Automatic recognition of Soundpainting for the Generation of Electronic Music Sounds

David Antonio Gómez Jáuregui, Irvin Dongo, Nadine Couture

Automatic recognition of Soundpainting for the Generation of Electronic Music Sounds

David Antonio Gómez Jáuregui
Univ. Bordeaux, ESTIA
Bidart, France
d.gomez@estia.fr

Irvin Dongo
Univ. Bordeaux, ESTIA
Bidart, France
Universidad Católica San Pablo
Arequipa, Peru
i.dongoescalante@estia.fr

Nadine Couture
Univ. Bordeaux, ESTIA,
LaBRI, UMR5800
Bidart, France
n.couture@estia.fr

ABSTRACT

This work aims to explore the use of a new gesture-based interaction built on automatic recognition of Soundpainting structured gestural language. In the proposed approach, a composer (called Soundpainter) performs Soundpainting gestures facing a Kinect sensor (® Microsoft). Then, a gesture recognition system captures gestures that are sent to a sound generator software. The proposed method was used to stage an artistic show in which a Soundpainter had to improvise with 6 different gestures to generate a musical composition from different sounds in real time. The accuracy of the gesture recognition system was evaluated as well as Soundpainter's user experience. In addition, a user evaluation study for using our proposed system in a learning context was also conducted. Current results open up perspectives for the design of new artistic expressions based on the use of automatic gestural recognition supported by Soundpainting language.

Author Keywords

Interactive system; Gesture recognition; Smart and Empowering Interfaces; Music; Soundpainting

CCS Concepts

•**Human-centered computing** → **Gestural input; Sound-based input / output;** *User studies;*

1. INTRODUCTION

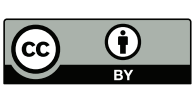
With the development of new technologies for interactive art, artistic expression has been propelled to new heights in which the public and the machine interact to produce, together and in real time, unique works of art. For this purpose, human-computer interaction (HCI) enables initiation and management of novel interactive processes [6]. Recently, gestures have been adopted as a new modality in the field of HCI in order to consider physical movements of the whole body, thus pushing human-system interaction limits by involving several human physical capabilities [7]. However, questions about HCI have to be considered for the creation of these new interactive artistic interactions. In particular, questions related to design of user experience, as well as artist's understanding and engagement, are particularly relevant [6]. The goal of this work is to explore the use of a new gestural interaction based on Soundpainting live composing language. Soundpainting is a sign language that was invented by Walter Thompson in 1974 in Woodstock. It is a multidisciplinary and universal language that enables live composition in real time with musicians, actors, dancers and visual artists [22, 23]. This language offers a well-defined grammar supporting interaction with a group of artists (e.g. musicians) as part of live performances in which artists themselves improvise (without using score).

We have developed a system capable of recognizing, in real time, Soundpainting gestures from a Microsoft Kinect®. The proposed system is used in a live demonstration, in which a Soundpainter improvises using Soundpainting gestures to produce musical sounds, where each sound corresponds to a different gesture, as illustrated in figure 1. Thus, a new context for Soundpainting is proposed. The gesture recognition system was evaluated, in real conditions, during a public live demonstration. System recognition accuracy has been measured and the feeling (user experience) of Soundpainters has been evaluated. The scientific challenges of this work are related to the accurate estimation of the pose allowing a strong coupling between the artist and the sound generation system, as well as recognition robustness using a real time art performance scenario. The purpose of the evaluation conducted in this work is, first, to determine whether the actions of Soundpainting are recognizable and treatable by an artificial intelligence system; then, whether quality and performance of Soundpainting gesture recognition are equivalent, superior or inferior to that of humans; finally, if the proposed gestural interaction is applicable in an artistic context.

Related work is presented in the next section. Soundpainting gesture recognition system and the music sound generation module are described in section 3. The live artistic demonstration used to evaluate the performance and the user experience of and with the proposed system are described in section 4. Results are presented section 5. Finally, section 6 concludes and proposes perspectives.

