

# Latency Tolerance for Gesture Controlled Continuous Sound Instrument without Tactile Feedback

Teemu Mäki-Patola\* and Perttu Hämäläinen†

Laboratory of Telecommunications Software and Multimedia, Helsinki University of Technology

\*tmakipat@tml.hut.fi

†pjhamala@tml.hut.fi

## Abstract

*This paper reports the results of an experimental study of human latency tolerance for gestural sound control without tactile feedback. 16 subjects played a Theremin that was routed through an adjustable delay buffer. It was found that when comparing to a reference with no latency, the just noticeable difference (JND) is between 20 and 30 milliseconds. After that, the probability of detecting latency increases linearly. It was also found that playing style strongly affects the detection of latency. When playing slow passages with vibrato, even high latencies were not noticed. The results also suggest that younger subjects detect latencies more accurately than older subjects. However, the subjects' activity with music and musical background did not seem to have an effect.*

## 1 Introduction

Currently, physical sound modeling is an active research area. Real-time sound production makes it possible to alter any parameter of the sound model while playing. This creates a need for controllers whose input flexibility matches the complexity of the sound model. Virtual reality input technology, such as data gloves and location/orientation trackers with gesture analyses, is one option that offers several degrees of freedom. We are currently experimenting with virtual reality interfaces for the control of physical sound models in a EU funded project called ALMA (2002).

Physical sound models are often computationally heavy, which introduces some latency. Additionally, a virtual reality system's reaction time to the user's actions always adds to the latency. Latency is also a key issue in networked co-operative playing. It is important to know the amount of latency that can be allowed for different control paradigms.

An article by Paradiso (1998) and a book edited by Wanderley and Battier (2000) offer a good overview of existing electronic interfaces and controllers. Many have been created, especially during the last few decades. However, only a few virtual reality interfaces for sound control have been made (Choi 2000, Mulder 1998). These interfaces have been interactive sound environments or

interactive filters rather than standalone instruments. The interfaces and alternative controllers have been reported mostly as case studies.

There seems to be a lack of quantitative comparisons of the suitability of different interfaces for controlling sound. A recent article by Wanderley and Orio (2002) offers some methodologies for evaluating input devices for musical expression. The importance of parameter mapping has only lately been considered (Hunt, Wanderley and Paradiso 2002, Hunt, Wanderley and Kirk 2000). Vertegaal and Eaglestone (1996) have made a comparison of three input devices for timbre space navigation. The preliminary parameter mapping observations offer some suggestions on the direction to proceed in. It would be beneficial to have similar guidelines based on the properties of available input technology and its suitability for control of different kinds of sound.

Earlier research suggests that tactile feedback improves the playing accuracy of an instrument (O'Modhrain 2000). Rovin and Hayward (2000) suggest use of a vibrotactile simulator as one possibility for including tactile feedback to open-air controllers. Still, tactile feedback is difficult to elegantly integrate into virtual reality interfaces. Thus, if we want to use virtual reality for controlling sound it is valuable to have an estimate for latency tolerance also in cases where the performer does not obtain tactile feedback while playing an instrument.

Several studies have shown that latency degrades user performance in virtual reality (Ware and Balakrishnan 1994, MacKenzie and Ware 1993). The degradation is gradual and depends of the task. The studies mentioned above concentrated on tasks of acquiring and reaching for targets. Feedback was visual and minimum latencies were higher than the maximum latency in our test. Similar results by Watson et al. (1999) show that latency slows down placement time and reduces placement accuracy when the task requires feedback. Watson et al. also studied the effect of variations in latency (1998), concluding that only variations with a standard deviation of above 82ms affect performance in a grasping and placement task.

A classical experiment conducted by Michotte and reported by Card, Moran and Newell (1983) shows that

humans perceive two events as connected by immediate causality if the delay between the events is less than 50ms.

