

## SONIFICATION OF MUSICIANS' ANCILLARY GESTURES

Vincent Verfaillie, Oswald Quek, Marcelo M. Wanderley

IDMIL, Music Technology Area – CIRMMT  
Schulich School of Music – McGill University  
555 Sherbrooke Street West  
Montréal, Québec, Canada H3A 1E3

### ABSTRACT

This paper describes the sonification of movements of three clarinetists<sup>7</sup>. Rather than quantifying the different kinds of movements and presenting such information using visual methods such as graphs or tables, sonification of such gestures provides a complementary way of analysing movements which is possibly more informative than visualisation of data. This paper describes the design methodologies, mappings, and synthesis techniques used in transforming a set of data markers each with  $x$ ,  $y$  and  $z$  cartesian coordinates into informative and intelligible sonifications.

### 1. INTRODUCTION

In this paper we are interested in using sonification as a complementary tool for analysing musicians gestures. Expressive movements of musicians<sup>1</sup> is being investigated with the goal of better understanding music performance. Through sonification, we hope to provide extra tools for analysing musicians ancillary gestures, *i.e.* gestures that are not directly related to sound production<sup>2</sup>.

We consider sonification as “*the use of nonspeech audio to convey information*”, or more specifically, “*the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation*” [1]. Among its many applications, sonification has been used for enhancing perception accuracy of jumps [2]. Interactive sonification has been used to enhance sonification by introducing its gestural control [3], for physiotherapy movement analysis [4], and for motion analysis in video games [5]. This however differs from our study as our first goal is not to gesturally control sound synthesis [6] or sonification [3], but to study musician’s gestures using sonification.

There are many possible gestures of musicians to analyse, representing a large dimensional data set. Our rationale for sonifying musicians gestures include the following:

1. Sonification is ideal for representing data sets with large numbers of changing variables and temporally complex information that, when changing fast, may be blurred or missed by visual displays but easily detectable if sonified [1].
2. While a visual presentation of data requires a particular orientation of the user, sonification frees the cognitive load of the listener and enables to focus on important aspects of the data, his/her eyes being already busy with another task [1].

<sup>1</sup>See <http://www.music.mcgill.ca/musictech/clarinet/>

<sup>2</sup>For clarinet players, lip and finger motions are effective gestures, whereas weight transfer and bell circular motions are ancillary gestures.

3. Sonification helps to reveal structures in data that are not at all obvious in the traditional visual-only analysis [4].

We discuss the sonification of musicians ancillary gestures: the choice of gestures and related sound synthesis techniques, and the mapping techniques to make the sonification more efficient. We also present some first results, further research directions and conclusions.

### 2. TOOLS AND SET UP

We used recorded musical performances using video cameras and a high-accuracy Optotrak system 3020 [7], an optical movement tracker with active infra-red markers to track musicians’ movements. While videos gave a holistic picture of the performance, the Optotrak system allows us to visualize in detail various markers (see Fig. 1) on the musicians’ body using *Matlab*. Three per-

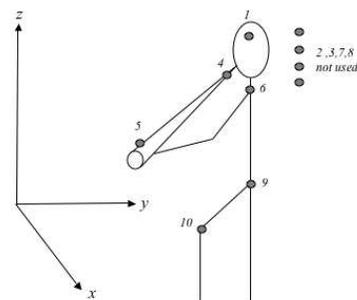


Figure 1: Optotrak marker placement, and the  $x$ ,  $y$ ,  $z$  directions.

formers played several times Stravinsky’s *Three Pieces for Solo Clarinet* with three expressive manners: normal, immobile and exaggerated [8]. Sonifications are in semi real-time, *i.e.* offline pre-processing is done using *Matlab*, and sonifications are played in realtime under *MaxMSP*.

In order not to overload the listener with too many data sets, we decided to sonify 4 ancillary gestures. A first thorough analysis of the videos using Laban-Bartenieff provided a list of ancillary gestures [9], and helped to choose the 4 following ones: the circular movement of clarinet bell, the body weight transfer, the body curvature, and the knee bend of a musician. Each synthesis technique must have unique features that makes it identifiable from the rest, so as to be able to simultaneously hear 4 different gestures. We chose Risset’s infinite glissandi sound [10] to sonify circular

Gesture Type	Markers	Synthesis Technique
Clarinet bell circular movements	5	Risset's infinite glissandi (pitch)
Body weight transfer	1, 7	Additive synthesis (tremolo rate)
Body curvature	1, 6, 9	Frequency modulation (brightness)
Knee bending	9, 10	Low-pass filtering (brightness)

Table 1: Summary of sonified gestures, affected markers, synthesis techniques and specific parameter mapped.

motions of the bell. Body weight transfer is sonified using a beat interference technique based on additive synthesis. The body curvature reflecting how ‘opened’, ‘straight’ or ‘rounded’ the body is, is mapped to brightness of frequency modulation sounds [11]. Knee bending was sonified using white noise filtering, and mapped to the filter cut-off frequency (and so forth to brightness). More details about mapping are provided in sec. 3. Table 1 summarize the selected gestures, the affected markers and the synthesis technique for each gesture.

