The Synekine Project

Grégory Beller
IRCAM
1 Pl. I. Stravinsky
75004 Paris, France
beller@ircam.fr

ABSTRACT

The Synekine project involves performative and scientific researches for the creation of new ways to express ourselves. The neologism "Synekinesia" would reflect our capacity to associate two or several sense motors. In the Synekine project, the performers develop a fusional language involving vocal gesture, hand gesture and movement in the space. An interactive environment made of sensors and other Human-Computer Interfaces increases this language.

Categories and Subject Descriptors

J.5 [Arts And Humanities]: Performing arts (e.g., dance, music))

Keywords

Speech Processing, Gesture, Movement, Performance, Voice

1. INTRODUCTION

The Synekine project possesses several dimensions; a dimension of scientific research on the phenomena "synaesthesia", the proprioception, the signal processing, the temporal topology and the functioning of the brain; a dimension of performative researches involving performers, around the notion of "Synekinesia" (central notion of the project explained below); a dimension of technical researches and developments for the creation of new Human-Computer Interfaces (HCI); an artistic and social dimension with a continuous publication of the project via the social networks, a cycle of performances and a show bringing in all the previous dimensions.

The neologism "Synekinesia", builds itself from the Greek terms "syn", which means "union", and "kinesis" which means "movement". This construction is similar to that of the more known term "synaesthesia" containing "aesthesis" which means "sensation". The synaesthesia is a neurological phenomenon by which two or several perceptive senses are associated. For example, in a type of synaesthesia known under the name of "graphemes-colors" synaesthesia (which would represent 64,9 % of the synaesthesia), the letters of the alphabet or the numbers can be perceived colored. In another type of synaesthesia, called "digital synaesthesia", the numbers are automatically and systematically associated with positions in the space. In another type of synaesthesia, called "synaesthesia of linguistic personification", the numbers, days of the week, month of the year evoke personalities. In other types of synaesthesia, the music and other sounds can be perceived colored, or having a particular shape.

By analogy in this phenomenon, the "Synekinesia" would be our capacity to associate two or several sense motors. Among these sense motors, the project is particularly interested in the vocal gestures and in the manual gestures, and implies more widely the movement of the body in the space. Much evidence has been found for pervasive links between the manual and speech motor systems, including evidence from infant development, deictic pointing, and repetitive tapping and speaking tasks [1]. The faculty to speak with hands would not only arise from cultural origins, but would also result from a deep neuronal relationship linking the speech to the gestures of the hands [2]. Certain admitted cultural or interpersonal variations would result then from the degree of progressive reduction of the physical movement accompanying the speech, the reasons of which could be numerous, as the annihilation of the redundancy of information, or the reduction of an "animality" probably bound to the expression of the emotions.

2. PROJECT PRESENTATION

The Synekine project, proposes in a first part, "to express this animality", by leading a research work with peformers, around the elaboration of a fusional language implying vocal gesture, manual gesture and movement in the space. Close from the theater of the cruelty, of A. Artaud [3], this performative research consists of the elaboration of a so-called Synekinesic expressivity through every kind of gesture. Inevitably personal because subjective (in a similar way in the synesthesic experience), this expressivity is made accessible through the self-exploration of one proprioceptive faculties, by the way of basic theater exercises: for example, to maintain a complete synchronization between manual gesture and vocal gesture, practicing the silence by the immobility and producing the sound of every movement. As a second step, the same exploration in a group of individuals, with or without situation of interaction, allows the identification and the definition of fundamentals on which can be built a shared performative language. It should be highlighted that this work differs from researches led on the sign language, by the negation of any semiotic element. Indeed, it is not a work on the meaning of the language, but on its capacity of expression.

The second part of this project consists in bringing in Human Computer Interfaces (HCI). Interactive systems are developed in the idea to propose to the performers new synekinesic possibilities, widening their field of expression (as allows it the practice of a musical instrument, for example). Certain interfaces

MOCO '14, June 16 - 17 2014, Paris, France.