Dahl and Bresin (2001) suggest that latencies of over 55ms degrade the performance of playing a percussion instrument without tactile feedback along with a metronome. The degradation was gradual. Only four professional musicians were tested with a baton instrument. The latency was increased in small steps while playing. Two subjects were also tested with tactile feedback (a MIDI drum), concluding that the standard deviation of the flutter of consequent hits increased with longer delay times - but again, the change is slow. The standard deviation is no larger at the latency of 50ms than at zero, but after 50ms it seems to rise gradually. However, the amount of subjects and samples in the test was small for strong conclusions. The study also verified a hypothesis that, when a performer has to synchronize their playing with other audio sources, they attempt to compensate for the delay by matching sound with sound. Finney has shown that delay in auditory response caused large errors in performance of pianists (1997). The main source of these errors seems to be discrepancy between sound and tactile feedback. There was no degradation if the performer did not receive auditory feedback.

A study by Sawchuk et al. (2003) shows that latency tolerance is highly dependent on the piece of music and the instrumentation. Collaborative playing over a networked system was researched. Somewhat surprisingly, the performers tolerated latencies of 100ms with a piano sound but only 20ms with an accordion sound in the same piece.

As professional piano players may perceive latencies of under 10ms, this amount is often suggested as the maximum latency for a music controller (Freed et al. 1997, Wright and Brandt 2001). However, latency tolerance is dependent on the type of music, the nature of the instrument's sound and the presence or absence of tactile feedback. For an extreme perspective, let us remind that latencies as high as several hundred milliseconds are not unusual for church organs, and they can still be played with practice.

For developing the ALMA project interfaces, we need to know the tolerable length of latency for instruments without tactile feedback. Dahl and Bresin's study offers a rough estimate for percussion instruments. Our study searches for a latency threshold for a continuous sound instrument.

### 3 User tests

The goal of our user tests was to find a threshold for noticing latency in a gesture controlled continuous sound instrument without tactile feedback. We tested for just noticeable difference (JND), which is the limit after which the answer distribution is clearly not achieved by guessing (Goldstein 1999). Our assumption is that if the performer does not perceive a nonzero latency in close comparison with zero latency, he is not likely to be distracted by it during a solo musical performance.

The tests consisted of series of playing tasks in which the subject was asked to reproduce an example passage played back by the test controller. In each task, each example was to be reproduced at two different latency settings, one of which was always zero latency. The subject then compared the two latency settings and evaluated which of the two settings had larger latency. The whole test contained 39 comparison pairs for each subject. The pairs were the same for all subjects but their order was randomized.

### 3.1 Subjects

16 students and researchers from the Helsinki University of Technology were chosen as test subjects. 14 subjects had at least six years of practice with a musical instrument, 10 subjects had more than 10 years of practice with several instruments, and three subjects had a music teacher's qualification. Five subjects practiced more than five hours per week, six subjects practiced for 1 to 3 hours and the remaining five practiced less than an hour or not at all. 10 subjects were 23 to 28 years of age, the rest were 30 to 50 years of age. 14 subjects were male. 15 subjects were right-handed. None of the subjects had prior experience with the Theremin instrument.

### 3.2 Test equipment

The test was conducted on an analog, solid state Theremin (Enkelaar). The Theremin is an instrument whose sound is controlled by the distance of the performer's hands from two antennas. The right hand's distance from a pitch antenna defines the pitch of the instrument's sinusoidal sound. The left hand's distance from a loop antenna controls the volume. The performer is in no physical contact with the instrument while playing.

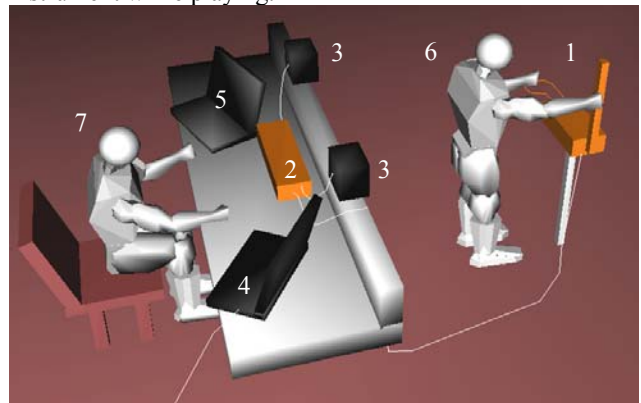


Figure 1. Test setting. The test subject (6) plays the Theremin (1) facing away from the controller (7). The Theremin's sound goes through the Effects Processor (2) and comes out of the speakers (3). One computer records test data (5), and another one (4) makes the example patterns. The example sound speakers were a few meters to the left of the subject (not visible in the figure).

