Visualizing Gestures in the Control of a Digital Musical Instrument

Olivier Perrotin
LIMSI-CNRS
BP 133 - F91403
Orsay, France
olivier.perrotin@limsi.fr

Christophe d'Alessandro
LIMSI-CNRS
BP 133 - F91403
Orsay, France
cda@limsi.fr

ABSTRACT

Conceiving digital musical instruments might be challenging in terms of spectator accessibility. Depending on the interface and the complexity of the software used as a transition between the controller and sound, a musician performance can be totally opaque for the audience and lose part of its interest. This paper examines the possibility of adding a visual feedback to help the public understanding, and add expressivity to the performance. It explores the various mapping organizations between controller and sound, giving different spaces of representation for the visual feedback. It can be either an amplification of the controller parameters, or a representation of the related musical parameters. Different examples of visualization are presented and evaluated, taking the *Cantor Digitalis* as a support. It appears the representation of musical parameters, little used compared to the representation of controllers, received a good opinion from the audience, highlighting the musical intention of the performers.

Keywords

Visualization, DMI, Performance, Gesture

1. INTRODUCTION

Two aspects are essential in an artistic performance when using a computer: the manipulation of the interface along with complementary gestures completed by the performer, and the effects ensuing this manipulation. Depending on which elements are shown or hidden to the spectators, a taxonomy has been established to describe a performance [11]. Whereas a magical performance with a hidden manipulation could sometimes be exciting, an expressive performance with revealed manipulation is often preferred in a musical context. Indeed, the decreasing intuitiveness of the relationship between one controller and its auditory effects may reduce the performance interest. Introducing visual feedbacks in the mapping process could benefit both to the performer and the spectator to better identify the functioning of the DMI. The experimentation of visual feedbacks on a virtual choir is presented in this paper.

Various visualizations have been implemented in the literature, at different locations of the mapping chain. Firstly, one of the most widely used type of visualization is displaying the controller of the interface. By filming the performer gestures or displaying the panel of controllers, it amplifies the manipulation but does not add any information. To that end, some choose to represent the controllers parameters. An example is the drawing on a graphical tablet used to control a sequencer [12].

These visualizations enhance the understanding on the performer's actions, but does not relate it to the sound production. Deeper in the mapping chain, a representation of intermediate parameters shows the musical intention of the performer and facilitate the understanding of sound production. The *Voicer* [8] visualizes the related-to-gesture parameters to help the performer himself controlling the instrument correctly. In case of complex mappings, the latter can be visually simplified representing direct connections between controllers and sound components to show the effect of the user interactions [3]. Some DMIs provide an interactive visual representation, where the controllers, the sound components and the link between them are represented at the same time [7].

These different representations help to understand mostly the interaction of the user with the interface, mainly focused on the control and gesture parameters. However, few visualizations have been done highlighting the musical parameters or the sound features (related-to-sound parameters) although it helps to focus on some features (timbre, pitch, loudness) [10]. Moreover, a whole sequence can be displayed entirely [2] keeping tracks of the previous moments. These examples of sound visualizations underline the importance of the representation of the musical parameters or sound features. However they have been realized in a context of a speech production, and were not applied during performances with DMIs.

This paper aims at using the *Chorus Digitalis* [5] as a support to explore the possibilities of visual feedbacks depending on their location on the mapping chain. This DMI takes advantage of strong analogies between drawing and singing [4] for making music. Hand gestures and vocal melodic motions seems to have comparable kinematics. In the present work, another aspect of this analogy is studied, i.e. hand gestures and visual patterns drawing, or visualization of hand motions for displaying musical motions. The latter have been proved efficient in audience understanding and have been few represented in performances with DMIs. Each visualization is evaluated on two aspects: the extra information brought to the audience, and its aesthetic quality. Section 2 presents the detailed mapping chain of the *Cantor Digitalis*. Several visuals are explored for each mapping layer in section 3. Discussions and conclusions are presented in section 4.