ACM 978-1-4503-2814-2/14/06...$15.00.
http://dx.doi.org/10.1145/2617995.2618007

can be also used by groups, in the idea to confer to the group, new ways of expression, based on its interaction.

The social and artistic dimension establishes the third part of this project, resulting from the belief that everyone might be interested is these experiences. Although the elaboration of this project requires experts' participation, the results can interest a wider public. That is why several forms are envisaged for the distribution and the communication of the project. First of all, an internet page[1] suggests following the temporality of the project, through playful videos of experiments. The use of the social networks allows every person interested in the project to subscribe and to participate themselves by supplying elements of research. Then, under the shape of isolated artistic performances, implying the performers, alone or by group, with interactive system or not. Finally, under the shape of a show that aggregates previous performances, implying several performers.

The following parts of this paper present various innovative HCI connecting voice to hand gestures and movement in the space. These interactive HCI define experimental scenarios for the stage or for installations for the general public. The gesture capture is mostly based on a combination of data coming from accelerometers and kinect.

3. GESTURE CAPTURE

3.1 Background

In the case of the acoustic instruments, the gesture is necessary for the sound production. This is not any more the case of the electronic instruments where the control of electronic sound processes is separated from the process of sound production. Several projects of research and creation in the IRCAM use the gesture capture. Whether it is to augment acoustic instruments, (Bogen Lied [4], Augmented Violin Project [5], augmented drums[2]) or to connect different worlds together such as music, video, gesture and dance for the creation (Glossopoeïa) or for the pedagogy [6], certain number of sensors was realized in the IRCAM. However, the gesture capture was neither used yet to control prosodic modifications nor vocal synthesizers in the spectrum of the contemporary creation in the IRCAM. It is thus a new way to use the gesture capture systems that we propose for this project of research and creation.

3.2 Accelerometers gloves

The technology of gloves wireless accelerometers / gyroscopes used [6] allows to measure the accelerations of both hands according to 6 axes (3 in translation and 3 in rotation with gyroscopes) (see figure 1). The raw data delivered by gloves are not necessarily easy to interpret. So a first stage of pre-processing / mapping allows to return more interpretable data.

The data resulting from the wifi receiver are transmitted via UDP all 1 ms. To synchronize them to the internal clock of Max / MSP, they are first median filtered (order 5) and sub-sampled by a factor 5. Thus we obtain a stable stream of data all 5 ms. Then various descriptors of the gesture arise from these pre-processed raw data: Variation of the Momentum, hit energy estimation and absolute angle of the hand.

The estimate of the immediate acceleration allows knowing, at any time, the variation of Momentum relative to the gesture. This Momentum, according to the laws of the classical mechanics, is directly proportional in the speed. The raw data coming from the sensor are at first "denoised" thanks to the average on the last 4 samples. The root of the sum of the square of these six filtered values allows obtaining a proportional quantity in the variation of Momentum of the gesture.

The hit energy estimation allows the immediate release from the observation of the variation of the Momentum of the gesture. Three values, delivered by the sensors of acceleration in translation, are stored in a circular buffer including all the time, 20 samples. Three standard deviations corresponding to these values are added, all the time, (norm I corresponding to the sum of the absolute values). This sum also allows representing the variation of Momentum of the gesture. To detect variation of this value, corresponding to abrupt variations of the gesture, it is compared all the time with its median value (order 5). When the difference between these two values exceeds certain arbitrary threshold, a discreet value appears to mean the presence of a fast change of the gesture. It allows, for example, emitting a regular click, when we beat a measure with the hand of bottom at the top, every time the hand changes direction.

The hit energy estimation is a process allowing generating discreet data from a gesture, by definition continuous. Indeed, of a continuous physical signal, it allows by thresholding, to define moments corresponding to the peaks of variation of the Momentum, which coincide, from a perceptive point of view for the user, in peaks of efforts (of acceleration). By this process, it becomes then possible to create precise air percussions either sounds activation at the moment when the hand of the user changes direction or accelerates surreptitiously.

