VIRTUAL AUDITORY MYOGRAPHY OF TIMPANI-PLAYING AVATARS

Florian Grond
AMI – CITEC
Bielefeld University
fgrond@techfak.uni-bielefeld.de

Alexandre Bouënard
IDMIL – CIRMMT
McGill University
alexandre.bouenard@mcgill.ca

Thomas Hermann
AMI – CITEC
Bielefeld University
thermann@techfak.uni-bielefeld.de

Marcelo M. Wanderley
IDMIL – CIRMMT
McGill University
marcelo.wanderley@mcgill.ca

ABSTRACT

Music performance is highly related to instrumentalists' movements and one of the biggest challenges is the identification and understanding of gesture strategies according to the plethora of musical nuances (dynamics, tempo, etc..) available to performers. During these past few years, a novel approach has been elaborated, consisting in studying movement strategies through auditory rendering. In this paper, we focus on the auditory analysis of timpani (percussion) gestures. We present a novel interface combining movement simulation and sonification as a means of enhancing the auditory analysis of timpani gestures. We further report the results from an evaluation of this interface, where we study the contributions of sonification to the multimodal display.

1. INTRODUCTION

Musical performances are usually and naturally associated with instrumentalists' gestures, the importance of which has been intensively documented and studied [1, 2]. Performer gestures can traditionally be divided into two categories: effective and ancillary ones. Effective gestures are those that are directly involved in the sound production, sometimes fingers and hands, as for the clarinet or the piano. Conversely, ancillary gestures are not directly involved in sound production processes, but are omnipresent in musical performance and are an integral part of the performers expressivity.

The existence of such variabilities in instrumental gestures are the common basis of gesture studies, usually conducted through the analysis of motion data captured from real instrumentalists performing with varying expressiveness and various gesture articulations. The availability of such data makes the analysis immediately relevant for the performing subject, and allows the exploration of the influence of music variations on instrumental gestures, such as gesture anticipation phenomena occuring under musical nuances and tempo changes [3, 4, 5]. However the inter-dependance and high dimensionality of gesture data makes it difficult to study the gesture-sound relationship apart from a divide-and-conquer approach.

In this work we propose an approach combining movement simulation and sonification for enhancing the perception of these subtle variations occuring in instrumental gestures. The interface allows for the selection of a wide range of gesture parameters as well as their mapping to sonification methods. In our case we study the gesture patterns of timpani players. With timpani players the movements of the whole mallet-arm-shoulder system participates to the effective gestures, and changes in gesture articulations can alter the gesture strategies involved in the system. For instance musical nuances can influence preparatory gestures preceding beat impacts, especially for mallet extremity trajectories [5]. The simulation of the instrumental gesture comes into play, as a possibility to systematically control and study a variety of movement parameters.

The paper is organized as follows: section 2 discusses the proposed work in the light of previous contributions, both regarding gesture sonification and modeling. The framework used to model and simulate timpani gestures is presented in section 3, providing the essential knowledge about the in-and-outs on which the sonification is based. Section 4 details the sonification process of our work, and particularly the sonification methods used. The user interface is presented in section 5, as well as relevant experiments that are possible to run with it. Section 6 presents a psychophysical experiment that has been conducted with our simulation/sonification framework. Finally, section 7 discusses our approach and concludes with further perspectives.

2. RELATED WORK

2.1. Gesture Sonification

Displays for movements of the human body are first and foremost visual, a self-evident approach, since the human visual and cognitive system seems highly adapted to perceive and interpret human motion [6]. However, sonification offers novel inspection techniques that complement visual analysis by transforming data into audible sound. Specific sonifications for this purpose have already been developed to some extent by the authors [7, 8, 9]. This perception mode is beneficial for the movement data of timpani players for two reasons: On the one hand, multivariate data sets with complex information such as fast transient motions can be perceived as one auditory stream. On the other hand the repetitive nature of timpani playing makes it easy for the listener to establish similarities and differences in what is presented acoustically, thereby making immediately sense of the auditory information channel. Sonifications in this field have so far only represented data that are directly related to the movement. However in simulations, there are additional interesting data-streams such as efforts for instance, which need to be displayed.

Simulated torques are such effort-related parameters, and can be understood as the turning force that moves the structure of the avatar, similarly to the force exerted through the muscles of the real performer onto its joints. This situates our work in a long tradition of sonification which is the auditory display of electromyography (EMG) data. Early descriptions, can be found in 1958 [10] where different conditions can be clearly distinguished acoustically. Recent research has been conducted in [11] where modern sonification methods are employed and the effect of the interactive nature of EMG sonification has been studied. In this paper we introduce sonification of torques from movement simulations and demonstrate on a selected example how they can be integrated as interesting dynamic features in an acoustic display.