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

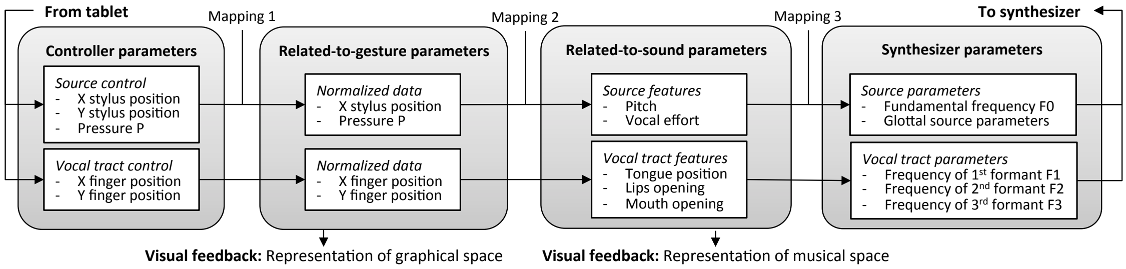

Figure 1: Evolution of gesture parameters through the different mapping layers of the Cantor Digitalis to control the glottal source and the vocal tract of a voice synthesizer with a graphic tablet.

2. CANTOR DIGITALIS

2.1 Cantor Digitalis and Chorus Digitalis

Voice synthesis has been improved considerably for the past 30 years, and has become an attractive medium for the conception of DMIs. Different interfaces of control were used such as mouse and keyboard, smart gloves, or graphic tablets. The latter proposes a wide range of possible gestures (stylus position, stylus tilt, stylus pressure, finger touch position, ...) and takes advantage of the expert gesture of writing or drawing. This make this interface both rich in terms of possibilities and easy to use, without the need of a consequent learning process. It has been used extensively in the past decades for controlling DMIs.

The instrument used in our study is the *Cantor Digitalis* [5]. It is based on a source+filter model, where the source models the vibration of the vocal folds using the RT-CALM model. The filters stand for the resonances of the vocal tract, modeled with parallel filters. Source and filter are modeled and controlled separately both with a graphic tablet. A Wacom Intuos® 5 touch tablet[1] is used, which detects both the stylus and the fingers on the surface of the tablet. The synthesizer is implemented with the Max real-time programming environment[2], and the patch *s2m.wacom*[3] is used to get the tablet data. Two hands are used to play with the tablet. The preferred hand holds the stylus, and the non-preferred hand touches directly the surface of the tablet. The stylus is used to control the glottal source: the melody (pitch) and the vocal effort. The fingers are used to control the articulation: the vowels. Details of the mapping are presented in section 2.2.

To test intensively this instrument, an ensemble of *Cantor Digitalis* was initiated, named the *Chorus Digitalis* [9]. It is a virtual choir where each of the 4 to 8 musician controls a single voice with a graphic tablet. Created in 2010, it had the opportunity to give around ten performances until now. The stage organization evolved to musicians seated behind tables to musicians seated on the floor with their tablet lying on a 20 cm-high support in front of them. However the surface of the tablet remains small to be observed from a certain distance. Then a visual feedback was added to make the gestures even more accessible, on a computer screen in front of the musicians, and projected on a wide screen behind the musicians for the audience.

2.2 Cantor Digitalis organization

A relevant mapping for a digital musical instrument has to contain successive layers, decoupling the gestural control on one side, and the sound production on the other side [1], [6]. This leads to a more complex control of the instrument which provides more expressivity and playing possibilities. The *Cantor Digitalis* has been built following this principle. A description of the mapping chain is detailed, to apply examples of visual feedbacks on different types of parameters. Figure 1 displays the different parameters of the *Cantor Digitalis* and how they are related through the different mappings. Four blocs are visible: the *controller parameters* measured by the graphic tablet; the *related-to-gesture parameters*, which are high level values representative of the user gestures; *related-to-sound parameters* which are high level musical values to be controlled, and the *synthesizer parameters*, which are low-level values to feed the synthesizer. Two chains are observed, the upper chain concerns the parameters related to the glottal source, and the lower chain concerns the vocal tract. The glottal source is controlled by the stylus on the tablet, using its X and Y position, and its pressure. After a first step of normalization, they are mapped to the pitch and the vocal effort which are the two variables controlled by the source. Then these values are mapped to the synthesizer parameters which defines the properties of the vocal fold vibration: the fundamental frequency, and other parameters detailed in [5]. Simultaneously, the vocal tract is controlled by a finger touching the tablet in a small area in the top left corner. The X and Y position of the finger are measured. They are normalized to remove the effect of the area on the tablet. In voice production, the three main movements of the mouth to produce the buccal french vowels are the mouth opening, the tongue position, and the lips opening. Therefore, the normalized position of the finger is mapped to these three high level parameters for articulation. Then it appears these three parameters are highly correlated with the center frequency of the three first resonators of the vocal tract called formants. A mapping is applied between them.