It is rather difficult, even impossible, to obtain a slow, quasi-static control or to define an absolute position in the space, from accelerometers and from gyroscopes. However, the Earth's gravitational field introduces an offset into the answer of the sensors, which can be exploited to deduct the absolute position of the hands, as well as the presence of slow movements. This quasi-static measure brings a continuous controller to the performer.

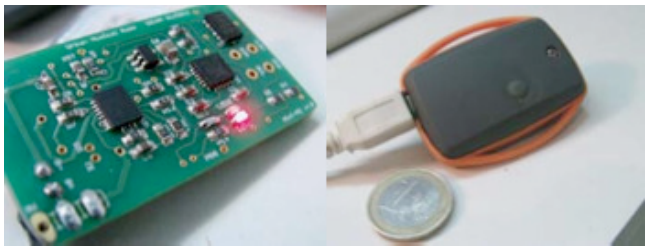

Figure 1. Accelerometers sensors gloves

3.3 Kinect

If accelerometers gloves provide a detailed description of the dynamics of gesture, they do not allow access to the absolute position of the hands. Indeed, a simple integration of the instantaneous velocities of the hands, does not allow the deduction of the absolute position of the latter as the signal / noise ratio of the sensor causes a drift in the estimation. So we cannot distinguish the same gesture performed under the basin or above the head. It is also difficult to define the magnitude of the gesture, by simple observation accelerometers.

The lack of absolute position seems to be confusing for performers who are used to define their actions with a combination of body posture and temporal dynamics.In order to address the lack of information on the absolute position of the hands, we decided to add to the accelerometer data, data from a

[1] http://www.gregbeller.com

[2] http://www.gregbeller.com/2014/02/augmented-drums

video capture. We chose to use a kinect with Synapse[3] software. This allows access, in real time, to the absolute positions of the joints of the body: hands, elbows, shoulders, head, neck, torso, hips, knees and feet. The combination of the accelerometers and the skeleton data, allows obtaining, in a synchronous manner, the position of the hands, their speed and their acceleration. In addition, data skeletons tell us about the postures adopted by the performers.

4. SPOKHANDS

In *Babil-on*[4] and in *Luna Park*[5], the percussionist Richard Dubelsky literally speaks with his hand gestures performing aerial percussions [7]. His hands are equipped with two movement sensors, developed by Emmanuel Flety [6], allowing him to trigger and to modulate various synthesizers in real-time. The generated sounds are the result of concatenating pre-recorded and pre-segmented sounds, rendered on the fly [8]. Instrumental sounds (flute) are randomly picked up whereas vocal sounds, such as syllables for example, are chosen wisely by a program following certain criteria [9]. Taking account these two fruitful experiences, the Synekine project wants to go further. Up to here, the main use of the sensors consisted in modulating or in activating prerecorded elements. Now, the Synekine project proposes to use the sensors to interact in real time with the voice, allowing the performer for exploration and improvisation. One fundamental improvement consists in triggering different sounds depending the position of the hands, using data from the kinect. The temporal accuracy of the hit energy estimation triggers different samples according to eight predefined zones (left / right and under basin / under shoulder / under head / above head). This makes spokhands a playable aerial drum machine.

5. HAND SAMPLING

One first experiment made in the Synekine project is to use hand gesture sensors to segment (to mark, or to sample) vocal gestures and to play sampled sound segments, simultaneously. Several paradigms of control are to be envisaged. For example, one hand can segment the voice, whereas the other one triggers immediately recorded segments. Interestingly, we found that it was more natural to segment the voice with the left-hand (for a right-hand writer) instead of using the right-hand, because to chop the voice with the right hand induces a more choppy voice than to chop it with the left-hand. A video[6] shows that paradigm in action. Another paradigm of control is to use a gesture implying both hands at the mean time to switch between recording mode and playing mode, and then to use one or the other hand to segment or to trigger voice samples.