2. RELATED WORK

Several authors studied and proposed solutions for body use and gestural interactions for music generation. Tanaka in 2000 [21] was one of the first to exploit muscles signals using the Bio-muse system of Knapp [12] to produce music in real time. Later, in 2004, Paine [14] proposed detection of body movement from two synchronized cameras in order to transfer movement and behavior of the artists towards synthesized sounds. In 2012, Clay et al. [5] used a motion capture combination of 17 sensors (accelerometers, gyro-



NIME'19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

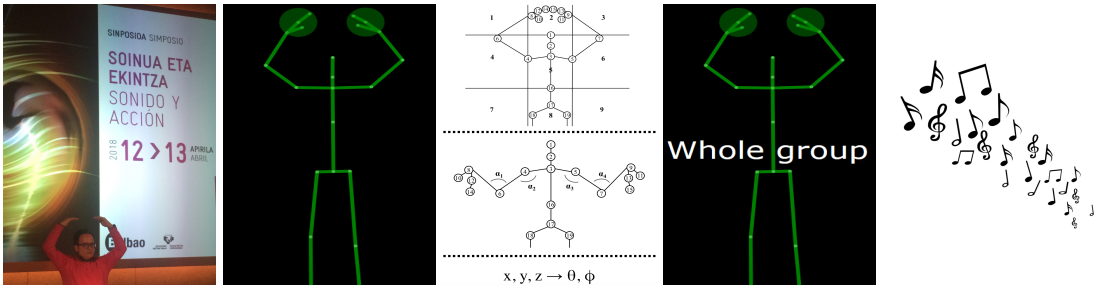


Figure 1: Gesture recognition and music production process, from left to right: the Soundpainter signing the gesture "whole group" of the Soundpainting; the skeleton recovered by the Kinect®; extracted features; gesture recognized and transmitted to the sound generation module; electronic music generation.

scopes and magnetometers) to accurately capture dancer's movements. A year later, Christopher et al. [3] placed a three-axis accelerometer on each finger and the hand of a musician to provide high-resolution details of expressiveness of his gestures. Expressiveness was used to control a synthesizer software. Recent research works have proposed musical interfaces based on direct mappings between body movements and sounds. In 2015, Françoise et al. [8] proposed a system where the voice is used in conjunction with body movement allowing users to consciously create their own personal motion-sound mappings Recently, Sarasúa et al. [19] proposed an interface based on the conductor-orchestra metaphor that allows to control tempo and dynamics and adapts its mapping specifically for each user by observing spontaneous conducting movements. Although all these previous work have proposed innovative and creative solutions for music generation by gestural interaction, they did not use a formal gesture language, which relies on a well-defined grammar, to generate music. In the proposed approach, we explore the use of automatic recognition of Soundpainting gestures for artistic music generation.

Automatic gesture recognition is a hot topic in computer science research that aims to interpret human gestures using machine learning algorithms from information provided by cameras or motion sensors. More recently, emergence of inexpensive cameras that combine RGB information with depth information (such as Kinect®) has led to significant advances regarding new methods of gesture recognition in real time. For instance [2] use depth information provided by these cameras as they are not impacted by sudden changes in lighting variation from uncontrolled environments. In addition, availability of 3D coordinates of joints extracted from this information enables having a precise description of human body configuration [4]. Nowadays most gesture recognition approaches use machine learning algorithms that train and learn from several features extracted from human body postures [4, 24]. These approaches have been used to recognize automatically musical gestures [10, 13]. Despite the large number of efforts on gesture recognition, to our knowledge, only Guyot and Pellegrini's work [11, 15] explored automatic recognition of Soundpainting language. They propose a system for automatic annotations of Soundpainting gestures allowing, for pedagogical purposes, to compare the same gesture made by several Soundpainters. However, performance of the proposed system was not validated quantitatively.