### 3. SONIFICATION STRATEGIES

In the following subsections we discuss the mapping strategies and synthesis techniques for each sonification.

#### 3.1. Clarinet Bell Circular Movement

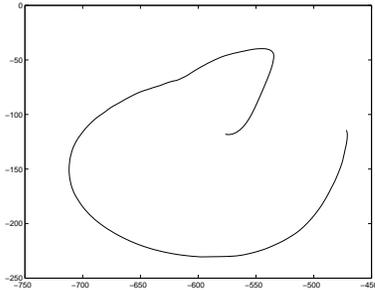


Figure 2: Typical circular motion of a clarinet bell,  $(x, y)$  plan.

Previous studies showed that circular bell motions gestures are important ancillary gestures affecting sound production [6]. A typical example of circular motion is depicted in Fig. 2. Many variations of such movement are possible, for instance arcs, ellipses, and small diameter circles.

A current mean position  $M(t)$  with coordinates  $(\bar{x}, \bar{y})$  is computed using the  $N$  last positions  $P(t-i)$  with coordinates  $(x(t-i), y(t-i)), i = 0, \dots, N-1$  of the current sample (at time  $t$ ) and previous samples. We then compute the 2-norm distances  $d_N(t) = \sqrt{(x(t) - \bar{x})^2 + (y(t) - \bar{y})^2}$  between this mean and the last point. We also compute the angle (clockwise direction) between the vertical line and the vector going from the current mean to the current position using trigonometric functions. For  $\bar{x} \neq x(t)$

and  $\bar{y} \neq y(t)$  is simplified as

$$\theta_N(t) = \frac{\pi}{2} i_1 + \pi i_2 + \tan^{-1} \left| \frac{\bar{y} - x(t)}{\bar{y} - y(t)} \right| \quad (1)$$

with the two indicators  $i_1$  and  $i_2$  defined as

$$i_1 = \text{indic}\{(\bar{x} - x(t))(\bar{y} - y(t)) < 0\} \quad (2)$$

$$i_2 = \text{indic}\{\bar{x} - x(t) > 0\} \quad (3)$$

Otherwise,  $\theta_N(t) = 0$  for  $\bar{x} = x(t)$  and  $\bar{y} \leq y(t)$ ,  $\theta_N(t) = \pi$  for  $\bar{x} = x(t)$  and  $\bar{y} > y(t)$ ,  $\theta_N(t) = \pi/2$  for  $\bar{x} < x(t)$  and  $\bar{y} = y(t)$  and  $\theta_N(t) = 3\pi/2$  for  $\bar{x} > x(t)$  and  $\bar{y} = y(t)$ . Optimal value for  $N$  is around 200 when calculating  $d_N(t)$  and 60 for  $\theta_N(t)$ .

Risset's infinite glissandi sound [10] are a continuous version of the Shepard tones [12]. We chose them to sonify circular motions of the bell, since it was suggested that circular gestures are convenient control for ‘circular’ sounds [13]. Shepard tones consist of 10 sinusoidal components that are linearly spaced in log-frequency (i.e. the  $i^{\text{th}}$  component's frequency is  $2^i f_0, i = 1, \dots, 10$ ). Pitch-shifting those sounds up or down gives rise to an eternally ascending/descending sound illusion. We linearly mapped  $d_N(t)$  to glissandi's amplitude  $a_{\text{gliss}}(t) = \frac{d_N(t)}{d_{\text{max}}}$  and the circular motion angle to the panning angle as  $\theta_{\text{gliss}}(t) = \frac{\pi}{4} \sin \theta_N(t)$ . The angle acceleration  $d^2\theta_N(t)/dt^2$  is low-pass filtered and mapped to the pitch shifting rate of the sinusoidal components as  $\frac{df_{\text{gliss}}}{dt} = h \left( \frac{d^2\theta_N(t)}{dt^2} \right)$ .