6. HOLD THE TONG

Another development consists in manipulating the temporality of the amplified voice in real time with the hands. The pitch and the intensity are easily conceivable prosodic variables that could be modulated in real time, using transposition and gain [10, 11]. Then, how to alter the perceived speech rate in real time? We developed a new paradigm allowing the modification of the perceived speech rate, by the use of a phase vocoder applied on a ring buffer, the size of which can be arbitrarily defined. The use of a ring buffer allows for the instantaneity of the transformation, while authorizing, at any time, to move back in the short-term past (the duration of which is defined by the size of the ring buffer). Using this process, we can locally lengthen certain portions of the signal as the vowels (using a voicing parameter to define a vowel) and cause at the listener the perception of a slowing down of the speech rate. If we cannot move forward, the return in the immediate time can be made in an accelerated way, causing the perception of an acceleration of the speech rate. As if the reading head of the buffer behaved as a rubber band which we locally tighten and relax. Extremely, it is possible to freeze the position of the reading of the buffer in a place, what causes a "stop on sound" which can give interesting effects (to hold a vowel). For example, using hand gesture control, it is possible to catch vowels with a hand and to transpose it with the other hand. Another paradigm of control is to apply the gestural energy to the speed of the reading head in the buffer, causing a slowing down when the user quickly moves his hands.

7. WIRED GESTURES

Using a multimodal probabilistic model for gesture-based control of sound synthesis [12], wired gestures is designed to allow the performer to record several synchronous (called wired) gestures constituted of vocal gesture and hand gesture produced in the mean time. After a recording phase and a learning phase, the performer scrubs the sound of the vocal gesture by using the associated hand gesture. Recording of four wired gestures allows the performer to set up a gesture landscape where he can then improvise using only hand gestures. Interestingly, while the performer explores this finite gesture landscape, choreography that is strongly related to the produced music appears. A video has been made to show the potential of wired gestures[7]. The other way round is also possible. Using a video capture of the performer, this one can scrub the pre-recorded video of a wired gesture, using only his voice. This second paradigm associated to the first one; bring together a nice feedback effect that can be composed. The gesture of the performer produces a sound that moves the video that produces, in its turn, more or less the same sound.

8. HYPER BALL

HyperBall is a sound design tool controlled by the position of the hands. The main idea is to control basic effects such as granulation, filter, reverb or phase vocoder using parameters derived from skeleton data. The capture of the hand positions is realized here by the use of a kinect, but could also be made thanks to a Leap Motion. One can imagine a dematerialized spherical zone (the virtual ball), which establishes a zone of control of a sound material. Added to a microphone and a system of sampling in the voice, we can then speak in the hands, the sound is sampled, held, frozen or stretched, what establishes a continuous sound. From there, various gestures as to turn hands around the virtual ball, to enlarge it or to concentrate it or to prick it with the fingers, cause the modulation of several relevant sound effects. An example using granular synthesis is shown by a video[8].

[3] http://synapsekinect.tumblr.com

[4] Babil-on. 2013. Composed by Grégory Beller. Performed by Richard Dubelski. World Premiere: CMMR2013, October 2013

[5] Luna Park. 2013. Composed by Georges Aperghis. World Premiere: CMMR2013, October 2013

[6] http://www.gregbeller.com/2014/02/hand-sampling

[7] http://www.gregbeller.com/2014/02/wired-gestures

[8] http://www.gregbeller.com/2014/03/hyper-ball

Figure 2. Performers playing with Hyper Ball

9. VIRTUAL CHOIR

Certain singers of Sardinian polyphonies give some indications about the harmony to the others by the gesture. Virtual choir originates from the idea to control a synthesized choir, produced from the voice of the performer, by the gesture. It allows the performer for leading alone a complete choir that can be synchronous or a-synchronous compared to his voice. In a previous installation[9], we designed such a virtual choir, but with a rudimentary way of controlling it. One potentiometer changes the major/minor aspect of a fixed tonality, a second button changes the acoustics (different reverberations) and two others modify the number of male singers and of female singers. In Babil-on[10], a more evaluated control has been designed using hand gesture sensors. While singing, the performer controls the tonality in one hand (mapped to the cycle of fifths), and the mode of the scale (major/minor) with the other hand. Even if it is a clever mapping, it still does not allow for an intuitive control of the harmony. One part of the Synekine project is to improve this mapping with singers to find a sweet spot to improvise with. To start, a simple mapping is to consider only two virtual voices that are transposed according to the gestures of the two hands of the singer.

