Methods for Effective Sonification of Clarinetists' Ancillary Gestures

Florian Grond[1], Thomas Hermann[1],
Vincent Verfaille[2], and Marcelo M. Wanderley[2]

[1] Ambient Intelligence Group, CITEC, Bielefeld University
http://www.techfak.uni-bielefeld.de/ags/ami
[2] IDMIL, CIRMMT, McGill University
http://www.idmil.org

Abstract. We present the implementation of two different sonifications methods of ancillary gestures from clarinetists. The sonifications are data driven from the clarinetist's posture which is captured with a VICON motion tracking system. The first sonification method is based on the velocities of the tracking markers, the second method involves a principal component analysis as a data preprocessing step. Further we develop a simple complementary visual display with a similar information content to match the sonification. The effect of the two sonifications with respect to the movement perception is studied in an experiment where test subjects annotate the clarinetists performance represented by various combinations of the resulting uni- and multimodal displays.

Keywords: sonification, 3D movement data, ancillary gestures, multimodal displays.

1 Introduction

Through advanced recording and simulation possibilities the amount of 3D movement data is constantly growing. The standard technique to investigate such data is the scientific 3D visualization of moving points or models, a self evident approach, since the human visual and cognitive system seems highly adapted to perceive and interpret human motion. However, the rather young research field of sonification offers novel inspection techniques that complement visual analysis by transforming data into audible sound. This perception mode is beneficial for several reasons: Firstly, sonification is ideal for representing dynamic patterns in multivariate data sets with complex information such as fast transient motions. Secondly, sound requires neither a particular orientation of the user nor directed attention. Thirdly, in many applications the eyes are already occupied with a specific task and have therefore limited capacity to focus on additional information. In this case however, the intricate interplay of multimodal displays, specifically the one of sound and moving image has to be taken into account. These benefits are of potential interest for the study of ancillary gestures of instrumentalists, in our case clarinetists.

S. Kopp and I. Wachsmuth (Eds.): GW 2009, LNAI 5934, pp. 171–181, 2010.
© Springer-Verlag Berlin Heidelberg 2010

Ancillary gestures are those body movements which are not directly involved in the sound production [1] [2] but are omnipresent in musical performance. For clarinet players, lip and finger motions are effective gestures, whereas motions like weight transfer and body curvature for instance are ancillary gestures. Their importance is due to the fact that they tend to align with musical motives in the score [2] and are therefore an integral part of the player's performance as these movement patterns show consistency even for various levels of expressiveness [3].

In this paper, we develop two different sonification strategies of ancillary gestures. The first is a direct mapping of marker velocities to sound, the second involves a principal component analysis as a data preprocessing step. Specific sonifications for this purpose have already been developed to some extent by the authors [4] [5] [6] [7] and others [8] [9] but there is still a lack of understanding, how sonification influences multimodal displays. and if different sonification strategies have noticeably different effects on how we perceive data in a in multimodal displays. Besides the development of sonification techniques, we put our focus on assessing the effects in a psychophysical experiment, in which we ask test subjects to identify self-chosen events and consistently detect their selection in repeated presentations.

2 The Vicon Motion Tracking Data

The motion capturing sessions were conducted in previous projects at the IDMIL. All performers were advanced instrumentalists playing an excerpt of Brahms' Sonata for clarinet op. 120 no 1. The movement data of the clarinetist were recorded with a VICON system 460 system using the standard plug-in-gait model [10], which provides 38 marker positions and gives a global description of the body posture. We removed from the resulting data redundant channels and decided to apply sonification to the posture information in terms of marker positions each with x, y and z Cartesian coordinates and its derivatives. This choice was motivated by that fact that we wanted to apply sonification to aspects of the data which could also be seen in a simple visual representation. The remaining set consisted of 18 markers, some of them were computed as combinations of originally measured data c.f. Table 1.

Table 1. On the left: the reduced dataset of 18 markers. On the right: clarinetists in the VICON motion capturing system.

body part	left	middle	right
head	front	back of the head	front
spine		neck C7	
		T10	
		end of spine	
arms	shoulder		shoulder
	Elbow		Elbow
	arm wrist		arm wrist
legs	hip		hip
	knee		knee
	ankle		ankle

The marker in the middle on the backside of the head corresponding to the two on the front, was computed as the mean of two left and right markers in the back. The marker at the end of the spine was the mean vector of the left right backside of the hips. The markers on the hips are the average of two markers back/front of each side respectively.