More specifically, torques have already been used in [12] as a sonification parameter, considering the avatar motion into a low-dimensional PCA-space and using torques at the global level of the human body. Our sonification work differently uses torques at the joint level and makes no dimension-reduction hypothesis of the avatar motion, thereby allowing the sonification of the motion of each individual joint and to perceive its contribution to the global motion. Additionally, we use *SuperCollider* as sound synthesis engine which allows for a more flexible sound design compared to the MIDI controled synthesis.

With respect to multimodal perception, interesting work can be found in [13] which shows that sound dominates the temporal perception and visualization dominates the spatial perception. The potential of sound to alter visual perception is demonstrated in [14, 15].

Despite all these initial contributions to gesture sonification and established results from perceptional psychology, audio-visual displays of complex and specific movements such as the avatar in our case have not been evaluated so far. This is why we decided to perform a psychophysical experiment to investigate the effect of sonification in a discrimination task.

2.2. Gesture Modeling

The modeling of instrumental gestures has often been under attention in computer music, mostly as a means of understanding the relationship between actions applied to music instruments and the resulting sounds. Based on models emulating real-world gesture mechanisms, another important benefit of such a modeling/synthesis approach is the interest to submit novel gesture parameters to sound rendering processes. Modeling gesture parameters can be more-or-less bounded to the gesture-sound interaction process. Strongly related to the interaction phase is the modeling for instance of contact [16] or attack [17] parameters. Other contributions aim at modeling complete instrumental gestures, including preparatory phases in addition to the interaction process. Early works have involved for instance MIDI-driven [18] or physically-simulated gestures [19]. More recent attempts include such modeling/synthesis approach using pre-recorded motion data combined with optimized dynamic systems [20], or statistical gesture models [21, 22].

Another promising approach is to combine real body motion data alongside with the simulation of physics laws, thereby maintaining the sensorimotor realism of synthesized gestures with regards to the performing task, as well as the responsive behavior of gestural actions towards the surrounding environment [23]. This has been successfully applied in the particular case of the modeling and simulation of percussion gestures to control sound synthesis processes [24]. Such an approach can also have benefits for the sonification process, as it provides not only kinematic clues about timpani gestures – such as position, velocity and acceleration – but also physical ones like forces and torques that can be related to effort parameters during simulated performances. Such a gesture model can therefore propose gesture parameters not easily available in motion capture settings, and also expand the reality in the virtual domain for creating novel instrumental situations.

3. GESTURE MODELING AND SENSORIMOTOR CONTROL

This section presents the framework used to model and simulate timpani gestures. This work is based on the sensorimotor model detailed in [23]. Here we only present the essential information about the model to help the reader to understand its underlying mechanisms, *i.e.* how pre-recorded real motion data used in conjunction with our model can be useful to recover estimated physical features about the original motion. We furthermore highlight the interest of such models for sonification purposes.

3.1. Modeling and Control

Figure 1 gives a summary of the data streams involved in the gesture model, first by invoking a physics model parameterized by anthropometric and biomechanical parameters coming from real motion data. A second step is to control the model through a sensorimotor control strategy based on the percussion gesture trace: mallet extremity trajectories.

The gesture model is represented by a skeleton $\boldsymbol{T}$ composed of a set of rigid bodies. The skeleton can cope both with static and dynamic parameters, cf. equation (1). Static parameters include parameters related to the human anthropometry, such as limbs' mass m_i, density ρ_i, inertia tensor I_i. Dynamic parameters can also be taken into account, such as the way skeleton limbs are articulated with each other. This involves especially how biomechanical parameters of joints like stiffness k_s^j and damping k_d^j model tension/relaxation properties of human joints.

$$\boldsymbol{T} = [\{m_i, \rho_i, I_i\}_{i\in[1...N]}, \{k_s^j, k_d^j\}_{j\in[1...M]}] \quad (1)$$

$$\boldsymbol{F_P} = m.\boldsymbol{\Gamma_P} \quad (2)$$

$$\boldsymbol{\tau_P} = I_P.\dot{\boldsymbol{\Omega}} + \boldsymbol{\Omega}.I_P.\boldsymbol{\Omega} \quad (3)$$

Once the presented static/dynamic attributes of the gesture model are parameterized, a realistic simulation can be settled, setting the model in motion according to the laws of physics. The physics simulation processes at every time-step the linear acceleration $\boldsymbol{\Gamma_P}$ and angular velocity $\boldsymbol{\Omega}$ of each rigid body composing the gesture skeleton, cf. equations (2) and (3). This process consequently needs the knowledge of physical forces $\boldsymbol{F_P}$ and torques $\boldsymbol{\tau_P}$ to be applied on the rigid bodies to put the gesture skeleton into motion.