The Theremin's output was routed to a Boss GX-700 Guitar Effects processor. Using the effects processor, the instrument's sound could be delayed for a specified amount of milliseconds. The effects processor was preprogrammed with patches that had only a delay effect active and no direct sound. The patches could be selected quickly, and the delay activated and deactivated with the press of a button. Our tests showed that the effects processor itself produced a delay of less than 1ms. The output of the effects processor was connected to a pair of loudspeakers (see Figure 1).

### 3.3 Test range

Before the user tests the authors experimented with two subjects to estimate a range where the JND threshold would be. The two subjects had a musical background and did not participate in the actual user tests. These preliminary tests suggested that latencies of over 60 ms were detected most of the time and latencies of 30 ms and less were not detected. Not detecting a latency means that the comparison test gives roughly an equal amount of correct and false evaluations.

A user test should not last more than 30 minutes to prevent fatigue from affecting the results (ITU-R 1997). Allowing every subject 10 minutes of practice with the instrument there were 20 minutes left for the comparison tests. We estimated each comparison to take about 30 seconds. This limited the maximum amount of comparison tests per subject to about 40.

We created three different sound examples for the tests: a broken chord (C4, E4, G4, C5), an alternation (C4, G4, C4, G4, C4, G4) and a slower passage (C4, G4, C4) where each note is faded in and out and played with vibrato. The vibrato used in the example sound was a 6Hz halftone vibrato. The first two sound examples consisted of quarter notes and the last example consisted of half notes in a 120 BPM tempo. We included the slower example to find out how playing style affects the just noticeable latency.

Based on the information from the preliminary test we constructed a test set of 39 comparison pairs. We chose eight latencies to be tested: 10, 20, 30, 40, 50, 60, 70 and 100 milliseconds. The preliminary tests suggested the JND to be between 20ms and 60ms. Thus, we emphasized that range by more comparisons. The amount of comparisons for each of the eight latencies is presented in Table 1. Latencies of 10ms, 70ms and 100ms were tested with one repetition on each subject on each of the three sound examples, the rest of the latencies with two repetitions.

As there was a 50% probability for making the right evaluation by chance, it was not unusual to get even all six samples per latency setting correct by guessing. The probability for this is 2%, 11% for guessing five or more correct. As the duration of the test should not exceed 30 minutes it was not possible to create accurate individual statistics of each subject's JND. However, from the 624 samples of the 16-subject population statistically accurate results could already be drawn.

### 3.4 Procedure

In the beginning of each user test, the subject was interviewed for his musical background. He was asked what instruments he plays, how many years of practice he has and how many hours in a week he has practiced during the last four years.

Latency	Comparisons per subject	Total comparisons
10	3	48
20	6	96
30	6	96
40	6	96
50	6	96
60	6	96
70	3	48
100	3	48
All latencies	39	624

Table 1. The amount of comparisons for each of the tested latencies. The preliminary tests suggested the JND threshold to be on the range of 20ms and 60ms. Thus, this range was emphasized with more samples.

After the interview the Theremin instrument was introduced to the test subject. The subject was given 10 minutes to play and experiment with Theremin without any latency. After the practice, the test was explained and the sound examples were introduced by playing each example once. The effect of latency was demonstrated to the subject with two practice comparisons. The two comparisons compared latencies of 170ms and 100ms with zero latency.

By zero latency we mean that there is no delay added by the effects processor. The performer is part of the solid state Theremin's oscillating circuit. Thus, the instrument's sound reacts immediately to motion. However, the speakers were about one meter behind the performer causing approximately 3ms of additional delay resulting from the limited speed of sound.