3 Data Preparation and Preprocessing

Before reducing the data as described above, we centered them between both feet. This was done such that for each time frame the center of origin of the coordinate system was moved to the middle between the left and right toe. In general, the clarinetists left their toes in one place during the whole performance (except one recording). Therefore, there was not much dynamics in the marker on the toe.

The left and right as well as the back and forth movements (weight transfer, WT) of most of the subjects dominated the whole dataset. For the PCA transform this lead to the fact that the first and second components consisted of the WT movement. In the velocity sonification approach, the WT was dominating in a similar way, when mapping velocity to sound. Therefore we removed this component from the data set. The WT can be understood as the moving center of mass (COM) of the performer, like a reversed pendulum pointing up, which is anchored between both feet of the performer and oscillates around the basis vector z (0,0,1) in the x,y plane. After removing this motion, the COM vector always remained parallel to (0,0,1).

4 Sound Synthesis and Mapping

4.1 Sound Synthesis

According to the Design Space Map [11] a recording rate of 100 Hz, suggested a continuous parameter mapping sonification as an appropriate strategy. The following two considerations were guiding us during the sonification design: The continuous sonification should automatically lead to acoustic articulations which segment the audio in a perceptually meaningful way according to the movement patterns. Further, the sonification should allow to distinguish movements from different parts of the body. Yet in order to know, what the test subjects would be listening to, if they directed their attention towards the sonification, we decided to design a single auditory stream. Therefore we constructed as the simplest sonification unit a source filter model. using white noise filtered through a resonant filter.

$$s(t) = \sum_{i=1}^{n} H_{Resonz}(\eta_i(t); f_i, rq, g) \tag{1}$$

Where H_{Resonz} is the resonant filter with the resonant frequency f_i, the gain g, the bandwidth is specified with the filter's reciprocal q -value rq and $\eta_i(t)$ is

a white noise source. Motion data are mapped to these parameters as follows. The resulting sonification as $s(t)$ is the sum over all n sonified data-features. To address the frequency loudness dependency we used basic psychoacoustic amplitude compensation.[1]

These sounds of filtered noise integrated nicely into one sound stream where the varying amplitude of the different resonant frequencies f_i could be distinguish.

4.2 Mapping

As mentioned two different mappings of the the data features to the sonification units have been applied.

Velocity Sonification: The first was a direct mapping of the velocities of all 18 data points to the sound parameters described above. Due to the noise in the data the derived velocity was smoothed with a rectangular window of 5 samples. Frequencies between 150 and 4000 Hz were assigned to each marker. The gain corresponded to the velocity exponentially mapped between 0.001 and 1. The velocity of each data point modulated the center-frequencies with ±5% and additionally the rq of each resonant filter was mapped exponentially between 0.001 and 0.1.

PCA Sonification: For the second mapping, we computed the principal components of the data set consisting of $3 * 18 = 54$ features over a complete performance.[2] Since the COM corresponding to the 54d vector of the mean posture has been subtracted before computing the data set covariance matrix $C = \frac{1}{N} \sum_\alpha x^\alpha x^{\alpha T}$ the principal components describe axis along coordinated activity. Then we took the first 6 coefficients corresponding to the largest eigenvalues of C that that cover approx. 85 % of the data set variance. It turned out, that taking components after the first 6, which described minute movement details could not be distinguish acoustically to our experience and were hard to identify visually too. The coefficients of the principal components were exponetially mapped between 300 - 2000 Hz.[3] The velocities of the coefficients were exponentially mapped to gain between 0.001 and 1. Frequency modulation corresponded to ±5% around the assigned frequency controlled by the time coefficients. The rq of the resonant filters corresponded to the principal components exponentially mapped between 0.001 and 0.1 In both approaches the resulting sound was the sum over all sonified data features as in eq. 1.

Data Selection for the Experiment: In order to cover different ancillary gesture patterns these sonifications were applied onto the data of the complete performance of 3 clarinetists, which have been selected since they exhibited noticeably

[1] For the details of the resonant filter and the amplitude compensation we refer to the implementation details of the `Resonz` and `AmpComp` class in `SuperCollider3`.