The described gesture model has several advantages, (a) it takes the anthropomorphic and biomechanical properties of the human body into account, and (b) it allows the modeling of gesture

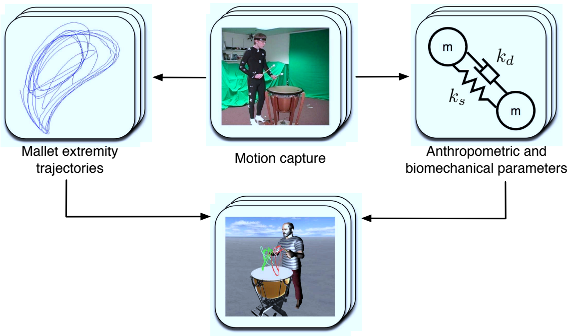


Figure 1: Sensorimotor simulation of percussion gestures: motion capture data from real timpani performers are used to identify mallet extremity trajectories as well as biomechanical parameters, which are the inputs of physics-sensorimotor framewrok for simulating timpani gestures.

responsiveness to external disturbances. However, these advantages turn rapidly into diffculties when the question comes to determining this huge set of parameters, especially considering the complexity of the human body.

A method is of course to consider a simplified gesture model with regards to the number of limbs and joints' degrees of freedom, and to relate this limited set to anthropomorphic tables. In our case, the anthropometry arises from pre-recorded motion data, and we consider a system representing the upper-body augmented with timpani mallets totalling 11 rigid bodies articulated by 33 degrees of freedom, and specifically for each arm, 4 bodies (clavicle, upper arm, forearm, mallet) and 12 degrees of freedom (neck, shoulder, elbow, wrist).

Another difficulty resides in the processing of forces/torques to set the gesture model into motion, especially when noticing that no unique couple (forces, torques) exists at each simulation step-time, so that the gesture model can achieve the task. We propose a control strategy composed of a cascaded combination of two inverse processes that infer an estimation of the needed forces/torques based on real motion data.

$$\boldsymbol{\Theta}_C = \mathcal{K}^{-1}(\boldsymbol{X}_C) \quad (4)$$
$$\boldsymbol{\tau}_C = \mathcal{D}^{-1}(\boldsymbol{\Theta}_C) \quad (5)$$

The first inverse process $\mathcal{K}^{-1}$ infers joint angle trajectories $\boldsymbol{\Theta}_C$ (neck, shoulder, elbow, wrist) from pre-recorded trajectories of mallet extremity $\boldsymbol{X}_C$, cf. equation (4). The second inverse process $\mathcal{D}^{-1}$ infers the torques $\boldsymbol{\tau}_C$ to apply on limbs (clavicle, upper arm, forearm and mallet) from the computed joint angle trajectories $\boldsymbol{\Theta}_C$, cf. equation (5). More details about the sensorimotor control scheme involving these inverse processes as well as algorithmic issues are available in [23].

3.2. Simulation Aspects for/in Sonification

The sensorimotor control of the proposed gesture model enables the simulation of whole arm gestures from the specification of mallet extremity trajectories. As a result, the sensorimotor control and simulation of the gesture model provides kinematic motion parameters such as limbs positions and orientations (and their respective time derivatives), as well as physical motion parameters such as estimated torques, as represented on Figure 2.

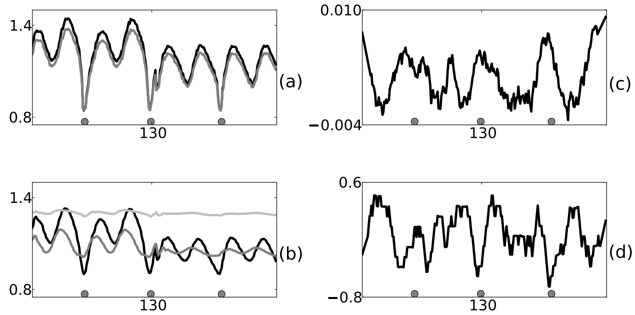


Figure 2: (a) Recorded (black) and simulated (grey) vertical trajectories of mallet extremity, (b) simulated vertical trajectories of wrist, elbow and shoulder joints (respectively black, dark grey and light grey), (c) simulated curves of wrist transversal-torque, and (d) elbow flexion-torque. Sphere markers on each timeline represent time occurences of beat impacts.

As an illustration of generated kinematic parameters, the pre-recorded height trajectory of the mallet extremity given as an input to our gesture model is depicted in Figure 2(a) (black line), in comparison to the simulated height trajectory (grey line). This shows an accurate tracking of the initial trajectory between the various beat impacts represented by the local extrema on the trajectories. Figure 2(b) presents the simulated height trajectories of arm joints: wrist (black), elbow (dark grey) and shoulder (light grey). Typically, one can notice the entrainment of wrist and elbow heights compared to the height of the mallet extremity, while the shoulder height is somewhat constant.