10. CONCLUSION

To connect the gesture with the voice seems to be an intuitive faculty of the human expression. However, in our modern cultures, it seems that this relation is rather reduced than developed. On one hand, the Synekine project suggests going to the reverse way, by submitting to performers, the exploration of this relation. To help them, new Human-Computer Interfaces, connecting the gesture with the voice, by the use of sensors and microphones, are developed. By the constraints that they bring, the use of these technologies urges the performers to reconsider, to discover, and to work various links between their vocal gestures and their manual gestures. On the other hand, a wider public as new way of interacting with sounds can then use the developments made to augment their expressivity. For instance, a hand gesture sound design system is a good candidate as a new HCI. The writing of this article is preliminary to the project, so it cannot bring elements other than forward-looking yet. Among the forward-looking elements, the exploration in a group will allow maybe the definition of common gestures or fundamentals in our capacity of expression. A webpage[11] suggests following the temporality of the project, through playful videos of experiments. The use of the social networks allows every person interested in the project to subscribe and to participate themselves by supplying elements of research. The creation of a show involving the performers of the project is envisaged, to show our researches to the general public.

[9] Exhibition “La voix, l’expo qui vous parle”, cite des sciences et de l’industrie, Paris, France, 2013

[10] Babil-on. 2013. Composed by Grégory Beller. Performed by Richard Dubelski. World Premiere: CMMR2013, October 2013

11. ACKNOWLEDGMENTS

Author would like to kindly thank partners of the project: IRCAM, Scene 44, Ubris Studio, LMA, GMEM; people who made it possible: N+N Corsino, G. Marchési, J. Diennet, C. Béros, A. Cont, E. Flety, J. Henrot, P. Bondu, F. Bévilacqua, N. Schnell, R. Borghesi and J. Françoise. And performers that were implied so far: Richard Dubleski, Stéfany Ganachaud, Jean-Charles Gaume, Dalila Kathir, Jean-Pierre Drouet, Lenny Barouk and Martin Seigneur.

REFERENCES

[1] Benjamin Parrell, Louis Goldstein, Sungbok Lee, Dani Byrd. 2014. Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics*, Volume 42, Pages 1-11

[2] Jana M. Iverson & Susan Goldin-Meadow. 1998. Why people gesture when they speak, *Nature* 396, 228

[3] Antonin Artaud. 1938. Le théâtre et son double. Gallimard, Paris.

[4] S. Lemouton. 2009. Utilisation musicale de dispositifs de captation du mouvement de l'archet dans quelques oeuvres récentes. In *JIM 2009*.

[5] F. Bevilacqua, N. Rasamimanana, E. Fléty, S. Lemouton, F. Baschet. 2006. The augmented violin project: research, composition and performance report. In *NIME 2006*.

[6] E. Fléty, C. Maestracci. 2011. Latency improvement in sensor wireless transmission using IEEE 802.15.4, in *NIME 2011*.

[7] Grégory Beller. 2011. Gestural Control of Real-Time Concatenative Synthesis in Luna Park. *P3S*.

[8] Grégory Beller. 2011. Gestural Control of Real Time Speech Synthesis in LunaPark, *SMC*.

[9] Grégory Beller. 2011. Gestural Control of Real Time Concatenative Synthesis, *ICPhS*.

[10] Grégory Beller. 2009. Analyse et Modèle génératif de l’expressivité : Application à la parole et à l’interprétation musicale. PhD thesis, Université Paris XI, IRCAM.

[11] Grégory Beller. 2009. Transformation of expressivity in speech. *Linguistic Insights*, 97:259–284.

[12] Jules Françoise, Norbert Schnell, and Frédéric Bevilacqua. 2013. A multimodal probabilistic model for gesture-based control of sound synthesis. In *Proceedings of the 21st ACM international conference on Multimedia* (MM '13), 705-70

[11] http://www.gregbeller.com