User Interface Comparison for Virtual Drums

Teemu Mäki-Patola
MSc, researcher
Laboratory of Telecommunications Software and Multimedia, Helsinki University of Technology
+358 9 451 5849
tmakipat@tml.hut.fi

ABSTRACT

An experimental study comparing different user interfaces for a virtual drum is reported. Virtual here means that the drum is not a physical object. 16 subjects played the drum on five different interfaces and two metronome patterns trying to match their hits to the metronome clicks. Temporal accuracy of the playing was evaluated. The subjects also rated the interfaces subjectively. The results show that hitting the drum alternately from both sides with motion going through the drum plate was less accurate than the traditional one sided hitting. A physical stick was more accurate than a virtual computer graphic stick. Visual feedback of the drum slightly increased accuracy compared to receiving only auditory feedback. Most subjects evaluated the physical stick to offer a better feeling and to be more pleasant than the virtual stick.

Keywords

Virtual drum, user interface, feedback, musical instrument design, virtual reality, sound control, percussion instrument.

1. INTRODUCTION

We have been experimenting with virtual reality (VR) interfaces for control of physical sound models for some years. In addition for making prototypes [9] we have studied the musical effects of the user interface properties. See our project website for additional information and videos [1].

Relatively few virtual reality interfaces have been reported for sound control [3], [6], [9], [13]. Most of these have been interactive sound environments or interactive filters rather than standalone instruments. Burgess and Mynatt [3] report a virtual drum used with a tracked virtual stick and Borchers [2] reports an infrared baton interface for control of percussion intruments. There is no performance analysis in the articles.

Dahl and Bresin tested a percussion instrument without tactile feedback [5]. The research suggested that latencies of over 55ms degrade the use of a nontactile percussion instrument. Only four musicians were tested.

Lots of literature exists about designing musical interfaces [4], [7], [8], [9], [13]. However, better understanding of individual interface properties is needed. Quantitative interface comparison studies for sound control are few [12], [13].

The objective of this study was to find out the best interface for controlling a virtual drum without including tactile feedback.

2. SUBJECTS

16 students and researchers were chosen as test subjects. They were purposely chosen to be people with some musical background. 12 of them had at least six years of practice with one or more musical instruments. All subjects were right-handed. No subjects had prior experience of the tested instrument. Five subjects had experience of drums. Two subjects were female.

3. TEST SETTING

The test was conducted in a Cave-like virtual reality room. The virtual scene is back-projected on three walls and the floor of the room. The subjects see three-dimensionally (3D) through shutter glasses. 16 loudspeakers are located around the virtual room. The test setting is presented in Figure 1.

The subject's task was to match his drum hits as close to metronome clicks as possible. The drum and metronome sounds emerged from loudspeakers that were at equal distances.

The subject hit the drum either with a virtual stick drawn to his hand or with a real wooden cymbal stick. Both sticks are located by a tracker piece attached to the end of the stick's handle held inside the hand. The same collision model was used for both sticks. The drum was tuned to create a collision and a resulting drum sound when the cymbal stick's tip appears to collide with the virtual plate. The virtual stick was created so that its collision model fit this requirement. It had a little bit larger ball in its tip than the physical cymbal stick.

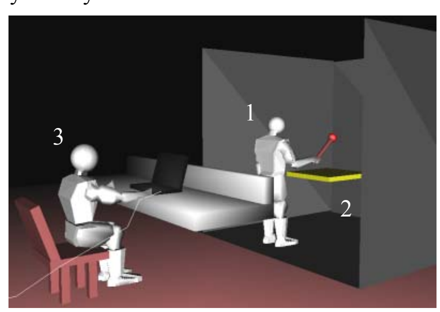

Figure 1. Test setting. The subject (1) plays the virtual drum (2) in the supervision of an instructor (3) who controls the test software.

The virtual drum object was one meter wide and long and five centimeters thick rectangular computer generated 3D plate located at the height of 1.2 meters. As the drum plate is a virtual object both sticks naturally pass through it unobstructed.