#### 3.2. Body Weight Transfer

Body weight transfer is the process of shifting the entire body from its original position to a certain direction: left, right, forward or backward. We used beat interference (resulting from the addition of two sinusoids of close frequencies) to give a distance measure between the musician's current position and his/her original position<sup>3</sup>. The weight transfer is mapped to the beats speed (or tremolo rate), increasing with the frequency difference. Panning is controlled from the  $x$  direction and amplitude from the  $y$  direction. A body weight transfer forward increases the loudness of the sonified sound, while backward decreases its loudness.

Fig. 3 shows the body movements of a musician based on the Optotrak markers on the head, shoulders and elbow. While all markers are highly correlated in the  $x$  direction (upper Fig. 3), the  $y$ -direction (lower Fig. 3) shows two different kinds of movements: genuine body weight transfer and breathing gestures. Breathing gestures are characterized by the elbow and head being  $\pi$  rad out of phase with each other; therefore we take the mean of markers 1 and 7 in the  $x$  and  $y$  directions which are sufficient to give a fairly accurate idea of where the entire body of the musician is (see Fig. 3). We map the  $x$ -distance to the panning angle as:

$$\theta_{\text{interf.pan}} = g \left( \frac{|x(t) - x(0)|}{d_{\text{max}}} \right) \in \left[ -\frac{\pi}{4}, \frac{\pi}{4} \right] \quad (4)$$

with  $g$  the non-linear warping shown in Eq. (5).

The  $y$  projected value is mapped to the amplitude  $a_{\text{interf}}(t) = \frac{y(t) - y(0)}{d_{\text{max}}}$ , and the norm 2 distance  $d_{\text{transfer}}(t)$  is mapped to the frequency difference  $\delta f_{\text{interf}} = \frac{d_{\text{transfer}}(t)}{d_{\text{max}}}$ .

<sup>3</sup>This additive synthesis has been used for surgical positioning [14].

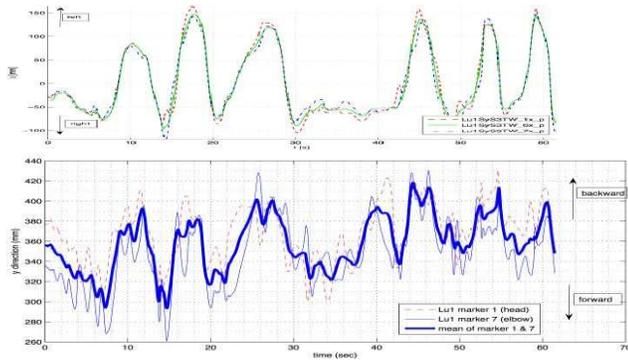


Figure 3: Examples of clarinetist movement in the  $x$  direction (upper figure) and  $y$  direction (lower figure: markers 1 & 7 and their mean in bold).

In mapping  $|x(0) - x(t)|$  to the panning angle, we used a nonlinear mapping function to highlight subtler body weight transfers, making the sonification more informative. The nonlinear function was given by noting  $z(t) = \frac{x(t) - 124}{173 - 75}$  as

$$g(z) = \frac{\pi}{2} \sin \left( \frac{\pi}{2} \sin \left[ \frac{\pi}{2} \sin (\pi z(t)) \right] \right) \quad (5)$$

for  $x(t) \in [75, 173]$ , otherwise  $g(z) = -\frac{\pi}{2}$  for  $x(t) < 75$  and  $g(z) = \frac{\pi}{2}$  for  $x(t) > 173$ . An amplitude parameter  $a_{\text{interf}}$  is introduced so that one only hears relevant changes in the body weight transfers (*i.e.* hear nothing when the musician is not moving) as  $a_{\text{interf}}(t) = h_{10} \left( \frac{\partial d_{\text{transfer}}(t)}{\partial t} \right)$  with  $h_{10}$  a order-10 mean average filter.

### 3.3. Body Curvature

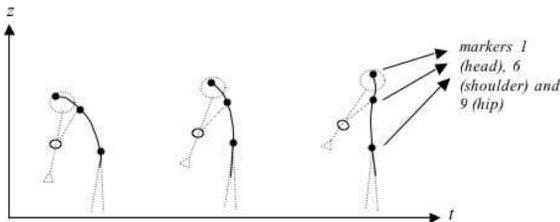


Figure 4: The body curvature of a musician.

We wish to hear a measure of the body curvature degree, from the markers situated on the head, shoulder and hip of a musician (Fig. 4). By analogy with the way a very directive instrument has its sound naturally lowpass filtered (and so ‘darker’) for a curved body, we chose to map curvature to the sound brightness. This can easily be achieved using frequency modulation [11]. Frequency modulation synthesis is a computationally efficient technique with 2 control parameters: modulation index (brightness control) and harmonicity ratio (timbre control).

