, the performer should explore sonorities in which pitch remain as a main parameter.

The technical issues presented in this section had a deep impact on the final format of the composition. They were an important part of the composition process through which we obtained a sensor network aimed at building the concept of Presence in the context of the piece. Further discussion regarding this process will be conducted in the next section.

3. DISCUSSION

The process of composing *MinDSounDS* was a cooperative process that integrated both the artistic and the technological points of view. As a result, we developed significant advances impacting both the final outcome – the piece itself – and the conduction of its composition process. Hence, we believe that *MinDSounDS* can be part of a base repertoire in future work.

In the cooperation process, we found problems that may arise in diverse environments. Since there was no prior guideline to follow, the group struggled to make *MinDSounDS* an artistic expression that comprised the desires of all musicians. This is partially inevitable, and an important part of the cooperation process, but we were able to detect some guidelines that may be useful in the future.

Group time management, in our process, was poor, which meant that in several ocasions there were scheduled activities that did not require the presence of some group members. This lead to a waste of time and contributed to the loss of focus. Although we were aware of this, it was not an easily avoidable situation because the objectives of each activity were not clear during the process.

Another issue related to time regards the fact that the musician that would perform with the sensor device was not the composer of the corresponding interaction. This implied in an interactive composition process in which there are two oposing points of view, one related to building a lean, usable system and the other related to constructing an artistically meaningful interaction. We adopted the solution of composing partial mappings, as discussed before, but this process had its own difficulties.

The interaction between the composer and the performer consisted of taking proposals by both musicians and trying to explore its possibilities (for the performer) or trying to implement it (for the composer). As the perform explores possibilities, new proposals arise, and the same holds for the implementation of the proposals by the composer. The first issue regarding this process involved finding proposals that could integrate the musical background of each musician, as well as the piece's proposal.

Another problem was related to the long time required for implementing proposals by each composer. This inevitably generated long periods of idle time, which had a negative impact on activity sessions and, ultimately, in the interaction process. Therefore, we detected a clear demand for a framework allowing these interactions to be built faster, so that the exploration and composition process may follow the musicians' pace.

We also faced problems regarding the construction of the piece's artistic proposal. Since we did not have a clear idea of what we were trying to implement, or even the musical language that we would follow, the final result emerged from our interactions. Following this proposal is advantageous in the sense that it allows experimenting a broader range of techniques, but also prevents a deeper individual experimentation on particular issues.

The indefinition of the expected result of a process is a known and well-studied issue both in music – for example, in improvised performance – and computer science – as there are specific software engineering technicques that deal with it. The case composing *MinDSounDS* is different from an improvised performance because the group was also responsible for building the musical instruments, and, moreover, each instrument had a deep impact on the others. Also, it was not the same as a software engineering case because the problem was not supplying functionalities for a client's demand, but building the demand from an initial, abstract idea.

Thus, it became clear that we lacked an effective process for communication of repertoire, expectations and analysis of the results. This points to a direction for future work, which is studying issues related to composing music in groups without a prior style agreement. In this sense, it is important to preserve artistic freedom and the feeling of participation, while introducing guidelines for cooperation.

Nevertheless, the piece was succesfully composed and presented, and is now a unit of structural cohesion. This property emerged from the composition process, generating a unique piece in which all parts involved presented important contributions. Also, this process was an important step towards understanding musical cooperation, and its analysis will have great impact on future work.

The mappings and algorithms we employed in the piece were also the result of this cooperation process. This process was different from two very frequent ones: the solo

musician that is both the composer and the performer, and the cascade workflow in which the perform executes instructions from the composer. Thus, composing lead to a greater understanding of each musician's role in the piece, and, from this point of view, this process was more important than the final result.

4. CONCLUSION

We described the process of composing the multimodal piece *MinDSounDS*, highlighting the technological and artistic issues that arised. We showed how each sensor was applied on the control of specific parts of the piece. Moreover, we discussed how the process of finding these mappings was relevant to the piece.

The piece was composed in a cooperative process, without the pre-definition of a final objective or an explicit artistic language. This gave rise to a series of problems, which were handled by the group and had a deep impact on the composition process. Finally, we finished and presented the piece, and also learned on aspects that could be improved in future work.

We take special care on presenting how each sensor cooperates to the piece. We discuss the algorithms and technological limitations of each sensor. As a result, the use of each sensor becomes differentiated, improving its contribution to the final artistic result.

Addressing technical and artistic limitations, especially the cooperation issues during composing and rehearsing, present a clear direction for future work. This direction should point at developing protocols that allow a creative interaction between composers and performers while providing and effective use of the team's time. These aspects are often conflicting, but this is a problem that must be studied in order to make cooperative composition a more efficient process.

Acknowledgments

The authors thank the Brazilian agencies FAPESP and CNPq for funding this research.

5. REFERENCES

[1] A. Mura, J. Manzolli, P. F. M. J. Verschure, B. Rezazadeh, S. L. Groux, S. Wierenga, A. Duff, Z. Mathews, and U. Bernardet, "re(per)curso: An interactive mixed reality chronicle," in *SIGGRAPH*, Los Angeles, 2008.

[2] R. A. Cornacchio, "Effect of cooperative learning on music composition, interactions, and acceptance in elementary school music classrooms," Ph.D. dissertation, Graduate School of the University of Oregon, 2008.

[3] M.-J. Yoo, J.-W. Beak, and I.-K. Lee, "Creating musical expression using kinect," in *Proceedings of NIME*, 2011.

[4] A. R. Jensenius, "Kinectofon: Performing with shapes in planes," in *Proceedings of NIME*, 2013.

[5] G. Odowichuk, S. Trail, P. Driessen, W. Nie, and W. Page, "Sensor fusion: Towards a fully expressive 3d music control interface," in *Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on*, Aug 2011, pp. 836–841.

[6] S. Sentürk, S. W. Lee, A. Sastry, A. Daruwalla, and G. Weinberg, "Crossole: A gestural interface for composition, improvisation and performance using kinect," in *Proceedings of NIME*, 2012.

[7] S. Trail, M. Dean, T. F. Tavares, G. Odowichuk, P. Driessen, A. W. Schloss, and G. Tzanetakis, "Non-invasive sensing and gesture control for pitched percussion hyper-instruments using the kinect," in *Proceedings of NIME*, Ann Arbour, Michigan, U.S.A., May 2012.

[8] M.-H. Hsu, W. Kumara, T. Shih, and Z. Cheng, "Spider king: Virtual musical instruments based on microsoft kinect," in *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*, Nov 2013, pp. 707–713.

[9] E. R. Miranda and B. Boskamp, "Steering generative rules with the eeg: An approach to brain-computer music interfacing," in *Sound and Music Computing*, 2005.

[10] ——, "Toward direct brain-computer musical interfaces," in *New Interfaces for Musical Expression*, 2005.

[11] S. Mealla, A. Väljamäe, M. Bosi, and S. Jordà, "Listening to your brain: Implicit interaction in collaborative music performances," in *New Interfaces for Musical Expression*, 2011.

[12] S. L. Groux, J. Manzolli, and P. F. Verschure, "Disembodied and collaborative musical interaction in the multimodal brain orchestra," in *New Interfaces for Musical Expression*, 2010.

[13] T. Mullen, R. Warp, and A. Jansch, "Minding the (transatlantic) gap: An internet-enabled acoustic brain-computer music interface," in *New Interfaces for Musical Expression*, 2011.

[14] B. Hamadicharef, M. Xu, and S. Aditya, "Brain-computer interface (bci) based musical composition," in *Cyberworlds (CW), 2010 International Conference on*, Oct 2010, pp. 282–286.