All subjects detected the latency of 170ms correctly. Five of the subjects made a mistake on the practice latency of 100ms. The subjects were told that the test consists of 39 similar comparison pairs with smaller latencies. After this, the test began.

In the beginning of each comparison, a sound example was played by a computer using a sinusoidal sound similar to Theremin. After listening to the example, the subject reproduced it on Theremin. He played the example a few times on the first latency setting (A). When ready, the subject notified the controller to switch for the second latency setting (B) and played the sample a few times again. After reproducing the sample on both latencies the subject evaluated (forced choice) which one of the latency settings (A or B) had larger latency. To eliminate time-consuming iteration, the subjects were not allowed to test again on the first setting after changing to the second.

It was randomly selected which one of the latencies in a comparison pair was zero. The other latency was one of the eight tested latencies. The test subjects had been told only that one of the latencies was larger than the other one.

After answering which of the settings had larger latency, the subject was asked to rate how certain he felt of his answer on a one to five scale. The subjects were asked also to give a subjective estimate if the larger of the latencies made playing more difficult. The questions asked after each comparison are presented in Table 2.

Question	Possible answers
Which setting had larger latency, A or B?	A B
How certain are you that your answer is correct?	1 just a guess 2 very uncertain 3 uncertain 4 somewhat certain 5 very certain
Did the larger of the latencies make playing the sample more difficult?	Yes No

Table 2. Questions asked after each comparison test. The results were collected using a computer program that formatted them to a convenient form for data analysis in MatLab.

## 4 Results

The results indicate that only the latencies of 10ms and 20ms were not detected at all. For them, the percentage of correct answers from the whole population does not significantly deviate from guessing. Statistical analyses show that latencies of 30ms and above are detected. Thus, the just noticeable difference at the accuracy of our test is 30ms. The percentage of correct evaluations does not make any dramatic jumps, but instead rises almost linearly as a function of latency, as shown in Figure 2.

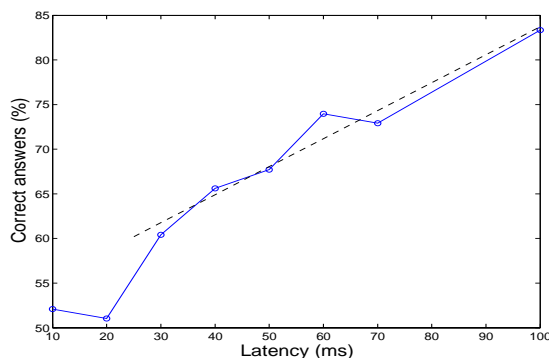


Figure 2. Percentage of correct answers in the whole population as a function of latency. Probability for answering correctly rises linearly after the JND threshold of 30ms. The dashed line in the figure is a linear regression model of the data.

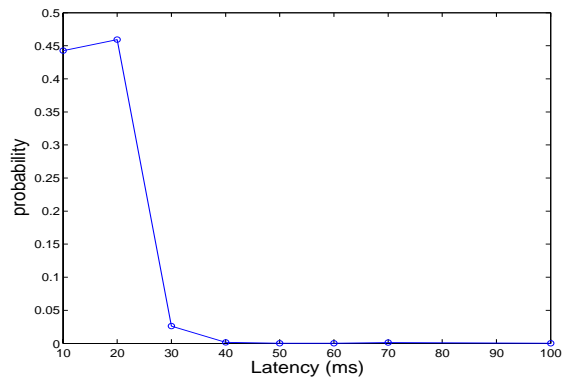


Figure 3. Binomial probability that the answer distribution can result from guessing. At 30ms the probability falls to 2.6% and approaches zero on larger latencies.

A latency of 30ms is detected correctly 60% of the time. To verify the detection of latency statistically, we calculated the probabilities for getting the resulting answer distribution by guessing (Figure 3), by using binomial probability. For example, in the case of 30ms, we calculated the probability of getting 58 correct answers out of 96 comparisons by guessing.

The probabilities of detecting latencies of 10ms and 20ms are not statistically significant but 30ms already is. We used the common risk margin of 5% for our hypothesis. After 30ms, the probabilities for getting our results by guessing are well below the risk margin.