3. SYSTEM IMPLEMENTATION

Our implementation is based on two systems (figure 2): a gesture recognition system, where gestures produced by a Soundpainter are analyzed in order to recognize predefined Soundpainting gestures, and a system for generation of elec-

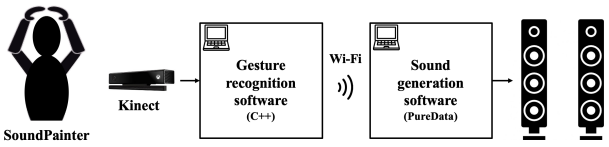


Figure 2: The overall architecture of the proposed system

tronic music sounds that will be played according to each recognized gesture.

Each system is executed on a different computer. We used a client-server architecture (UDP), where the gesture recognition system acts as a client and the music generation system as a server. The client sends an identifier according to the recognized gesture and the server receives this identifier to perform a defined action (the generation of a particular sound). The first line of images in Figure 3 shows the six Soundpainting gestures we considered in this study, extracted from [22, 23]. These gestures were selected because of their performance characteristics allowing them to be accurately recognized by a Kinect® sensor (high quality of data) and because they offer sufficient expressive power to the Soundpainter for its composition. A gesture is a sequence of key postures. To reduce the complexity and optimize the recognition process, we have selected for each gesture the most representative key position to identify the gesture. A gesture is therefore reduced for our study to a single key position. The key postures are illustrated on the first line of images in the figure 3.

3.1 Gesture recognition software

Our gesture recognition system is capable of analyzing and recognizing predefined gestures in two steps (figure 4): extraction of gesture features (e.g., articulation joint angle) and gesture classification performed by the Soundpainter. Through the use of a Kinect®, system input is a set of 3D Cartesian coordinates that are analyzed by the feature extraction module and classified by the classification module. This classification module uses the model trained by a learning module. The output of the system is the gesture recognized in the list of possible gestures. The proposed system is able to recognize the gestures in real-time (30 fps). The proposed approach does not work with multiple users in front of the Kinect. In the following sections, we describe each of the modules.

3.1.1 Feature extraction module

The Kinect® sensor is used to capture gestures performed by the Soundpainter. This sensor returns a vector with the Cartesian 3D coordinates of the 25 body joints with

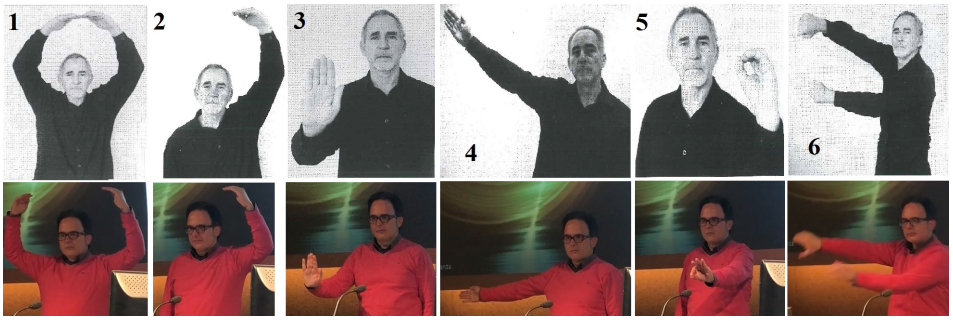


Figure 3: The 6 Soundpainting gestures performed during the live artistic demonstration: 1. Whole group, 2. Rest of the group, 3. Wait, 4. Scanning, 5. Silence, 6. Off. The first line shows the gestures made by the creator of the Soundpainting language (Walter Thompson), images extracted from [22, 23]. The second line shows the actions performed by the Soundpainter during the show/experiment.

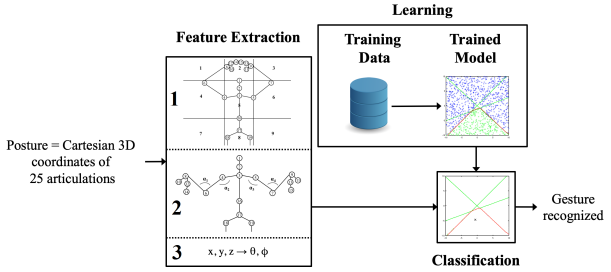


Figure 4: Architecture of gesture recognition software

a frequency of thirty frames per second. To recognize the posture of the captured person, we have created a morphology independent extraction module, while preserving a low computation time.