3. VISUALIZING THE CANTOR DIGITALIS

From the point of view of the controller, the stylus gestures on a graphic tablet are easily assimilated to a drawing or writing. In the musical space, including the pitch and vocal effort, the gestures have a musical meaning. These parameters are essential components of a musical performance. Therefore it is possible to represent the gestures either in a graphical space taking the related-to-gesture values, or in a musical space taking the related-to-sound values. The visual feedback consists in the visualization in real-time of our gestures on the tablet. Including the time variable T, each gestural representation (graphical and musical space) is characterized by 4 dimensions. Although very complete and sometimes aesthetic, a visual four-dimension representation is heavy both visually and in terms of computation. This is why we focus on simpler representations, keeping up to 3 variables each time.

[1] http://www.wacom.com/en/gb/creative/intuos-pro-m
[2] http://cycling74.com
[3] http://metason.cnrs-mrs.fr/Resultats/MaxMSP/

3.1 Representation in the graphical space

The simplest way to represent the musician's gestures in the graphical space is displaying a shape (a dot, a circle, a stain) on the screen corresponding to the state of the stylus (resp. finger): the normalized X and Y position of the stylus (resp. finger) are mapped to the X and Y position of the shape on the screen, and the normalized stylus pressure is mapped to the size of the shape (resp. a unique size for the finger). This is similar to freehand painting applications. Depending on the control, several context are imagined.

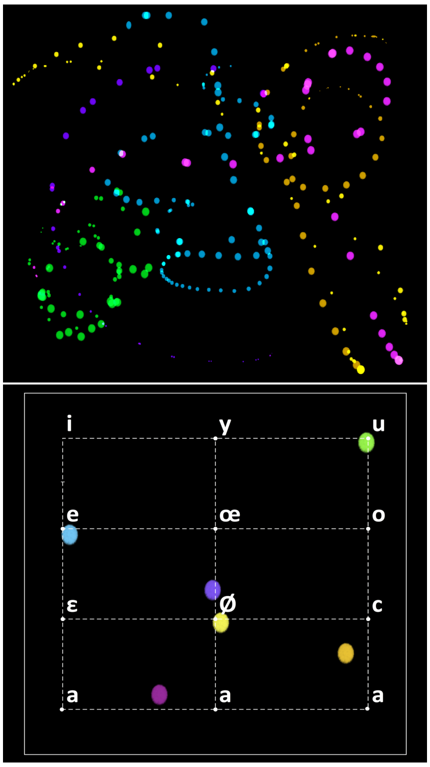

Figure 2: Visual feedbacks of the related-to-gesture parameters. The source control is represented on the top (state of the stylus), and the vocal tract control on the bottom (state of the finger)

3.1.1 *Source*

The stylus is used to control the source. The X-position is essential because it controls directly the pitch. The Y-position is let free to the user, as a mean of expressivity. Therefore, the performer can either draw straight horizontal lines to target pitch values, or draw different contours with more natural movements. The top of figure 2 shows an example of such a visualization where each color corresponds to a musician. For the audience, this representation of the gesture is in some way an amplification of the tablet gesture. However, contrary to a simple camera placed above the tablet, this adds some information not visible on the tablet, such as the overlapping of the musician trajectories. It is a calligraphic metaphor which reflects the musicians ability to draw shapes in a space. This gives a multimodal mean of expressivity, mixing drawing or painting and music.