[2] PCA was computed via pythonSC [12].

[3] Frequencies and range were chosen to yield an acoustically rich result.

different movement patterns. Clarinetist 1 and 3 exhibited very pronounced ancillary gestures with lots of weight transfer movements, whereas clarinetist 2 was only occasionally moving the whole body, but instead made rapid movements with the elbows and the arm wrist.

5 Considerations on Sound and Image

For the design of multi modal displays it is important to consider that human perception is not merely a superposition of moving image plus sound. The fact that sound can give an added value to the image has already been extensively studied by Chion [13] and Flueckiger [14] for the cinema. Some findings from these works are particularly instructive for the design of sonifications for multi modal displays. One is that we always look for a visual cause that explains our acoustic impressions. Therefore the sound always seems to come from the image and is together with a visual representation not perceived in isolation. On the contrary, if no cause can be found in the visual display, no integrated perception emerges. This implies in turn that complementarity receives a particular meaning for multi modal displays. Sonification can most effectively be used when it emphasizes information that is already in the visual display and can be more precisely identified there. Unfortunately this "searching" for a visual cause can also lead to unwanted results as the following experience shows: In [7] a framework for ancillary gestures sonifications for clarinetists was developed, where in a first trial sonifications were combined with videos from the clarinet players clearly showing the finger movements. The effect was that some test subjects asked if the finger movements (effective gesture) caused the sound, where in fact the sonification was representing only ancillary gestures such as weight transfer and body curvature, which were less visible.

Given these considerations about the interplay of sound and image, the main focus was if different sonifications of ancillary gestures change the way how we perceive them and if both modalities can be efficiently integrated. By giving the test subjects an open task we investigated if sonifications can contribute towards a more unanimous interpretation of perceived movements amongst test subjects.

5.1 The Visual Display of Body Movements

Because of the aforementioned considerations we developed a visual display which only shows ancillary gestures[4]. Figure 1 shows an abstract stick figure in profile and from the front omitting details such as finger movements. In small preliminary tests we received a good feedback for this display design in terms of a consistency between sonification and image, yet people still reported difficulties to identify which part of the body was responsible for a certain sound. Therefore we added, at least for the velocity sonification, a red glyph to each sonified joint, which changed in size and color (by varying the alpha channel) according to the velocity. This provided visually similar information, as mapped to sound in the velocity sonification.

[4] The visualization in `SuperCollider3` was implemented in `SCgraph` [15].

Fig. 1. Visualization of the clarinetists: (a) simple stickman (b) enhanced stickman where the velocities of joints are additionally highlighted with red glyphs

6 Experiment

6.1 Task

Given the diverse background of the test-subjects, we designed an open task for the experiment, which consisted of identification and detection of events in stimuli for selected sequences of movements of the three clarinetists. All stimuli represented the same musical phrase consisting of 4 distinct melodic units. The test subjects were asked to look and listen to each stimulus 9 times and to identify events that they encountered and to mark their choice in real time by mouse clicks. The test subjects were told to consider the first two runs as test runs and to try to repeat their selection of events consistently in the subsequent 7 runs. At the end of all trials the test subjects were asked to fill out a questionnaire about the different stimuli and about their musical experience.

6.2 Setup

The combination of the two visual representations and the two sonifications resulted in 7 different stimuli, which were presented to the test subjects.

id	sonification	vizualization	acronym
1	direct	-	A1 V0
2	direct	stickman	A1 V1
3	direct	stickman + glyphs	A1 V2
4	-	stickman	A0 V1
5	-	stickman + glyphs	A0 V2
6	PCA	-	A2 V0
7	PCA	stickman	A2 V1

In the case of the PCA based sonification we omitted the highlighting red glyphs since those aspects were not sonified. For each test subject the order of the stimuli was randomized ensuring that the 3 clarinetists were interleaved. For clarinetists 1 and 3 the velocity sonification resulted in a very structured sound that was easy to connect with the visual representation. The PCA sonification for the selected region appeared less structured and was therefore more difficult to connect with the visualization. This was reversed for clarinetist 2, where the PCA sonification lead to more structured sounds.