As for physical parameters, Figures 2(c) and 2(d) show torque curves, respectively for the wrist (transversal component) and the elbow (flexion component). These two torque curves reveal two behaviors which differ from their macro/micro action on the motion control. On the one hand, the flexion component of the elbow torque curve, Figure 2(d), shows a macro action in the sense that forearm movements show a sort of pendulum motion. During preparatory phases (between two beat impacts), this macro action is characterized by an increase of flexion torque during the first half for elevating the forearm, and a decrease during the second half since the weight of the forearm helps in gaining momentum in the view of the next beat impact. On the other hand, the transversal component of the wrist torque curve, Figure 2(c), shows a micro action in the sense that wrist torques seem to act on motion control around beat impacts. Wrist torques indeed rapidly increase slightly before impacts, remain almost constant during impact time, and rapidly decrease slightly after impacts. This micro action could for instance be explained by the need of fine control interaction mechanisms during beat impacts.

Combining such kinematic (limbs position, orientation) as well as physical parameters (torques) can finally be of great interest since all of these parameters are not easily visually accessible. Particularly in the case of estimated torques, while such parameters could be involved in playing strategies and variations in gesture articulations, findind an auditory representation appears to be a much more relevant approach, than a visual representation that

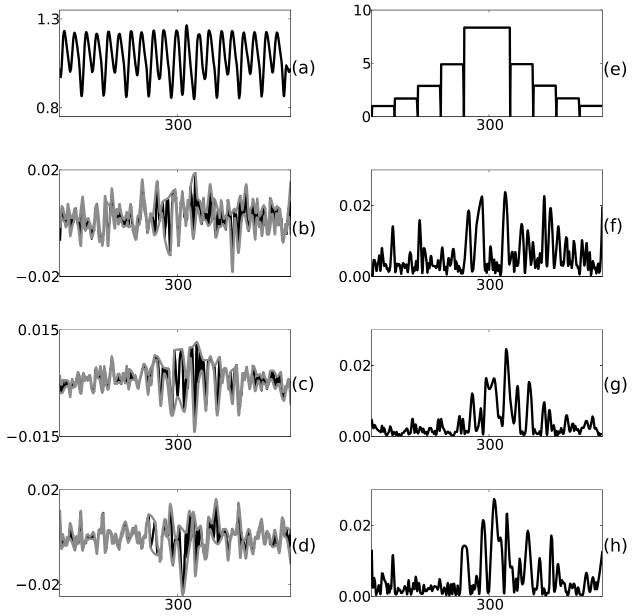


Figure 3: Investigation on the effect of mallet mass on wrist torques – simulation of a single roll played legato, mallet height extremity trajectory (a), under a mass variation of the mallet (e). Simulated curves of wrist transversal-torque (b), flexion-torque (c) and twist-torque (d) with their up-down envelopes, as well as their respective ranges (f-g-h).

would probably just overload the visualization.

As an illustration, we investigated this point by studying the effect of physics parameters of a simulated timpani performance on the generated torques. We specifically investigated the effect of a variation of the mallet mass on the generated torques at the wrist level. Figure 3 reflects this study in the case of a simulation of a single roll played legato, and especially shows wrist torque trajectories in relation with the mass variation. Figure 3(a) represents the height of the mallet extremity during the single roll simulation, while Figure 3(e) shows the piecewise increase/decrease of the mallet mass along the simulation, with a maximum mass factor of eight times the original mallet mass. Figures 3(b-c-d) show respectively the corresponding torque trajectories at the wrist level (for each component transversal-flexion-twist) in response to the mallet mass variation, as well as with their up-down envelopes. The difference between up-down envelopes, *i.e.* torque ranges curves, are represented for each component in Figures 3(f-g-h). Interestingly, these range trajectories reflect the torque increase for each component involved during the simulation as a means of counteracting the growing mass of the mallet. This same cause/effect relationship can also be seen during the decrease of the mallet mass, where torque ranges decrease progressively.

4. GESTURE SONIFICATION

The gesture sonification was designed by carefully mapping movements to sound parameters based on our earlier experience of movement sonifications [9]. We further decided to leave the control of some aspects of the sonification to the user. The guidelines of the mapping design can be summarized as follows:

1. Sonification of the motion data: all moving joints, shoulder, elbow, wrist as well as the mallet should be included into the sonification. The various moving parts should be as good as possible distinguishable in the resulting sound stream so that one can perceive the different movement patterns as acoustically well articulated and distinct. The sonification should lead to an *auditory gestalt* that allows to add complementary information from the simulation.
2. Since the movement directions from the timpani player such as up-down, back-forth and left-right correspond to the base vectors of the coordinate system, the joint positions (x, y, z) and their velocities can be directly mapped to sound parameters. The up and down movement should be acoustically distinguishable.
3. The amplitude of the velocity of each joint component was normalized for each channel over different gesture articulations. This decision was made since we wanted to make sure that even minute movements from the shoulder would be loud enough in order to compare it with the strong amplitude variations on the wrist or mallet. Furthermore the normalization factor was the same for the left and the right part of each component, so that asymmetries in the movement could be detected in the sonification. The scaling was performed over all the different gesture articulations.
4. The torques of shoulder, elbow and wrist should be integrated into the sound stream. Since the torques are vectors whose orientation is of interest, they should be mapped to a sonification parameter that still allows to attribute the resulting sound to the same joint.
5. The sonification of the left and right sides of the avatar should be distributed to the stereo panorama, as this mapping allows for a natural separation and distinction of the alternating dynamics of the arm movement.
6. The user should have control over relevant sonification parameters. More specifically the user should be able to either explore only the sonification of the movement patterns or mix them together with the sonification of the torques.