The location and orientation of the tracker pieces is sampled at the rate of 100Hz. The responsiveness time of the system from user action to sound reaction was measured to be 71ms with a standard deviation of 6ms. This is a high latency for instrument playing. However, our study is of a comparative nature.

The sound software [1] ran on 44.1kHz frequency generating all sounds in 2.9ms long buffers. It was controlled over a local network from the SGI mainframe running the virtual reality application. When a command for making a metronome sound and a drum sound were sent after each other the standard deviation of the starting time difference of the sounds was 3ms with a mean of 0.41ms. Thus, sometimes the messages were received so that one of them was delayed to the next buffer. The frame rate of the visual feedback was 30Hz during all tests.

A physical membrane sound model was used for generating the drum sound [11]. It was tuned to make a quickly decaying sound similar to hitting a tight membrane. Hitting velocity was mapped to the amplitude of the sound. The lead metronome click in a measure was a short rim-shot and the rest were clap sounds.

4. COMPARED INTERFACES

Three interface properties were studied: hitting style, visualization of the drum plate and having a real or virtual stick. As the performer can hit through the plate it was tested if two sided hitting would be reasonable for such a drum. The subjects hit the drum either just from one side, usually swiftly stopping their hand after each hit or from both sides in a more continuous motion. The drum's collision model reacted either only to hits with a downward velocity or to hits from both sides. Visualization of the drum plate tested how much it helped the subjects to see the drum compared to refining their hits based solely on aural feedback. The drum plate was either visible or the subject saw nothing and found the drum from the air using the aural feedback. The cymbal stick was either drawn to the subject's hand as a virtual object or the subject used a wooden cymbal stick. In the latter case no virtual stick was drawn.

Thus, the altered variables in the interfaces were visualization of the drum plate (present/absent), two different cymbal sticks (virtual/real) and the hitting style (one-/two-sided). All eight combinations were not tested. When the drum plate is not visible it is of little interest to visualize a virtual stick as it does not offer any useful feedback. Also the combination of visible drum plate, real stick and two sided hitting was not tested. The effect of the two sided hitting could be deduced already from the other two combinations where it was used. The remaining five combinations were tested. They will be referred to as VV1, VV2, VC1, IC1 and IC2. The first letter represents if the drum plate is visible or invisible (V/I). The second letter represents if the mallet is a virtual object or the wooden cymbal stick (V/C). And the third letter is a number representing if the collision model of the plate is one or two-sided (1/2).

To make the results less influenced by the metronome pattern the interfaces were tested on two rhythmically different patterns. The first rhythm was a steady 120BPM rhythm with a click interval of 500ms. The second rhythm was a discontinuous 80BPM rhythm with three faster hits each 375ms apart after which there was a pause of 750ms.

Five interfaces, two patterns on each, resulted in ten tests for each subject. To avoid subject fatigue the whole test situation was made to last at most half an hour. The metronome pattern repeated for the duration of each individual test. The recording of the data started after the first time that the subject hit the drum. Before that he could listen to the metronome as long as he wanted. When started the data recording continued for 50 seconds after which the sound model stopped to respond.

5. TEST PROCEDURE

The test was first explained to each subject. The interfaces were described and the subject was told to hit the drum sound as close to each metronome click as possible. The two sided drum was instructed to be played by rhythmically hitting through it from both sides in turns. All five interfaces were practiced each for one 50 seconds long test period. Three of the training periods were carried out on the continuous rhythm.

After the practice the test began. The subject remained inside the virtual room for the duration of the ten tests. Before each test he was told which interface he should pick for use.

The order of the ten tests was randomized for each subject by shuffling the names of the tests with Matlab. The subject was told to wear the data glasses during all tests regardless of if there was any visualization to keep the test situations as similar as possible.

After testing, the subject was interviewed for his musical background and asked to rate which interface he considered to be the best and why. Free comments were also encouraged.

6. RESULTS

Each of the ten test recordings for every subject contained about one hundred metronome clicks and matching drum hits. The processing time of each click and hit was recorded into a file. The recorded data was analyzed using Matlab. The closest drum hit was searched for each metronome click. If the closest hit was not inside a predefined time tolerance of 200ms the hit was marked as missed. A missed hit meant either too large time difference or that the subject stopped his hand before the mallet collided with the plate. By average less than two hits were missed in a test.