The first feature extracted is the **wrists positions**, since in the Soundpainting language, the wrists are the parts of the body that contain the most variations. Considering their positions as characteristics is therefore essential. We used the method of Chang et al. [1] which consists of dividing the space around the body in 9 quadrants defined by 4 lines (Figure 4; feature extraction).

The second extracted feature is the series of four **angles generated by the arms** (Figure 4; feature extraction). These angles have the advantage of being invariant to the changes of scale and the translation [24], thus independent of the morphology. In addition, they complete the information of the position of the wrists. Finally, to optimize the recognition, we do not consider the 8 joints numbered from 18 to 25, since the legs do not intervene in the considered gestures. We keep the articulations from 1 to 17. Moreover, according to [20], to simplify the process of normalization and to make the skeleton independent of the morphology and the size, we converted the Cartesian coordinates of the points of articulation in **spherical coordinates** in order to not consider the radial distance which depends on the morphology.

Each posture thus generates a vector of size 40: a quadrant for each wrist, 4 angles from the arms and 34 (2x17) spherical coordinates.

3.1.2 Learning module

In the learning module written in C++, we have used a machine learning library specially designed for real-time gesture recognition, the GRT library (*Gesture Recognition Toolkit*) [9]. In our implementation, we chose the Decision Tree algorithm. The main advantage of this algorithm is that the model is particularly fast at classifying new input samples. 9 Soundpainting gestures (key postures) were learned in the system: the six Soundpainting gestures used to create the musical composition (figure 3), a resting gesture (arms lie along the body) and two gestures of intermediate postures to avoid confusion in the recognition.

For the classification module, the Decision Tree (the trained model), built during the learning, makes it possible to decide in real time to which class of gestures belongs the input vector.

3.2 Sound generation software

For music generation, Pure Data software is used, the visual programming language developed by Miller Puckette [17] for interactive multimedia and music creation. The Pure Data software communicates with the gesture recognition software through a socket UDP Client-Server architecture. Once a Soundpainting gesture is recognized by the gesture recognition software, a gesture code is sent to the Pure Data software through the socket. This code is analyzed and the sound action associated with the recognized gesture is started. The following table indicates the Soundpainting gesture, the code of the gesture (received in the socket) and the corresponding sound action:

Gesture	Code	Actions
Whole Group	1	Range of six ascending and descending notes
Rest of the Group	2	Set of simultaneous joint sounds
Wait	3	Increase in volume
Scanning	4	Harmonic series of musical sounds
Silence	5	Decrease in volume
Off	6	Stop sounds

4. TWO INTERACTION SCENARIOS

In order to study the potential of the proposed gestural interaction, our system was first tested in a public artistic demonstration, and second in a learning context. For each of them we are interested in the performance of the proposed system and in the User experience of the Soundpainter.

4.1 Performance interaction scenario

The purpose of this demonstration was to evaluate two dimensions: the user experience of the Soundpainter in a real scenario of artistic composition and the accuracy of the Soundpainting gesture recognition system (described in the previous 3.1 section). During this demonstration the Soundpainter performed a musical composition in real time using the 6 gestures recognized by the proposed system. At the same time, two musicians sitting in the audience who knew the language Soundpainting, annotated manually (using an Android tablet application) the actions performed by

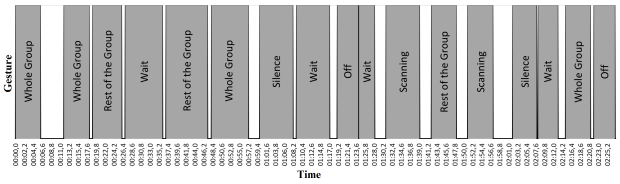


Figure 5: The sequence of gestures performed and their duration (ground truth)

the Soundpainter in order to compare their accuracy in the gesture recognition process with the accuracy given by our proposed system.