3.1.2 *Vocal tract*

The control of the vocal tract to produce vowels is done with two parameters, the X and Y position of the finger on the tablet. The representation of the normalized finger position on the screen is inspired from the vocalic triangle representing the vowels in a 2D plan: tongue position vs mouth opening. The lips opening variable is projected on this plan. This representation is also closely related to the related-to-sound and synthesizer's parameters. The visual representation consists in displaying this vocalic triangle with the position of exact vowels, and the current articulation by a disc associated to each musician on this plan, either matching a specific vowel or in intermediary positions. Unlike the pitch trajectories, the articulation tends to vary less in time. The most important information is the position of the finger, related to a particular vowel. The bottom of figure 2 shows an example of visualization. Each color represents a musician.

3.2 Representation in the musical space

The graphical space visualization only brings forward the gesture on the tablet and is not related with the musical production. To help the user understanding how the control of the sound is made, i.e. to highlight which sound features are driven by the performer, it is also possible to display the musical or related-to-sound parameters.

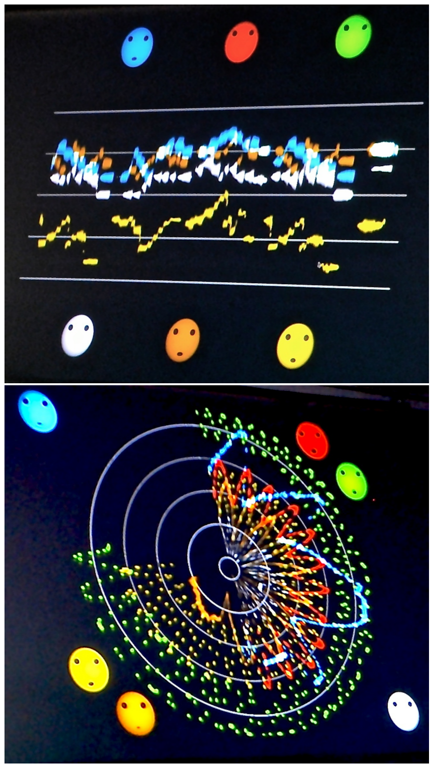

Figure 3: Visual feedbacks of the related-to-sound parameters. The lines represent the pitch versus the time, linearly on the top, and circularly on the bottom. The avatars represent the articulation control. Each color corresponds to a musician.

3.2.1 *Source*

In the *Cantor Digitalis*, the source is controlled among others with the pitch and the vocal effort. Moreover, although in a drawing context the time parameters have less importance, time governs music. This is why we choose to represent the pitch versus the time on the visual feedback. The current time is fixed to a position on the screen, so as the representation scrolls giving a continuous pitch trace, one for each musician. The vocal effort is mapped to the width of the trace. The time can be either represented linearly, scrolling from right to left as shown on the top of figure 3, or circularly, imitating the time on a clock [2] and scrolling clockwise. It is represented on the bottom of figure 3. To help the audience to identify who is actually playing and what they are playing, the color of the pitch trace is associated to the color of the musician's shirt.

3.2.2 Vocal tract

While the vocalic triangle is related to signal analysis, the related-to-sound parameters are more meaningful as for the articulation. They represent the implicit physical gesture one have to make to produce a vowel with its real voice. The representation of these parameters is a physical metaphor of the mouth opening, tongue position and lips opening using avatars. For each musician, spheres are used to represent head and eyes, and ellipses for the mouth and the tongue. This gives a simple avatar articulating the vowels played by the user. These avatars are visible on figure 3. The colors of the avatars are also associated to the musician shirts, to keep coherence with the pitch representation.