In the implementation of these guidelines, we decided for a mix between additive and subtractive synthesis schemes. For the velocity of the components (x, y, z) of each joint a sound was generated with a base frequency and a set of 5 overtones, whose gain was decaying with reciprocal to its order, $1/n$. The set of overtones ensured that the sound of each joint did not disappear in the overall sound stream, and could still be perceived as a tone with a defined pitch. The base frequency of the tone was modulated by the corresponding component of the torque. The gain of this sound was controlled by the velocity of the joint component. Analogous to a whitening procedure the mapping of the normalized data was given by velocity[0.0 ... 0.4] $\rightarrow$ gain [-60 ... 0] dB, using a linear, clipped mapping function with the given range limits. The mixing between additive and subtractive synthesis was controlled by the sign of the correspondent velocity component.

The implementation in the *SuperCollider* [25] sound synthesis program reads:

```
LinXFade2.ar(
    Klank.ar(‘[partials, gain, ring], noise, freq),
    Klang.ar(‘[partials, gain, phase], freq),
    mix, gain) * AmpCompA.kr(freq, freqmin);
);
```

Herein *LinXFade2* is a unit generator that fades between the two synthesis unit generators *Klank* for substractive synthesis and *Klang* for additive synthesis. The unit generator Klank implements an array of two pole resonant filter based on [26] where the bandwidth is given as 60dB ring decay time.

The input signal for the subtractive synthesis *noise* is pink noise. The parameter *ring* is an array with the 60dB ring decay time for each filter partial. The parameter *mix* ranging from -1 to 1 was mapped to the torques, while *gain* corresponded to the component velocity, both with flexible scaling.

Additionally we added a basic psychoacoustic amplitude compensation as provided by SuperCollider with the Class *AmpCompA*, implementing an ANSI A-weighting curve. Here *freq* is the frequency for which the amplitude is compensated for and *freqmin* is the lowest frequency for which the compensation function applies.

We developed two frequency schemes that were mapped to the 12 components of the moving joints. The frequencies were linearly mapped on the pitch scale with equidistant intervalls between 3 octaves and 7 semitones covering a frequency range between 100 and 1200 Hz. In the first mapping we grouped the frequencies with respect to joints starting with the shoulder's x, y, z component for the 3 lowest frequencies the middle ones to elbow and arm-wrist ending up with the 3 highest assigned to the x, y, z component of the stick.

The second scheme grouped the frequencies according to the x, y, z directions of movement, by assigning the 4 lowest frequencies to all the x components of shoulder, elbow, arm-wrist and stick. The middle and highest 4 frequencies to the y and z component respectively.

5. THE USER INTERFACE

For a better control of the sonification process we developed a minimal graphical user interface. In this interface a rendered video of the avatar was played back simultaneously with the sonification of the movement data, as depicted on Figure 4.

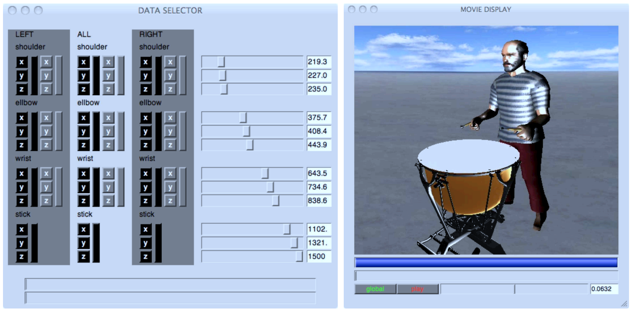
The sonification of each component for each joint can be switched on and off individually. For each joint all the 3 components can also be switched on simultaneously. The individual and grouped component for each joint can also be controlled for both sides of the timpani player, Figure 4(a). The amount of mixing between additive and subtractive synthesis can also be adjusted by a slider (bottom part of Figure 4(a)).

Figure 4(b) further shows that the playback range of the data could be set as well as the playback speed. A play button starts and stops the sonification. A second button next to it allows to change the scaling of the data from individual to global maxima. Additionally the two frequency mapping schemes can be selected through a toggle. In the background there is a fullscreen graphical display of the velocity of each component of the joint. Depending on what data is selected for sonification the corresponding part of the graphical display is highlighted.