4.1.1 Participants

Three volunteers participated actively in the study. The first is the composer (Soundpainter) aged 44 years with 6 years experience as a Soundpainter. The other two participants are two musicians aged 52 years and 54 years respectively. Both play the piano and have 1 year and 3 months experience with the Soundpainting Language. Finally, there were thirty people in the audience (participants of the Symposium "Sonido y Action").

4.1.2 Procedure

The Soundpainter, located about 2m in front of the Kinect sensor, was tasked to perform, for a few minutes, a composition using the 6 gestures recognized by the system (figure 3). For a few seconds, the Soundpainter discovered the system by performing the 6 gestures sequentially. For each performed gesture, he had a feedback message indicating the gesture recognized (see image 4 in figure figure 1). These images were displayed in front of him, on the computer screen hosting the gesture recognition System. After this discovery step, the live demonstration began and the Soundpainter performed, with the 6 gestures, an improvised musical composition lasting 2 minutes and 25 seconds. For every gesture recognized, the system recorded in a log file, the name of the gesture and the precise moment (the minute and the second) where the gesture was recognized.

A video capture of the Soundpainter was recorded during the live demonstration. Throughout the demonstration, the two musicians, sitting in the audience, were instructed to note the gestures of the Soundpainter they recognized using an application installed on a tablet. After the demonstration, we gave the Soundpainter a user experience questionnaire ([18]) to evaluate how it felt about the system.

The hypotheses in this study were as follows: **(H1)** The quality and performance of the Soundpainting's gestural recognition system is similar to that of the Human. **(H2)** The proposed gestural interaction is applicable in an artistic context.

4.1.3 Results

For the performance of the proposed system, the figure 5 shows the sequence of gestures performed by the Soundpainter during the live demonstration/experiment. This sequence of gestures represents the ground truth because it was extracted from the video captured during the live demonstration. We find that 17 gestures were made in total with a duration of between 4 and 9 seconds. In order to evaluate the performance of our Soundpainting gesture recognition system, we calculated an F-measure. The F-measure is an indicator of a test's accuracy commonly used to evaluate machine learning algorithms [16]. It is calculated from the following equations:

$$P = \frac{VP}{VP + FP} \in [0, 1] \quad (1)$$

$$R = \frac{VP}{VP + FN} \in [0, 1] \quad (2)$$

$$F\text{-}measure = \frac{2 \times P \times R}{P + R} \in [0, 1] \quad (3)$$

where VP(True Positive) represents the number of gestures correctly recognized by the system, FP(False Positive) is the number of gestures wrongly recognized by the system (bad Recognition) and FN(False Negative) is the number of gestures not recognized by the system. The precision P indicates how often a gesture is recognized correctly. In our case, the reminder R is always equal to 1 because our system always associates a posture with a learned gesture class (see paragraph on the classification module).

During the live demonstration, for some gestures, the Soundpainter made several attempts (small variations of the gesture) so that his gesture was correctly recognized by the system. These attempts were also recorded by the log file. The *F-measure* as defined above does not measure these attempts. In order to penalize these attempts, we introduce a new measure, which we call F-measure penalized ($F\text{-}measure_p$) which consists of dividing the computed f-measure for each gesture by the number of attempts made for this gesture.

$$F\text{-}measure_p = \frac{F\text{-}measure}{number\ of\ attempts} \quad (4)$$

From the log files generated by the automatic recognition system, by the tablet application of the musician #1 and by the tablet application of the musician's #2, we computed the $F\text{-}measure_p$. We find that the gestural recognition system obtained an average $F\text{-}measure_p$ measure of 68.15%, the first and second musicians obtained a $F\text{-}measure_p$ of 100% and 88.24% respectively. Both $F\text{-}measure_p$ to 0.0% of the musician #2 are due to the fact that she forgot to indicate the gesture on the tablet. Figure 6 shows the evolution of the performance ($F\text{-}measure_p$) of the proposed system for each of the 6 gestures used during the musical performance.