7 Experimental Data Evaluation

We recorded and analyzed data for 12 test subjects aged from 22 - 33, 11 male, 1 female, 7 of them playing an instrument. Since the movement patterns are very different for each clarinetist, resulting in very different direct and PCA sonifications, the analysis was made for each clarinetist individually.

Average Click Frequency: At first we are interested in the average number of clicks given for each stimuli. The results are compiled in Figure 2. In all the subsequent figures the results for the different stimuli are represented showing the audio only condition on the left (A1V0, A2V0) and the visual only condition on the right (A0V1, A0V2) , leaving the middle for the combined stimuli (A1V1, A1V2, A2V1).

Although the click-frequency does not vary significantly across the conditions, it is interesting to note that the highest click frequency was obtained with stimuli that at least included a visual representation. The lowest click-frequency always appeared with stimuli that contained an audio only condition. For clarinetist 3, who had a very structured performance with pronounced ancillary gestures, the standard-deviation is much smaller for most of the conditions comparing it with the other two clarinetists. This suggests that the intersubjective convergence in the perception of ancillary gestures is first and foremost influenced by how pronounced the ancillary gestures themselves appear.

Kernel Estimated Click Density: In order to compare the different stimuli along the performance of the clarinetists we visualized the results by computing a kernel estimated click density. Selected intervals of the click densities are depicted in figure 3.[5] The three selections (a, b, c) from figure 3 are examples for patterns we noticed in the plots. Selection a shows that in some cases multimodal conditions led to a noticeable delay in the reaction of the participants (we also found a case where visualization only, V2A0, was triggering clicks faster, than in the multimodal case V2A1). We hypothesize that integrating two perceptual streams increases the cognitive load and therefore causes the delay. Selection b shows that the velocity sonification seemed to dominate in clarinetist 2 and made the test-subjects ignore an event that was selected in A0V1 and A2V1. The last two conditions are very similar in the click-profile (except the multimodal delay in A2V1), the sonification of A2 therefore seemed not to overrule the visualization. Around second 6 in selection b we found that A1V1 made the test-subjects clearly differentiate two events which we could identify as two quick arm-wrist movements. Selection c clearly shows a peak at sec. 3 for the stimulus A1V2. At this moment clarinetist 3 made a step, which was interestingly not noticeably marked by clicks in any V1 or V2 only condition.

In Table 2 you find the results of the Kolmogorov Smirnov test over the click trains comparing different stimuli. An interesting because most general

[5] For an overview about all kernel estimated click densities please follow this URL: `http://www.techfak.uni-bielefeld.de/~fgrond/GW2009/`

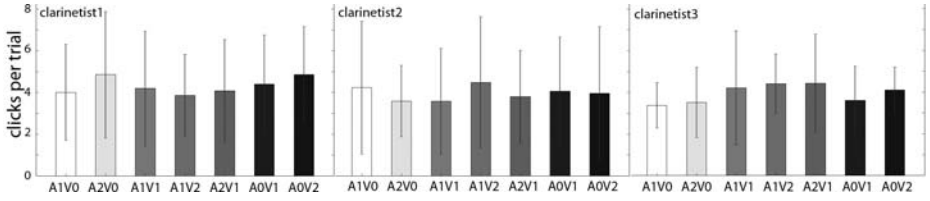

Fig. 2. Average number of clicks per trial

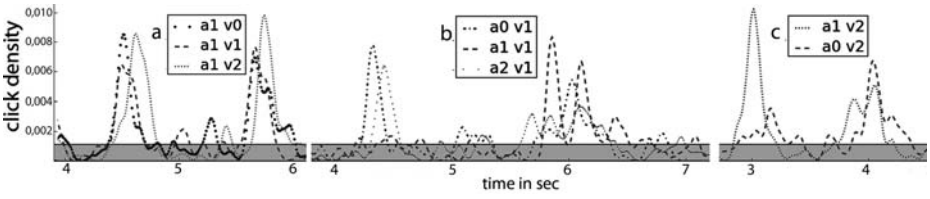

Fig. 3. Kernel estimated click densities. *a*,*b* and *c* are selected intervals of clarinetist 1, 2 and 3 respectively. The grey horizontal bar indicates the average click density $\int = 1$.