Video examples of the application together with examples from the subsequently described user test can be found on the web: `http://tinyurl.com/2vo3k4w`

6. EVALUATION OF THE MULTIMODAL DISPLAY

While developing the application our assumption, that sonification together with a visual representation influences the perception



(a) (b)

Figure 4: (a) The data selection interface together with (b) the visualization of the avatar.

of the avatars' movements was supported through working experience. After some time of learning how to read the visual representation of the avatar, different gesture articulations could be distinguished. Sonification however seemed to highlight immediately the differences in the movement patterns. We were interested to evaluate the impact of sonification on untrained users' capability to discern gesture articulations and therefore set up a psychophysical experiment.

As a null hypothesis, we assumed that the display contained no relevant information which would help the subjects to distinguish different gesture articulations. Hence deciding wether two gesture articulations were the same or different would correspond to a discrete random process. By testing this on each of the three stimulus cases – *i.e.* auditory, visual and audio-visual individually – we wanted to find out in which particular way sonification contributes to the perception of the avatar's movements in the multimodal display.

6.1. The Experiment

The subjects were confronted with pairs of stimuli and were asked to rate if they perceived them as same or different. The stimuli consisted of 4 seconds of the the avatars movements, the starting point of which were slightly varied. Therefore the subjects could not base their decision on auditory or visual cues from the beginning or the end of the stimuli.

At the beginning of the experiment, the subjects were briefly instructed how to proceed by showing them one short excerpt of the kind of stimuli they would encounter. This instruction phase was kept short in order to avoid learning effects to set in. Further, the subjects were instructed that they would have to distinguish four different gesture articulations of the avatar: accent (A), legato (L), staccato (S), vertical accent (V).

The combination of all four gesture articulations gives 16 pairs, 4 of them with the same stimulus and 12 with different stimuli. In order to balance the ratio between same and different stimuli pairs we added the 4 pairs with the same stimulus a second time, giving a ratio of 12 different to 8 same stimuli pairs. For the sonification the second frequency scheme as described above was chosen. Playback speed was set to the original speed of the simulation. 12 tests subjects were recorded, 2 of them were female. The age ranged from 22 to 34 years. The three stimuli condition auditory, visual, and audio-visual were systematically varied. We therefore recorded two subjects for each of the 6 possible permutations of the 3 stimuli pairs. The sequence of stimuli within one condition was randomized. The whole experiment lasted about 20 minutes for a single subject. The visualization of the avatar was presented from within the application on a flat screen and the display of the time-series and the data-selection interface was minimized, the sound was displayed via AKG K 272 HD headphones. After the experiment the subjects were asked if they play an instrument and how the different presentation forms helped them to accomplish the task.

6.2. Results

The evaluation of the results from the experiment are compiled in Figure 5. The histograms show in the vertically axis a scalar value for similarity $(0-1)$ which was the mean rating of all subjects for a given stimuli condition. The first four vertical bars represent the cases where the same stimuli in one pair were compared by the subjects. The fact that none of these four combinations were rated unanimously as same is due to the variation in the beginning of the playback as explained earlier. This phase-shift seems to noticeably influence the perception and gestalt formation of rhythmic structures. The following 12 pairs are those where the subjects were confronted with different stimuli. They are ordered such that the combination AL and its inverse LA are next to each other. The null hypothesis represents the horizontal line at 0.5, which is equivalent to random guessing. The other dashed horizontal lines represent the 95% confidence interval. It differs for the first four cases since they appeared twice as often in the stimuli. Hence the total numbers of samples n and therefore the probability of deviating from the null hypothesis differed.

In the case of the auditory-only display, we find that the 4 stimuli pairs with the same gesture articulations received high mean values for similarity. Most of the pairs with stimuli of different gesture articulations received small similarity values, except the first AL and LA (outside the confidence interval) and the last two (SV, VS). In fact, these two pairs contained gesture articulations, which are very similar to each other. The similarity rating for both was around random rating except for the case LA, which had a tendency to be wrongly rated as similar.

In the case of the visual-only display the subjects had apparently more difficulties to recognize the same gesture articulations. Also different gesture articulations were recognized with less confidence except for the combination SL. Interestingly when comparing with the auditory-only condition, combinations AL and LA made of visually similar movements became less similar although found with ratings close to random rating. However VS was significantly wrongly rated as similar.

In the case of audio-visual display, we see how the good recognition of the same gesture articulations from the auditory part was adopted. Also the discrimination of the very distinct gesture articulations was apparently easy. The similar gesture articulation pairs (AL, LA, SV, VS), however could not be distinguished adequately compared to auditory-only and visual-only conditions. They were even more wrongly rated as similar.

The feedback from the subjects was very diverse. Most mentioned that they found it easier to use the sonification for the discrimination task. Some mentioned that they would have preferred to practice before the actual experiment. Subjects also reported that it was difficult to follow two features in the visual display at the same time, such as the movement of shoulder and hand wrist. One subject explicitly reported that the unimodal auditory and visual stimuli, were easier for the given task. In the multimodal case this subject said it seemed to have overheard certain acoustic differences.