The User experience of the Soundpainter was evaluated using the French version of the UEQ questionnaire (*User experience questionnaire* [18]). The scales in this questionnaire are designed to cover the complete and accurate impression of the User's Experience. The questionnaire format enables the user to immediately express his or her feelings, impressions and attitudes when using a product or System. The UEQ contains 26 elements that enables the evaluation of the following 6 scales: Attractiveness (overall impression of the product), Perspicuity (easiness to get familiar with the product), Efficiency (solving the task without unnecessary effort), Dependability (feeling in control of the interaction), Stimulation (motivation to use the product) and Novelty (innovation and creativity). Each scale varies from -3 to +3. Thus -3 represents the most negative response, 0 a neutral response and +3 the most positive response. Scale values greater than +1 indicate a positive impression of users for this scale, with values less than -1 a negative impression. The results of the Soundpainter user experience evaluation are shown in Figure 7.

4.2 Learning interaction scenario

A user evaluation was conducted to determine the potential of the proposed interaction for learning the Soundpainting

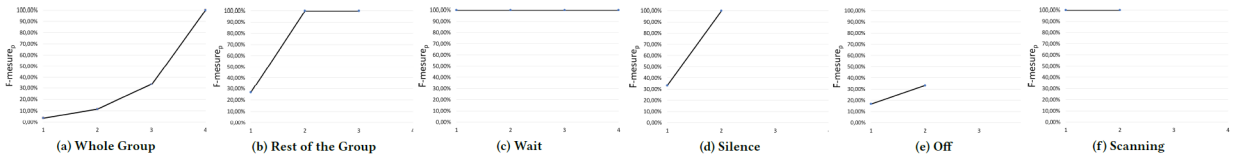


Figure 6: Performance ($F\text{-}measure_p$) of the proposed gesture recognition system by each of the 6 gestures. X-axis: number of uses of the gesture by the Soundpainter during the live demonstration.

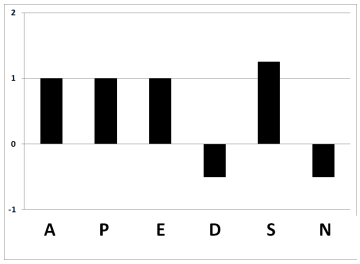


Figure 7: User experience of the Soundpainter. The labels are: Attractiveness (A), Perspicuity (P), Efficiency (E), Dependability (D), Stimulation (S) and Novelty (N).

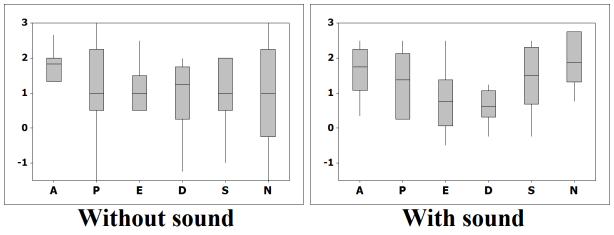


Figure 8: User experience (UEQ scores) in a learning context.

language and the influence of the sound generation during this learning process.

4.2.1 Participants

For this evaluation, art students from the University of the Basque Country (UPV/EHU). These students had few experience (from 1 day to 1 month) using the Soundpainting language and they did not know correctly how to perform each Soundpainting gesture. Thus, thirteen students took part in the study, aged from 21 to 38 (mean=23.3, sd=5.01), 6 males and 7 females.

4.2.2 Procedure

Participants were asked to improvise using the six trained Soundpainting gestures during 2 minutes and 25 seconds (similar to the conditions of the live performance described in the previous section). In order to study the influence of the sound generation two conditions were tested: improvisation without sound and improvisation with sound. Thus, for this experiment, 7 students improvised with no sound generation and 6 students improvised with sound generation. At the end of the experiment, each student filled the Spanish version of the UEQ questionnaire [18].

4.2.3 Results

For the performance of the proposed system we compute the measure ($F\text{-}measure_p$) of the gesture recognition system was computed for each gesture and each student. The average measure obtained by the students without sound condition was 0.49 (sd=0.14) while the average measure obtained by the students with sound condition was 0.50 (sd=0.08). Regarding the $F\text{-}measure_p$, the sample size estimated is 13 with a margin of error of 0.072 and a confidence level of 95%. The number of gestures performed by each student was also considered: 52 (sd=7.74) average number of gestures performed by the students without sound condition and 36 (sd=10.07) average number of gestures performed by the students with sound condition. Regarding the User Experience of the students, the UEQ results obtained from each condition are presented in figure 8.