The data obtained from individual subjects contained a fair amount of noise. Only four subjects had 15 or more answers out of 18 correct on the 50ms to 100ms range. As data from 10ms and 20ms samples did not deviate from guessing we have calculated regression lines for the figures by using results only from the last six sample points.

### 4.1 Effect of playing style

The sound samples that the subjects tried to reproduce contained two fast samples and one sample with slower changes and vibrato. Our hypothesis was that latency is harder to notice with the vibrato sample.

The hypothesis was verified even more strongly than expected. Figure 4 shows the correct answer percentage for each of the three sound examples, indicating that none of the tested latencies were clearly noticed when playing the vibrato sample. On the broken chord example, topmost in Figure 4, no one did a mistake on 100ms of latency.

Because no latencies were detected on the third sound example, we continue by analyzing data only from the first two sound examples. Figure 5 shows the combined answer distribution for the first two sound examples. The binomial probabilities of Figure 3 were very similar for the data of sound examples one and two. 10ms and 20ms were not detected, 30ms already was.

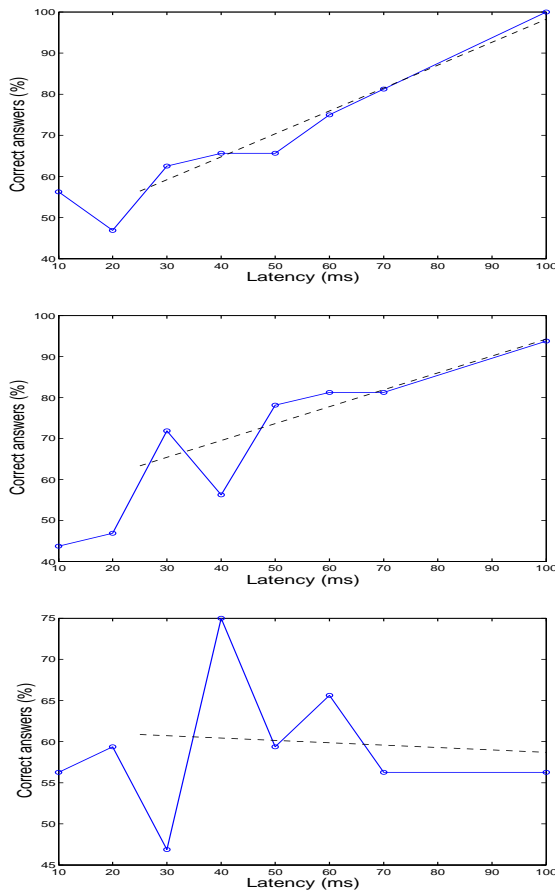


Figure 4. Percentage of correct answers for the different sound examples (top: broken chord, middle: alternation, below: slow notes with vibrato). The latencies were not detected while playing the slower vibrato.

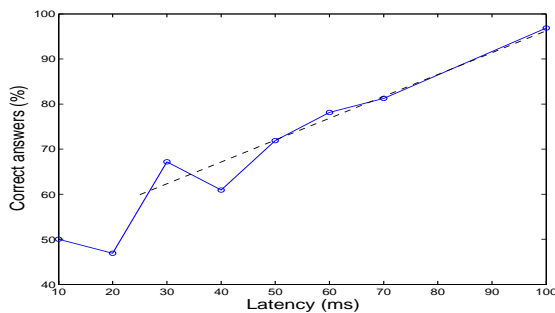


Figure 5. Percentage of correct answers from comparisons using only the two faster sound examples.

## 4.2 Correlation with subjective certainty

The percentage of correct answers of those of which the subject was certain or somewhat certain was 11% higher on average than of all answers (Figure 6). The latency detection correlates significantly with subjective certainty. The anomaly at 10ms is caused by there being only four certain

or somewhat certain answers. Three of these are correct, but this happens easily by chance. Out of all answers for each latency, only a small fraction were certain or somewhat certain (Figure 7). Below 60ms, more than four fifths of the answers were uncertain. Some subjects never felt certain of their answers.