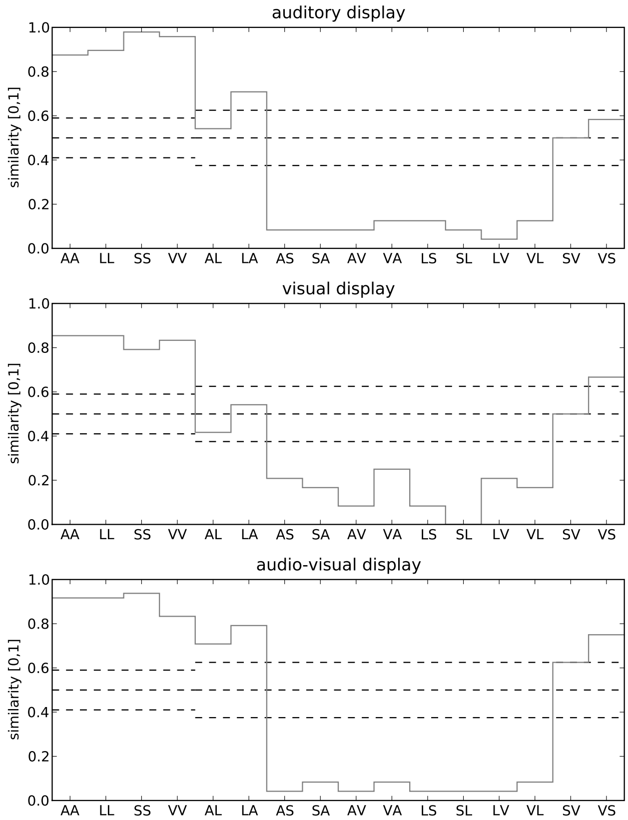


Figure 5: Histograms of the rating from the pairwise stimuli for the three conditions auditory, visual and audio-visual.

6.3. Conclusions from the Experiment

The evaluation of the experiment shows audio and visual displays contains relevant information for the subjects to partly succeed in solving the given discrimination task. The phase in which a repetitive movement is started seems to have a noticeable effect for the auditory as for the visual gestalt formation in perception. The combined visual and auditory information in the display however did not lead to a better discrimination, but rather emphasized the confusion of gesture articulations (AL, LA, SV, VS) detaining common features. Therefore similar gesture articulations had a tendency to be rated as same. Sonification seemed to influence the subjects' confidence to recognize gesture articulations as different, for those cases where it was more obvious. We think that this can be explained within the framework of cognitive load theory [27, 28], where evidence can be found that if we split our attention onto two perceptional modes, like auditory and visual, both of these modes have only limited resources and therefore have difficulties to perform well for a given task.

7. SUMMARY AND FUTURE WORK

In this paper we have introduced a method and tool for the sonification in combination with the visualization of simulated avatar movements. We have given examples where data from the simulation like torques can be interestingly integrated into the sonification concept. In a psychophysical experiment we investigated some basic properties of the presented audio-visual display. The results from the experiment show that sonification is at least as useful for discriminating movement patterns as the visual representation of the avatar. However the combined multimodal display does not provide the user with more clues to discriminate the movements.

Our results open up some interesting research questions to be answered in the future. The first one is if we can design a multimodal interface with a better perceptual integration of both visual and auditory modalities. If so, can similar gesture articulations be better differentiated? Furthermore, it would be interesting to find out how experience with the multimodal display changes with learning and how we make use of the auditory and visual information if they are presented at the same time. We also plan to evaluate systematically how the sonification of torques is integrated with the visual display, knowing that they are practically invisible in the movement itself. We further want to investigate how people make use of the data-selection interface in order to better understand the different gesture articulations. Here we are particularly interested how many data channels need to be sonified in order to evoke noticeable differences. This is important in so far as in a visual display we can decide where to look, whereas it is far more difficult to willingly isolate selected features in an auditory stream. This is where the interactive aspect of the sonification interface can have a beneficial role.

8. ACKNOWLEDGMENTS

The authors would like to thank all the people who have kindly accepted to take part to the user test experiments, as well as for all of their feedback. This work was partially supported by the *Natural Sciences and Engineering Research Council of Canada (Discovery and Special Research Opportunity grants)*.

9. REFERENCES

[1] M. M. Wanderley, B. Vines, N. Middleton, C. McKay, and W. Hatch, “The Musical Significance of Clarinetists’ Ancillary Gestures: An Exploration of the Field,” *journal of New Music Research*, vol. 34, no. 1, pp. 97–113, 2005.

[2] M. Nussek and M. M. Wanderley, “Music and Motion: How Music-related Ancillary Body Movements Contribute to the Experience of Music,” *Music Perception*, vol. 26, no. 4, pp. 335–353, 2009.

[3] S. Dahl, “Playing the Accent: Comparing Striking Velocity and Timing in Ostinato Rhythm Performed by Four Drummers,” *Acta Acustica united with Acustica*, vol. 90, no. 4, pp. 762–776, 2004.