The time difference between the metronome click and the drum hit was evaluated for the matched hits. A subject's data then consisted of ten roughly hundred dimensional vectors marking the time differences between the onsets of the metronome clicks and the onsets of the drum hits. A negative value thus meant that the matching drum hit came after the metronome click.

Standard deviation (STD) was chosen as the main precision measure for performance. The smaller the STD the smaller the spread of the hits around the metronome clicks. Mean was chosen as another measure. By average it should be close to zero as people are known to match sound with sound. However, we wanted to see if there would be differences between the interfaces.

A box plot is a common way to present statistical distribution of variables. It marks the median with a horizontal line. The boxes extending above and below the median line mark the range that contains half of the values. The whiskers reaching from the boxes show the extent of the rest of the data. Outliers are marked outside the whiskers. The notches in the boxes represent an estimate of the uncertainty of the medians. Here box plots are used to present the test population distributions for the chosen variables.

6.1 Standard Deviation of Onset Differences

The STD values for the ten tests of each subject were evaluated stored into a 16 value population vector for each interface on both rhythms. Figure 2 presents box plots of the individual STDs for the different interfaces. The spread in the hits is clearly larger for the two-sided interfaces VV2 and IC2.

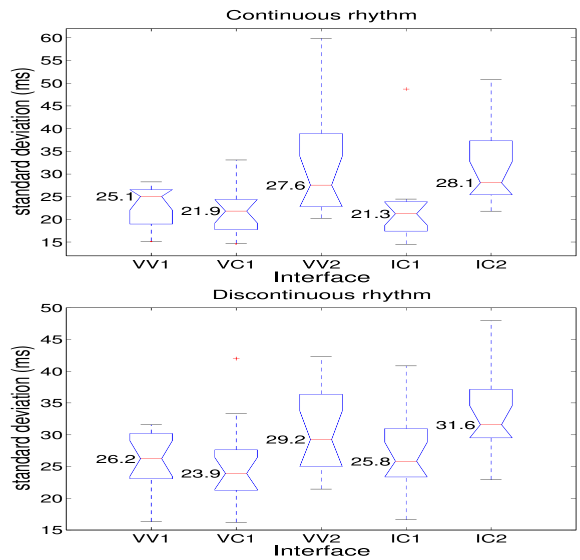

Figure 2. Box plots of the spread of the hits for each interface. The median values are written in.

For statistical analyses the interfaces were first compared within each subject. The interfaces were ranked in order of accuracy marking the most accurate with one and the least accurate with five resulting in a rank distribution for the population. This way also the poorly performing subject's STD values had the same weight in the population than the values of any other subject. The ranks were stored in 16 dimensional vectors. Box plots of the population ranks for each interface are presented in Figure 3.

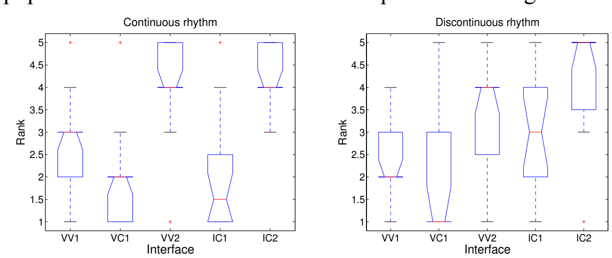

Figure 3. STD accuracy rank distribution box plots for both rhythms. Smaller rank is better.

As the ranks of the STDs do not follow a normal distribution, a nonparametric method (Kruskal-Wallis) was used to evaluate if the performance differences were significant between the interfaces. The Kruskal-Wallis method performs a nonparametric one-way analyses of variance (ANOVA) comparing the medians of two groups of data. It returns a probability (p-value) for the null hypothesis that the medians of the groups are equal.

To evaluate significant differences between the interfaces the rank distribution of each interface was compared pair-wise against all others. The common 5% risk margin was used for rejecting the null hypothesis. Rejection means that the median STD values of the compared interfaces differ significantly.