4. DISCUSSIONS AND CONCLUSION

To evaluate the relevance of our visual representations as for the enlightenment of the instrument control, questionnaires were given during a performance of the *Chorus Digitalis*[4]. The first 15min were played without visual, then the linear and circular visuals of the related-to-sound parameters were displayed alternatively during the last 30 min. All the questions had to be rated with a note between 1 (not at all) and 5 (clearly). The questions were: Did you understand how the voices were controlled; Could you distinguished who was controlling each voice before projection; Could you distinguished who was controlling each voice after projection. 18 persons took time to answer the questionnaire. The average age was 38 years. Only 4 persons had already been to a *Chorus Digitalis* performance, although the visual projection was new to this concert. 8 subjects were musicians, 2 of whom were singers and 5 were familiar with voice synthesis. The main impact brought by the visual feedback was the identification of the musicians. Two thirds of the audience could not distinguish decently the different players before the visual was applied, and 44% could not at all. Nevertheless, when applying the visual, all but two subjects could differentiate decently the players. It must be noted that the relationship between the trace color and the musicians color were not told before the show.

Our usual use of visualization is to add extra information and expressivity to a musical content. However, one can imagine a reverse process with a performance based on "drawing" on the visual feedback, whereas the music is added afterwards to add expressivity. It is all the more constraining so as the performance has to remain musically meaningful. Thus the visual composition needs to be created thinking of the musical rendering. An example of visual composition was experimented during the same performance. The main feedback was that it was too short, that the performers should have played more with the visual. This shows the interest of the public in this new way of creating music. Finally, 4 subjects found the visual the most noteworthy part of the performance, while the main purpose was the singing synthesis.

To conclude, the visualization of mapping parameters aims at improving the audience understanding of a digital musical instrument during a performance. The decomposition of the mapping chain of a DMI enables to target spaces of representation: the controller space, and the musical space. We mainly focused on the musical space, as it is little represented compared to the controller space. Visual representation was designed using the *Cantor Digitalis* and the ensemble *Chorus Digitalis* as a support. Several examples of representation both in graphical space (controller space) and musical space were given. The representation of musical space was well received by the audience and proved helpful the understanding of the performance.

To complete the visualization of parameters at every step of the mapping chain, the representation of synthesizer parameters such as formant frequencies for instance might be explored in future work. Moreover, a more dedicated platform of programming for graphics representation could be preferred to extend the aesthetic possibility of the visual.

[4]http://www.youtube.com/watch?v=9XpnDiJJyMk

5. REFERENCES

[1] D. Arfib, J.-M. Couturier, L. Kessous, and V. Verfaille. Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organized Sound*, 7(2):127–144, 2002.

[2] T. Bergstrom and K. Karahalios. Conversation clock: Visualizing audio patterns in co-located groups. In *Hawaii International Conference on Systems Science (HICSS)*, page 78, January 3-6 2007.

[3] F. Berthaut, M. T. Marshall, S. Subramanian, and M. Hachet. Rouages: Revealing the mechanisms of digital musical instruments to the audience. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2013.

[4] C. d'Alessandro, L. Feugère, L. B. Sylvain, O. Perrotin, and L. Feugère. Drawing melodies: Evaluation of chironomic singing synthesis. *Acoustical Society of America*, In print, 2014.

[5] L. Feugère. *Synthèse par règles de la voix chantée contrôlée par le geste et applications musicales*. PhD thesis, Université Paris VI, September 26 2013.

[6] A. Hunt and M. M. Wanderley. Mapping performer parameters to synthesis engines. *Organized Sound*, 7(2):97–108, 2002.

[7] S. Jordà. Interactive music systems for everyone exploring visual feedback as a way for creating more intuitive, efficient and learnable instruments. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC)*, August 6-9 2003.

[8] L. Kessous. Gestural control of singing voice, a musical instrument. In *Proceedings of Sound and Music Computing*, October 20-22 2004.

[9] S. Le Beux, L. Feugère, and C. d'Alessandro. Chorus digitalis: Experiments in chironomic choir singing. In *Proceedings of Interspeech*, August 28-31 2011.

[10] M. Pietrowicz and K. Karahalios. Sonic shapes: Visualizing vocal expression. In *Proceedings of the International Conference on Auditory Display*, pages 157–164, July 6-10 2013.

[11] S. Reeves, S. Benford, C. O'Malley, and M. Fraser. Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 741–750, New York, NY, USA, 2005. ACM.

[12] M. Zadel and G. Scavone. Different strokes: A prototype software system for laptop performance and improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, NIME '06, pages 168–171, Paris, France, 2006. IRCAM; Centre Pompidou.