5. CONCLUSIONS AND PERSPECTIVES

This multidisciplinary work in art and science has given rise to the realization of a functional prototype, robust, and fluid, as our in-situ experiments during a public performance have proved. We have thus lifted a technological lock by proposing a technique of automatic recognition of Soundpainting gestures from a single video stream and in real time. The first objective of this study was to assess the accuracy of the Soundpainting gesture recognition System. The performance results showed that the overall performance of the proposed recognition system is lower (68.15%) than the recognition made by the humans (100% and 88.24%). This result therefore rejects the hypothesis **H1**. However, we see in Figure 6 that the performance of gesture recognition (except if it is optimal from the beginning) consistently improves over time. The reason is that Soundpainting language is not limited to a precise form of realization of each gesture which is clearly visible on the images in figure 3 (difference in execution between W. Thompson and the Soundpainter in the live show). Some gestures were not properly recognized in their first executions since the Soundpainter was accustomed to performing it in a different way from the trained posture. However, given that the Soundpainter had a visual feedback (e.g. skeleton + recognized gesture message) of the recognition of his own gesture when he realized it, he was able to adapt (in live and in few seconds), during the performance, the realization of his gesture to be correctly recognized by the system. The second objective of this study was to evaluate user experience of Soundpainter in order to determine the relevance of our approach in an artistic context of musical composition. Results of the UEQ reported in Figure 7 show that Soundpainter has positively evaluated the proposed gestural interaction (positive Attractiveness). These results also indicate that Soundpainter has been able to quickly familiarize itself with the gesture recognition system (positive Perspicuity) by making its musical composition effortless (positive Efficiency). He also found the stimulating and motivating interaction (positive Stimulation). Thus, the **H2** hypothesis is accepted.

Regarding the user evaluation for students of Soundpainting language, results (Figure 8) showed that participants found more innovative (higher Novelty) and motivating (higher Stimulation) the process of learning the Soundpainting with

sound generation. Results suggest, however, that participants using the proposed system with sound generation felt lower control (lower Dependability) during the interaction compared to the participants that used the system with no sound generation. This may be due to the fact that participants with no sound were more engaged in the visual feedback message during the interaction because of the instantaneous response of this visual feedback, whereas the response time of the sound feedback was slower. In addition, the results obtained from the second study (section 4.2.3) showed that the accuracy of the proposed system was similar for both conditions (with sound and without sound). However the number of gestures performed by the students was larger without the sound condition. This may due to the fact that the students with sound condition had to wait until the sound ended before performing the next gesture. This suggests that the sound was not ignored by the students and that they really felt as if they were performing an artistic composition.

We are essentially looking at two technological perspectives in this work. First, that a gesture is no longer characterized by a key posture but by the complete movement of the gesture considering the temporality. This will make it possible to discriminate in a more precise way gestures that are similar. Second, it would be useful in order to offer more creativity to the composer, to increase the dictionary of gestures and to move from 6 gestures to 30 then to a hundred while using complementary sensors and by proposing a recognition based on the merging of multimodal information. Once these technological challenges are solved, we plan to set up a live show in which musicians and machines (robots, drones, synthesized sound generators as in this article) can collaborate in order to generate together, by improvisation, an original musical composition from the language of Soundpainting.

6. ACKNOWLEDGMENTS

The authors would like to thank Josu Rekalde from University of the Basque Country for developing the sound generation software. Special thanks also to Cristina Arriaga and Baikune de Alba (both from University of the Basque Country) for their support and collaboration in the study.

7. REFERENCES

[1] M.-S. Chang, J.-H. Chou, and C.-M. Wu. Establishing a natural hri system for mobile robot through human hand gestures. 9th IFAC, pages 9–12, 2009.