Combining the Kruskal-Wallis results with the box plot information of Figure 3 we have deduced if the medians that differ significantly are smaller or larger than that of the compared interface. The results are presented in Table 1. L/S in the table means that the interface on the left has significantly larger/smaller median than the interface on top. A minus sign means that there was no significant difference and the slash (/) separates the results from the two different rhythms. Continuous rhythm comparison results are before the dash.

Table 1. Results from pair-wise comparisons of the STD accuracy. The best performing interface is shaded.

	VV1	VC1	VV2	IC1	IC2
VV1	- / -	L / -	S / S	L / -	S / S
VC1	S / -	- / -	S / S	- / -	S / S
VV2	L / L	L / L	- / -	L / -	- / S
IC1	S / -	- / -	S / -	- / -	S / S
IC2	L / L	L / L	L / L	L / L	- / -

Table 1 shows eg. that the spread (STD) in the hits around the metronome clicks on interface VV1 is smaller than on the two-sided interfaces on both rhythms, larger than on the other one-sided interfaces on continuous rhythm and not significantly different with them on the discontinuous rhythm.

On the faster discontinuous rhythm the differences between the interfaces are smaller. The biggest difference compared to the continuous rhythm case is that interface IC1 does not perform as well. It does not differ significantly even from the double sided interface with visualization (VV2). IC2 is clearly the worst interface. Its medians are significantly larger than those of all other interfaces. By the spread of the hits VC1 performs the best on both rhythms but not significantly better than IC1 on both rhythms and VV1 on the discontinuous rhythm.

6.2 Mean Onset Differences

The same analyses done for the STD values were carried out also for the individual mean values. The means were calculated for each test and these were again ranked similarly for each subject.

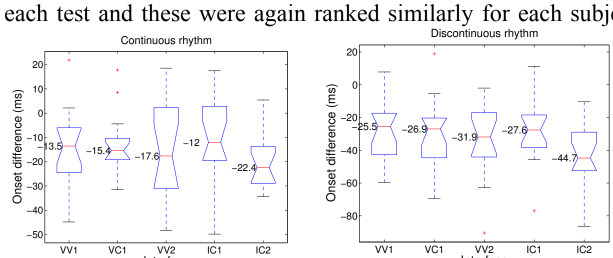

Figure 4. Box plots of the subjects' means for each interface.

Figure 4 shows the distribution of the means for both rhythms. There are no large differences between the interfaces as the extents of their data overlap strongly. The rank box plots do not give much more information and were thus left out of the paper.

A pair-wise Kruskal-Wallis analysis of the ranks produced only one significant difference on the continuous rhythm: IC1 had significantly better median rank than VC1. On the discontinuous rhythm IC2 was significantly worse than all other interfaces. No other significant differences were found.

6.3 Subjective Evaluations

88% of the subjects preferred the real stick over the virtual one. They felt that it was good to have something tangible in their hands. The real cymbal stick provided them with inertial tactile feedback compensating slightly the fact that the drum plate did not recoil when hit.

All but one subject preferred the one sided hitting over the two sided. Four subjects felt that the hitting style did not matter at all. Many speculated that the two sided hitting requires getting used to as it is quite different from the real world.

75% of the subjects preferred the visualization of the drum plate over the bare audio feedback. Yet, its importance was ranked low. The VC1 interface was ranked best by most subjects.

7. DISCUSSION

The subjects noticed the system latency clearly, especially with the virtual stick as also the visualization had latency and the stick followed slightly behind the real location. However, some subjects said that with the real stick and the visible drum plate the latency was almost unnoticeable. The stick was perceived to hit the plate when the sound came.

Some of the subjects who preferred the real stick estimated that they might favor the virtual one if the frame rate was higher and if the visualization had less latency. As a result of the virtual stick's latency the subjects tended to use it cautiously. They evaluated it to feel limp and sluggish and felt a need to make larger motions with it compared to the real stick. However, some subjects who complained it to feel disturbing in the beginning claimed to like it after getting used to. Interestingly the virtual stick performed well on the discontinuous more complex rhythm.