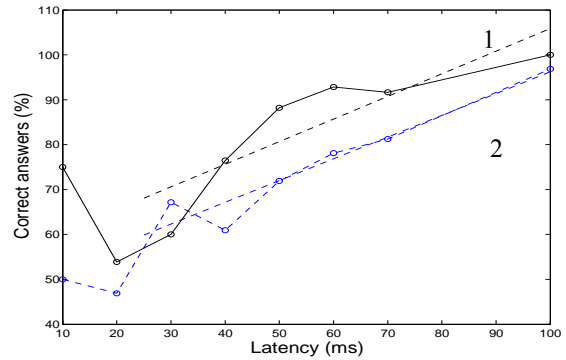


Figure 6. The percentage of correct answers of those of which the subject was certain or somewhat certain (1) vs. percentage of correct answers from all answers (2). At the range of 30ms to 100ms, the first curve is 11% higher than the curve of all answers.

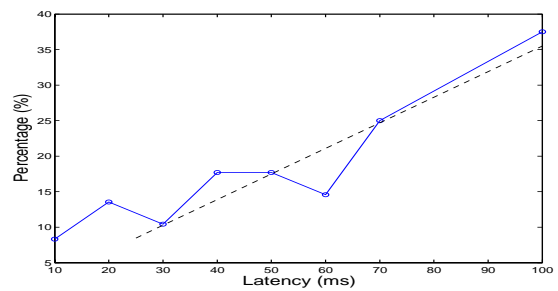


Figure 7. Percentage of certain or somewhat certain answers from all answers as a function of latency.

## 4.3 Effect of musical activity

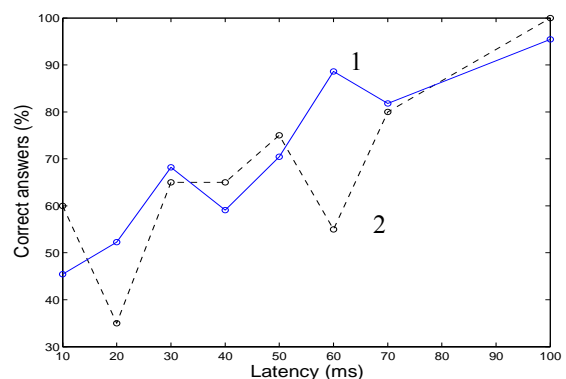


Figure 8. Percentage of correct answers for the five subjects who practiced five or more hours per week (1) vs. the 11 subjects who practiced only few hours or not at all (2).

Figure 8 suggests that there was no significant difference in detecting latency between the musically skilled subjects and those less skilled. This was a slightly surprising result. However, it should be noted that almost all subjects had had at least six years of training with a musical instrument at some point in their lives. The data from the two subjects that had not had any practice with an instrument were not enough for significant statistics.

There is a strong anomaly in Figure 8 at 60ms. At that latency, the difference of correct answers between the two groups is 34%. A 5% significance t-test rejects the null hypothesis that the two groups would be equally accurate at that latency ( $p=0.002$ ). With the other latency values, the null hypothesis is valid. As we have no explanation for the gap at this latency, we assume that it is a result of random chance – despite the 0.2% probability for that.

#### 4.4 Effect of age

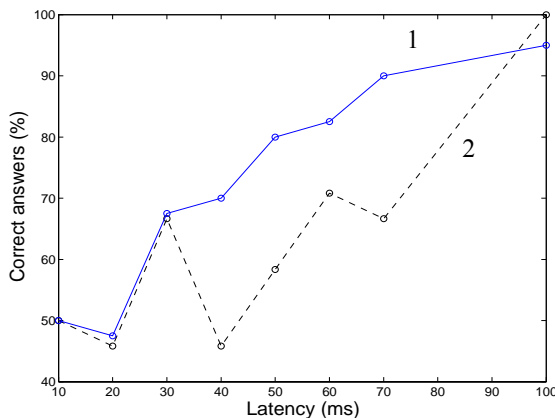


Figure 9. Percentage of correct answers with subjects of less than 30 years of age (1) and those over 30 (2). Young subjects are significantly better at detecting latency than the older ones.