The curvature  $k$  of the second degree polynomial  $f(x)$  fitting to three points (markers 1, 6 & 9) is measured [15] as

$$k = \frac{f''(x)}{(1 + [f'(x)]^2)^{3/2}} \quad (6)$$

It was evaluated at the shoulder marker coordinate. Examples of curvatures for three musicians are depicted in Fig. 5. We linearly

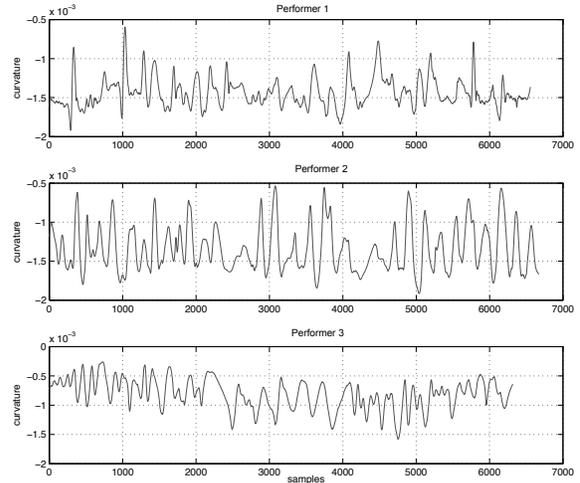


Figure 5: Curvature values for the musicians.

map curvature  $k \in [-0.002, 0]$  to the modulation index as follows:  $m(t) = 10 \frac{k(t) + 0.002}{0.002} \in [0, 10]$  and the amplitude depends on the curvature absolute derivative as  $a(t) = \frac{\exp(600 \left| \frac{dk}{dt} \right|) - 1}{5.05}$ , enabling to only hear relevant changes in the data. The exponential weighting better lowers the volume of the low curvature values.

### 3.4. Knee Movement

When considering the sonification of knee movements, we decided to use the low frequency range, as the knees are at the lower body. Sound is synthesized by low pass filtering a white noise using a biquad filter. The hip to knees distance  $d_{\text{knees}}$  increases as the knees bend and decreases as they straighten. Once again, we only want to hear relative changes in the knees movement so we used the distance absolute derivative to compute the cut-off frequency (after applying an order-30 low-pass filter in order to remove jitter):

$$f_c = 100 \cdot h_{30} \left( \left| \frac{\partial d_{\text{knees}}}{\partial t} \right| \right) \in [0, 200] \text{Hz} \quad (7)$$

where the low-pass absolute derivative varies in  $[0, 2]$ . An amplitude control  $a(t)$  is given loosely by  $a(t) = 1/x$ , using  $x = \frac{1}{100} f_c + 1$  to keep the energy constant. Thus a bigger knee movement corresponds to a higher cut-off frequency.

## 4. DISCUSSION

The auditory scene aspect of sonification – in terms of frequency, amplitude, panning and timbre modulations – needs some special care so as to efficiently represent high-dimensional data sets. The frequency bands activated are an important feature: the sinusoidal components of Risset’s glissandi have their greatest amplitude in the  $[100, 1000]$  Hz range. Body weight transfer sonification has a constant 440 Hz fundamental frequency; the FM sound carrier frequency was set to 318 Hz; and the low-pass filter cut-off frequency used for knee movement ranged in  $[0, 200]$  Hz. Despite that all the four sounds lie within the low and intermediate frequency range,

each of them was clearly distinguishable from the others, mainly because they were modulated in different ways: pitch for Risset's glissandi, brightness or harmonic FM sounds, tremolo for additive synthesis, and brightness or noise for low-pass filtering. Sound panning was used only for gestures where panning was involved (body weight transfer and clarinet bell circular motions). Amplitude was modulated for all 4 sonifications to emphasize changes and avoid ear saturation for continuous sounds.

During the first steps of building our sonification system, mapping settings were particularly good for the first performance it was set with, and less valid for other performances. This suggested that interactive (and so forth fully real-time) sonification would help to avoid this problem by involving several users' points of view in the mapping settings.

The initial tests indicate that sonification of clarinet bell circular motion stresses motions that could not be observed in the video but that existed in the *Optotrak* data. The timbre and amplitude changes of frequency modulated sounds (body curvature sonification) very effectively reflected the body curvature, possibly the most obvious gesture from the video as it involves the entire upper body. Therefore, this system could help performers to be aware of their ancillary gestures (such as body weight transfer or clarinet bell motions), in order to be more conscious of those movements, train to modify them, etc. It also seems that listening to the sonifications could help in analysing correlations between gestures.