The real stick conflicts with the perception of depth as it occludes the wall images and is visible even when under the plate. Yet, most people preferred the interface with the real stick and the visible drum plate as the best one.

The continuous rhythm was so repetitive that the subjects quickly learned it kinaesthetically. In the case of the discontinuous rhythm the subjects likely needed to rely more on the feedback. The significantly smaller means for the steady rhythm seem to support this assumption.

The metronome click preceded the drum hits by 20ms by average. This is interesting as humans tend to match sound with sound when they get no tactile feedback. The rim-shot sound of the metronome started instantly. However, the clap sound contained five claps that started consecutively during 40ms. It seems that although the first clap started instantly the clap sound's perceptual center was in the beginning of the strongest clap 30ms in the sound and the subjects tried to match the drum hit to this location.

In the future we plan to include tactile feedback as physical objects and use motion prediction to compensate for the latency.

8. CONCLUSIONS

Five different interfaces were compared for controlling a virtual drum. Virtual meaning that the drum plate was not a real physical object and did not offer tactile feedback. Changing variables were hitting with a virtual or a real mallet, having visible or invisible drum plate and using one-sided or two-sided hitting style. The playing accuracy of each tested interface was measured.

For both rhythms tested, the smallest spread in the hits was produced with the real stick and a visible plate. Most subjects rated the real stick to offer the best feeling. It responded as they expected and gave a slight tactile feedback with its inertia.

The results suggest that hitting a virtual percussive objects with a real object rather than with another virtual object results in better temporal accuracy and is more appealing.

9. ACKNOWLEDGMENTS

This research was supported by Pythagoras Graduate School funded by the Finnish Ministry of Education and the Academy of Finland and by an EU IST program (IST-2001-33059). The author would like to thank also Perttu Hämäläinen, Tapio Lokki and Tapio Takala for help and ideas.

10. REFERENCES

[1] ALMA project instruments page: (visited April 2004) http://www.tml.hut.fi/~tmakipat/alma/almawebisivu/HUTTMLIndex.html

[2] Borchers, J. O. WorldBeat: Designing a Baton-Based Interface for an Interactive Music Exhibit. *Proc. CHI.* 1997.

[3] Burgess, D., Mynatt, E. 3-D interactive percussion: the virtual drum kit. *Proc. CHI.* 1994.

[4] Cook. P., "Principles for Designing Computer Music Controllers", *NIME Workshop – CHI*, 2001.

[5] Dahl, S. Bresin, R. Is the Player More Influenced by the Auditory Than the Tactile Feedback from the Instrument. *Proc. Conference on Digital Audio Effects 2001*.

[6] Lanier, J. "Virtual Reality and Music" http://www.advanced.org/jaron/vr.html (visited 20.1.2005)

[7] Hunt, A. 1999. *Radical User Interfaces for Real-time Musical Control*. PhD Thesis, University of York UK.

[8] Machover, T. Instruments, Interactivity, and Inevitability. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME02)*. 2002.

[9] Mulder, A. *Design of Virtual Three-Dimensional Instruments for Sound Control*. PhD Thesis, Simon Fraser University. 1998.

[10] Mäki-Patola, T., Kanerva, A., Laitinen, J., and Takala, T. Experiments with Virtual Reality Instruments. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME05).* Vancouver, Canada, May 2005.

[11] Trautmann, L., Petrausch, S., Rabenstein, R. "Physical Modeling of Drums by Transfer Function Methods" *Proc. Int. Conf. on Acoustics, Speech & Signal Processing (ICASSP)*, Salt Lake City, Utah, May 2001.

[12] Vertegaal, R., Eaglestone, B. (1996). Comparison of Input Devices in an ISEE Direct Timbre Manipulation Task. *Interacting with Computers 8*, 1, pp.113-30

[13] Wanderley, M., Orio, N. (2002) Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, 26:3, pp. 62-76, Fall 2002.

[14] Wanderley, M., Battier, M. (2000) Eds. *Trends in Gestural Control of Music*. Ircam - Centre Pompidou - 2000.