Figure 9 compares the percentage of correct answers from subjects over 30 years of age to subjects under 30 years of age. Six subjects were over 30 years old with an average age of 36.3 years. 10 subjects were under 30 years of age with average age of 25.4 years. Both groups had similar musical background.

On average, the percentage of correct answers was 13% higher by younger subjects than by the older subjects. In the range of 40ms to 70ms their percentage was 20% higher. A t-test rejects the null hypothesis that the two groups would be equally accurate ( $p=0.008$ ,  $p=0.001$  in the range of 40ms to 70ms). Thus, we conclude that younger subjects notice latencies significantly better than older subjects. Strangely, the older subjects noticed the 30ms latency equally well and deviated only on larger latencies. However, as there are only six older subjects the probability to get their resulting 30ms answer distribution by random guessing is still 8%. Thus, the high value can be by chance. However, the probability

for the consistently lower scores, in the range of 40ms to 70ms, to result from random chance is well below the 5% risk margin.

#### 4.5 Subjectively disturbing latencies

Figure 10 shows that only a small fraction of latencies was rated disturbing for playing. Some subjects never rated any latency as disturbing. However, the fraction clearly rises as a function of latency as expected.

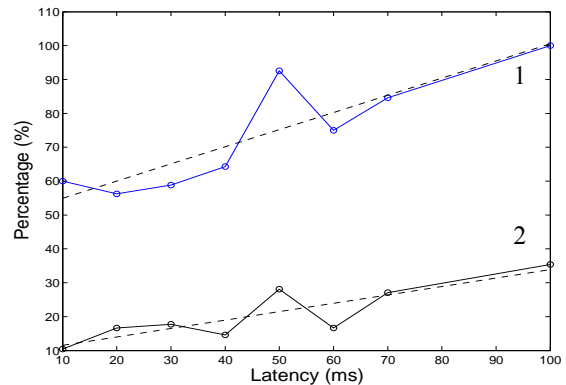


Figure 10. Correct answer percentage among latencies rated disturbing for playing (1) and the percentage of all tests rated as disturbing on each latency (2).

#### 4.6 Possible sources of errors

Some possible sources for errors could not be eliminated from the results. For instance, different subjects had different playing styles. Most subjects played with their pitch hand open but a few played with a fist. Some subjects used the volume control hand more than other subjects. Some subjects used more time for playing before giving their answer. There were subjects who played the example only once or twice for each comparison and those who played it several times. Some subjects concentrated more on getting the notes correct than others although they were told to concentrate on the latency. Thus, they played a bit slower. A few subjects played constantly in a slightly faster tempo although they were told to follow the tempo of the example sounds. Subjects seemed to be also annoyed to answer that the latencies did not disturb their playing several times in a row. Sometimes subjects played from higher or lower notes compared to the example. Because the Theremin's range is less dense in the low pitch end, a lower sound may seem to react more slowly although that there is no difference in latency. The same person controlled all user tests so there should be minimum errors resulting from differences in conducting the tests.

It is important to note that the test subjects did not have prior experience of the Theremin instrument. It seems likely that a skilled performer who has years of practice with Theremin would detect even smaller latencies than our test subjects. However, our assumption was that latencies, which



are not detected, are not likely to distract a performer who is new to such an instrument. If the performer practices several years or plays together with other instruments even the now undetected latencies might become an issue.

The majority of the subjects thought that the test was difficult. Most of the time the subjects felt uncertain of their answers. It was not unusual for the subjects to answer that both of the compared latencies were noticeable although that one of them was always zero. It even happened that a subject claimed both of the compared latencies to be disturbing, even when one was zero and the other one was as low as 10ms. A few comparisons later the same subject rated a latency of 60ms as unnoticeable and to feel clear and immediate. However, the noise in the data was mostly filtered away by large population size and by randomizing the comparison order. The trends are clearly visible from the analyzed data.