Table 2. comparing the clicktrains by the Kolmogorov Smirnov test. Values below **5%** are indicated. The first block of 3 stimuli pairs (line 1-3) compares A1 in combination with V0 V1 V2. The second block of 3 stimuli pairs (line 6-9) compares V1 in combination with A0 A1 A2.

compared stimuli	clarinetist 1	clarinetist 2	clarinetist 3
A1V0 — A1V1	0.233	**0.032**	**0.011**
A1V1 — A1V2	**0.010**	0.493	0.302
A1V2 — A1V0	0.180	**0.010**	0.078
A2V0 — A2V1	0.226	0.086	0.248
A0V1 — A0V2	0.144	0.105	0.441
A0V1 — A1V1	0.618	**0.018**	0.127
A1V1 — A2V1	0.223	**0.007**	0.360
A2V1 — A0V1	0.487	0.508	0.368
A1V0 — A2V0	0.086	**0.001**	**0.012**
A1V2 — A0V2	**0.006**	**0.006**	0.104

result (last row) is that adding the velocity sonification to V2 made a significant difference for clarinetists 1 and 2, even for clarinetist 3 the value of 10% is low. Particularly for clarinetist 2 the click distribution of many stimuli pairs were significantly different accepting a threshold of 5%. A qualitative analysis as done for figure 3 reveals however that also for clarinetist 1 and particularly for 3 different choices were made by the test-subjects depending on the modality of the stimuli. In order to illustrate those differences we plotted the click time versus the click number as shown in figure 4. Comparing the conditions A1V0, A1V2, A0V2 shows how adding the velocity sonification A1 to an already enhanced visualization V2 led to a smaller distribution around the diagonal of succeeding clicks. The condition A1V0 on the left shows the most pronounced diagonal with

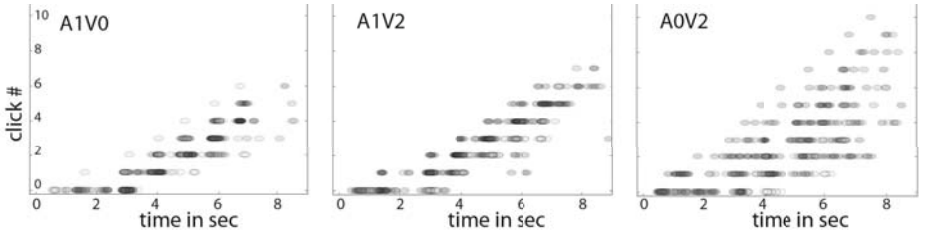

Fig. 4. Clarinetist 3, click time versus click number

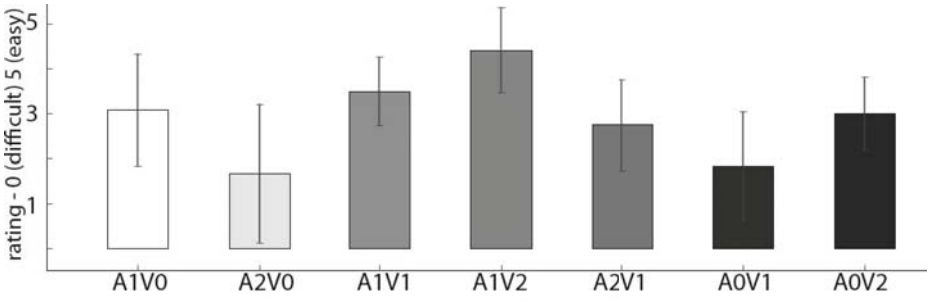

Fig. 5. Self rating of test subjects. x shows the stimuli, y the mean ± standard error.

few outliers, in the condition A0V2 we find less coherence, the plot in the middle A1V2 lies in between, this can be interpreted as intersubjective convergence, which is that velocity sonification alone or together with a visual display forced the test-subjects to select similar events in a similar order.

Self Rating: At the end of the experiment the participants were asked to rate all the stimuli between 1 (difficult) and 5 (easy) with respect to how much they helped them in achieving the given task of consistently selecting events in the display. The results are compiled in figure 5. The multimodal conditions involving velocity sonification A1V2 A1V1 were first and second rated. Interestingly less preference was given to the PCA sonification, which is consistent with the findings in analyzing selection *b* in figure 3.