[4] N. Rasamimanana and F. Bevilacqua, “Effort-based Analysis of Bowing Movements: Evidence of Anticipation Effects,” *journal of New Music Research*, vol. 37, no. 4, pp. 339–351, 2008.

[5] A. Bouënard, M. M. Wanderley, and S. Gibet, “Gesture Control of Sound Synthesis: Analysis and Classification of Percussion Gestures,” *Acta Acustica united with Acustica, Special Issue on Natural and Virtual instruments: Control, Gesture and Player interaction*, vol. 96, no. 4, pp. 668–677, 2010.

[6] G. Johansson, “Visual Motion Perception,” *Scientific American*, vol. 232, pp. 76–80, 85–88, 1975.

[7] T. Hermann, O. Höner, and H. Ritter, “Acoumotion – An Interactive Sonification System for Acoustic Motion Control,” in *Gesture in Human-Computer Interaction and Simulation*, N. Courty S. Gibet and J. F. Kamp, Eds. 2006, LNCS 3881, pp. 312–323, Springer Verlag.

[8] V. Verfaille, O. Quek, and M. M. Wanderley, “Sonification of Musician’s Ancillary Gestures,” in *Proc. of the International Conference on Auditory Display*, 2006, pp. 194–197.

[9] F. Grond, T. Hermann, V. Verfaille, and M. M. Wanderley, “Methods for Effective Sonification of Clarinetists’ Ancillary Gestures,” in *Gesture in Embodied Communication and Human Computer Interaction*. 2010, LNCS 5934, pp. 171–181, Springer Verlag.

[10] R. V. Miller, “ELECTROMYOGRAPHY – Uses and Limitations,” *California Medicine*, vol. 89, no. 4, pp. 250–252, 1958.

[11] S. Pauletto and A. Hunt, “Interactive Sonification of Complex Data,” *International journal of Human-Computer Studies*, vol. 67, no. 11, pp. 923–933, 2009.

[12] A. Effenberg, J. Melzer, A. Weber, and A. Zinke, “MotionLab Sonify: A Framework for the Sonification of Human Motion Data,” in *Proc. of the International Conference on Information Visualisation*, 2005, pp. 17–23.

[13] S. Guttman, L. Gilroy, and R. Blake, “Hearing What the Eyes See,” *Psychological Science*, vol. 16, no. 3, pp. 228–235, 2005.

[14] H. McGurk and J. MacDonald, “Hearing Lips and Seeing Voices,” *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[15] R. Sekuler, A.B. Sekuler, and R. Lau, “Sound alters Visual Motion Perception,” *Nature*, vol. 385, no. 6614, pp. 308, 1997.

[16] F. Avanzini and D. Rochesso, “Physical Modeling of Impacts: Theory and Experiments on Contact Time and Spectral Centroid,” in *Proc. of the International Conference on Sound and Music Computing*, 2004, pp. 287–293.

[17] M. Demoucron, *On the Control of Virtual Violins: Physical Modelling and Control of Bowed String Instruments*, Ph.D. thesis, Université Paris VI, France, 2008.

[18] R. Hänninen, L. Savioja, and T. Takala, “Virtual Concert Performance - Synthetic Animated Musicians playing in an Acoustically Simulated Room,” in *Proc. of the International Computer Music Conference*, 1996, pp. 402–404.

[19] S. Gibet and P. F. Marteau, “Gestural Control of Sound Synthesis,” in *Proc. of the International Computer Music Conference*, 1990, pp. 387–391.

[20] N. Rasamimanana, *Geste Instrumental du Violoniste en Situation de Jeu : Analyse et Modélisation*, Ph.D. thesis, Université Paris VI, France, 2008.

[21] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous Realtime Gesture Following and Recognition," in *Gesture in Embodied Communication and Human Computer Interaction*. 2010, LNCS 5934, pp. 73–84, Springer Verlag.

[22] E. Maestre, M. Blaauw, J. Bonada, E. Guaus, and A. Perez, "Statistical Modeling of Bowing Control Applied to Violin Sound Synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 855–871, 2010.

[23] A. Bouënard, *Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis*, Ph.D. thesis, European University of Brittany, France, 2009.

[24] A. Bouënard, M. M. Wanderley, and S. Gibet, "Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis," in *Proc. of the International Computer Music Conference*, 2009, pp. 255–261.

[25] J. McCartney, "Rethinking the Computer Music Language: SuperCollider," *Computer Music journal*, vol. 26, no. 4, pp. 61–68, 2002.

[26] K. Steiglitz, "A note on Constant-Gain Digital Resonators," *Computer Music journal*, vol. 18, no. 4, pp. 8–10, 1994.

[27] S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller, "The Expertise Reversal Effect," *Educational Psychologist*, vol. 38, no. 1, pp. 23–31, 2003.

[28] S. Oviatt, R. Coulston, and R. Lunsford, "When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns," in *Proc. of the International Conference on Multimodal Interfaces*, 2004, pp. 129–136.