Further works include formal experiments in order to evaluate the interest of sonification to study musicians gestures. We set up a pilot experiment, where sonifications were presented alone or with a video. The one corresponding to the sonified data, or another video from the same performer with another expressive manner, or another performer with the same expressive manner. Sonified gesture data are time-warped so as to be synchronized with the corresponding video. Such an experiment will hopefully help to answer research questions such as: 'can one recognize a gesture, a performer, an expressive manner or the performer expertise level? What can be learned from the sonification alone? From the video and the sonification?'

Other further works include checking if four and more simultaneous sonifications do not overload cognition. If they do not, then sonification of musicians ancillary gestures could prove to be a good complementary method of presenting high dimensional data sets in addition to visualisation. For example, some weight transfer motions are subtle and not visible from the video since they happen in a direction we cannot observe, but clearly visible in the *Optotrak* data. However, the user cannot at the same time watch four curves and detect specific patterns, whereas the sonification creates an auditory display of 6 curves through 4 sonifications. In such a case, sonification will not only bring some complementary information to the video but also facilitate the extraction of information from the markers' data.

## 5. CONCLUSIONS

Looking for new tools to study musicians' ancillary gestures, we started to investigate the sonification of markers data obtained from previous studies. When building this sonification system, we focused on the selection of gestures, the choice of appropriate synthesis techniques, and then built adequate mappings between them. An important aspect we focused on was to build an efficient auditory scene. Preliminary observations have indicated that sonification of a set of up to 4 musicians ancillary gestures can be heard

clearly. Further improvements will involve interactive sonification in order to provide more efficient mappings. Sonification is a complementary tool to identify and qualitatively analyse musicians ancillary movements. However, it is a research in progress: formal experiments have now to be conducted, in order to reveal if listeners can identify gestures, performers and performance manners.

## 6. REFERENCES

- [1] S. Barrass and G. Kramer, "Using sonification," *Multimedia Systems*, vol. 7, pp. 23–31, June 1999.
- [2] A. O. Effenberg, "Movement sonification: Effects on perception and action," *IEEE Multim*, vol. 12(2), pp. 53–9, 2005.
- [3] K. Beilharz, "Wireless gesture controllers to affect information sonification," in *Proc. 11th Meeting Int. Conf. Auditory Displays, Limerick*, 2005, pp. 105–112.
- [4] S. Pauletto and A. Hunt, "Interactive sonification in two domains: helicopter flight analysis and physiotherapy movement analysis," in *Proc. Int. Workshop on Interactive Sonification, Bielefeld*, January 2004.
- [5] T. Hermann, O. Höner, and H. Ritter, "Acoumotion - an interactive sonification system for acoustic motion control," in *Proc. Int. Gesture Workshop, (GW 2005), Vannes*, 2005.
- [6] M. M. Wanderley, Ph. Depalle, and O. Warusfel, "Improving instrumental sound synthesis by modeling the effect of performer gestures," in *Proc. Int. Comp. Music Conf.*, 1999, pp. 418–21.
- [7] NDI, *Optotrak*, "<http://www.ndigital.com/certus.php>," 2006.
- [8] M. M. Wanderley, B. W. Vines, N. Middleton, C. McKay, and W. Hatch, "Expressive movements of clarinetists: Quantification and musical considerations," Tech. Rep., MT2004-IDIM01, IDMIL, McGill, Oct. 2004.
- [9] L. Campbell, M.-J. Chagnon, and M. M. Wanderley, "On the use of laban-bartenieff techniques to describe ancillary gestures of clarinetists," Tech. Rep., IDMIL, McGill, 2005.
- [10] J. C. Risset, "Pitch control and pitch paradoxes demonstrated with computer-synthesized sounds," *Jour. Ac. Soc. of Am.*, vol. 46, no. (A), pp. 88, 1969.
- [11] J. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Comp. Music Journal*, vol. 1, no. 2, pp. 46–54, 1977.
- [12] R. Shepard, "Circularity in judgments of relative pitch," *Journal of the Acoustical Society of America*, vol. 36, no. 12, pp. 2346–53, 1964.
- [13] D. Arfib, "Le geste créatif en informatique musicale," Tech. Rep., *Conseil Général des Bouches du Rhône*, 2002.
- [14] E. Jovanov, K. Wegner, V. Radivojevic, D. Starcevic, M. Quinn, and D. Karron, "Tactical audio and acoustic rendering in biomedical applications," *IEEE Trans. on Information Technology in Biomedicine*, vol. 3, no. 2, June 1999.
- [15] H. S. M. Coxeter, *Introduction to Geometry*, 2nd ed. New York: Wiley, 1969.