8 Conclusion

In this paper we implemented a velocity and a PCA based sonification for ancillary gestures and studied them as a stand alone as well as in combination with a visual displays. Sonification as a stand alone display directed the focus on similar events. These events were marked in their timely order in a more unanimous way amongst test subjects. The velocity sonification seemed to be

more efficient in highlighting otherwise overseen aspects of the gestures and was preferred by the test-subjects. We hypothesize that the coordinated directions of movement as extracted through the PCA do not necessarily correspond to what we perceive in a visual display and we encounter therefore difficulties in connecting these two display modes. The fact that the PCA extracts for each clarinetist different coordinated directions of movement depending on their idiosyncratic patterns means that the "meaning" of the PCA, i.e. how it connects with the visual display, has to be learned by the test subjects for each clarinetist anew. For clarinetist 3 we found that, even if the visualized information was similar to the sonification, displaying ancillary gestures visually made the subjects select different events compared to the unanimous choice in the A1V0 and A1V2 case. Although we set up a very open task without testing reaction time, we found evidence that multimodal displays are processed slower by the user. Further we found evidence that sonification has the potential to effectively guide attention to information that is present in the visual display, but not necessarily in the focus of attention. This might be of particular interest for annotation tools in gesture analysis. We therefore see the main purpose of sonification in this particular setting in guiding attention rather than adding information that is not present in the visual display or not perceivable there.

Acknowledgments. This work was supported through a Short-Term Scientific Mission at IDMIL CIRMMT McGill University by the COST Action IC0601 on Sonic Interaction Design and later by an NSERC Discovery grant. For interesting discussions and various support we like to thank Alexandre Savard, Ulf Grosskatehöfer, Till Bovermann and Florian Paul Schmidt.

References

1. Cadoz, C., Wanderley, M.M.: Gesture - Music. In: Wanderley, M.M., Battier, M. (eds.) Trends in Gestural Control of Music. Ircam – Centre Pompidou (2000)
2. Wanderley, M.M., Vines, B.W., Middleton, N., McKay, C., Hatch, W.: The musical significance of clarinetists' ancillary gestures: an exploration of the field. Journal of New Music Research 34(1), 97–113 (2005)
3. Wanderley, M.M.: Quantitative analysis of non-obvious performer gestures. In: Gesture and Sign Language in Human-Computer Interaction, pp. 241–253 (2002)
4. Hermann, T., Höner, O., Ritter, H.: Acoumotion - an interactive sonification system for acoustic motion control. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 312–323. Springer, Heidelberg (2006)
5. Nusseck, M., Wanderley, M.M.: Music and motion—how music related ancillary body movements contribute to the experience of music. Music Perception 26(4) (2009)
6. Verfaille, V., Quek, O., Wanderley, M.M.: Sonification of musicians' ancillary gestures. In: Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20 -23 (2006)
7. Savard, A.: When gestures are perceived through sounds: A framework for sonification of musicians' ancillary gestures. Master's thesis, IDMIL CIRRMT McGill University (August 2008)

8. Effenberg, A.O.: Movement sonification: Effects on perception and action. IEEE Multim. 12(2), 56–69 (2005)
9. Goina, P., Polotti, M.: Elementary gestalts for gesture sonification. In: NIME 2008, pp. 150–153 (2009)
10. Ferrari, A., Benedetti, M.G., Pavan, E., Frigo, C., Bettinelli, D., Rabuffetti, M., Crenna, P., Leardini, A.: Quantitative comparison of five current protocols in gait analysis. GAIT POSTURE (2008)
11. de Campo, A.: Toward a data soni cation design space map. In: Scavone, G.P. (ed.) ICAD 2007, Schulich School of Music, McGill University, Schulich School of Music, McGill University, Montreal, Canada, pp. 342–347 (2007), http://www.icad.org/Proceedings/2007/deCampo2007.pdf
12. Hermann, T.: PythonSC, http://www.sonification.de/projects/sc3
13. Chion, M.: Un art sonore, le cinéma, le phrasé audio visuelle. Cahiers du cinéma, pp. 541–561 (2003)
14. Flückiger, B.: Sound Design Die virtuelle Klangwelt des Films. Schüren, Marburg (2001)
15. Schmidt, F.P.: Design and implementation of a realtime 3d graphics server. Master's thesis, Bielefeld University, Germany (April 2007)