[2] C. Chen, K. Liu, and N. Kehtarnavaz. Real-time human action recognition based on depth motion maps. *J. Real-Time Image Process.*, 12(1):155–163, 2016.

[3] K. Christopher, J. He, R. Kapur, and A. Kapur. Kontrol: Hand gesture recognition for music and dance interaction. NIME'13, pages 267–270, 2013.

[4] E. Cippitelli, S. Gasparrini, E. Gambi, and S. Spinsante. A human activity recognition system using skeleton data from rgbd sensors. *Computational Intelligence and Neuroscience*, 2016, 2016.

[5] A. Clay, N. Couture, M. Desainte-Catherine, P. Vulliard, J. Larralde, and E. Decarsin. Movement to emotions to music: using whole body emotional expression as an interaction for electronic music generation. NIME'12, 2012.

[6] E. A. Edmonds. Human computer interaction, experience and art. In L. Candy and S. Ferguson, editors, *Interactive experience in the digital age: evaluating new art practice*, chapter 2, pages 11–23. Springer, London, 2014.

[7] D. England. *Whole Body Interaction*. Springer Publishing Company, Incorporated, 1st edition, 2011.

[8] J. Françoise and F. Bevilacqua. Motion-sound mapping through interaction: An approach to user-centered design of auditory feedback using machine learning. *ACM Trans. Interact. Intell. Syst.*, 8(2):16:1–16:30, 2018.

[9] N. Gillian and J. Paradiso. Grt - gesture recognition toolkit. `http://www.nickgillian.com/wiki/pmwiki.php/GRT/GestureRecognitionToolkit`, 2013. Online; accessed 2018-05-05.

[10] N. E. Gillian, R. B. Knapp, and M. S. O'Modhrain. Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping. In *NIME*, 2011.

[11] P. Guyot and T. Pellegrini. Vers la transcription automatique de gestes du soundpainting pour l'analyse de performances interactives. JIM'16, pages pp. 118–123, Albi, France, Mar. 2016.

[12] B. R. Knapp and H. S. Lusted. A Bioelectric Controller for Computer Music Applications. *Computer Music Journal*, 14(1):42–47, 1990.

[13] C. P. Martin, H. J. Gardner, and B. Swift. Tracking ensemble performance on touch-screens with gesture classification and transition matrices. In *NIME*, 2015.

[14] G. Paine. Gesture and musical interaction: Interactive engagement through dynamic morphology. NIME'04, pages 80–86. National University of Singapore, 2004.

[15] T. Pellegrini, P. Guyot, B. Angles, C. Mollaret, and C. Mangou. Towards soundpainting gesture recognition. AM'14, pages 1–6. ACM, 2014.

[16] D. M. W. Powers. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[17] M. S. Puckette. Pure data. In *ICMC'97*, 1997.

[18] M. Rauschenberger, M. Schrepp, M. PÃl'rez Cota, S. Olschner, and J. Thomaschewski. Efficient measurement of the user experience of interactive products. how to use the user experience questionnaire (ueq). example: Spanish language version. *International Journal of Artificial Intelligence and Interactive Multimedia*, 2(1):39–45, 2013.

[19] A. Sarasua, J. Urbano, and E. Gomez. Mapping by observation: Building a user-tailored conducting system from spontaneous movements. *Frontiers on Digital Humanities*, 2019.

[20] A. Taha, H. H. Zayed, M. Khalifa, and E.-S. M. El-Horbaty. Human activity recognition for surveillance applications. ICIT'15, pages 577–586, 2015.

[21] A. Tanaka. Musical performance practice on sensor-based instruments. In M. M. Wanderley and M. Battier, editors, *Trends in Gestural Control of Music*, Science et musique, pages 389–405. IRCAM - Centre Pompidou, 2000.

[22] W. Thompson. *Soundpainting: the art of live composition. Workbook 1*. 2006.

[23] W. Thompson. *Soundpainting: the art of live composition. Workbook 2*. 2009.

[24] S. Zhang, X. Liu, and J. Xiao. On geometric features for skeleton-based action recognition using multilayer lstm networks. WACV'17, pages 148–157, 2017.