## 5 Future work

A natural continuum for this paper is to study how latency affects the playing accuracy. It seems likely that small latencies, such as 30ms, do not impede the playing of gesture controlled continuous sound instruments that lack tactile feedback. It will be interesting to see how much noticeable latencies increase playing errors. We want to establish knowledge for creating new instruments. Thus, it is good if the test subjects do not have prior experience of the instrument without latency. In Finney's test for pianists (1997) all of the subjects had trained years with immediate response piano. Introducing then latency changes the whole instrument for them.

It would be interesting to know also the effect of background rhythm to latency tolerance. In the ALMA project, we are making also percussion instruments with virtual reality interfaces. We plan to study how tactile feedback could be incorporated into the interfaces. We will test dummy objects of which place is known also in the virtual reality presentation. This kind of "augmented virtual reality" would contain tactile feedback and yet have the degrees of freedom of a virtual reality interface. Especially in the case of percussion interfaces, being able to hit something concrete is likely to help attain better temporal accuracy.

Dahl and Bresin (2001) used both a tactile and a nontactile percussion interface. However, they did not compare the results. The study had also too few subjects for strong conclusions. In our tests it is easy to keep the interface otherwise the same but add tactile feedback. By testing users on both we can compare how tactile feedback affects the temporal accuracy and latency tolerance.

Virtual reality interfaces always introduce some latency. However, they might still suit even for percussion interfaces. The inertia of the hand makes it possible to predict the motion some 50ms ahead. An article by Dahl (2000) reported a study of professional percussionists. The

results show that averaged flutter ranged between 10 and 40ms, between 2-8% of the associated tempo on professional musicians. This suggests that the latency variance of the virtual reality system may not be a large problem as the gaps between hits flutter much even on their own.

Virtual reality instruments are not likely to become appreciated by professional musicians before the technology gets faster, better and affordable. However, as virtual reality technology will eventually come to ordinary homes it is interesting to study how it could be used for controlling sound synthesis.

## 6 Conclusions

It was found that the just noticeable difference (JND) for latency in a continuous sound instrument without tactile feedback is about 30ms. Interpolating the binomial probability distribution presented in chapter 4 and adding the 3ms delay resulting from the limited speed of sound, we estimate the exact threshold for the tested population to be much closer to 30ms than to 20ms. After the threshold the probability of detecting latency rises gradually.

Playing style strongly affects the detection of latency. While playing slowly with a vibrato, latencies of even 100ms were not detected. Such playing is characteristic for the Theremin instrument used for the test.

The results suggest that younger people notice latencies better than older people. By average, their detection probabilities were 13% higher. However, it seems that the musical background of the subjects does not have a significant effect on the JND. We did not get results for women vs. men as we had only two female subjects. The one left-handed subject was allowed to turn the Theremin around, thus switching the pitch and volume hands.

To see the effect of learning during the test, we divided the data into two groups. The groups consisted of the results from the first 19 comparisons and from the last 20 comparisons of each subject. We found that the latency of 50ms was not detected during the first half of the test but it was well detected during the second half of the test. The latency of 60ms behaved similarly but the effect was not so strong. As was seen in chapter 4, there were also other anomalies that were difficult to explain around the latencies of 50ms and 60ms. It could be that these time constants have some special characteristic in the human physiology, but from our part this is a matter of further research.

## 7 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 authors would like to thank also Tapio Takala and Tapio Lokki for guidance, help and ideas and Aki Kanerva for proof reading.

## References

- ALMA project home page (2002):  
<http://ftp-dsp.elet.polimi.it/alma/>
- Card, S. Moran, T. Newell, A. (1983) "The Psychology of Human- Computer Interaction." *Lawrence Erlbaum Associates*.
- Choi, I. (2000) "A Manifold Interface for Kinesthetic Notation in High-Dimensional Systems." *Trends in Gestural Control of Music*.
- Dahl, S. Bresin, R. (2001). "Is the Player More Influenced by the Auditory Than the Tactile Feedback from the Instrument." *In proceedings of the Conference on Digital Audio Effects*.
- Dahl, S. (2000). "The Playing of an Accent – Preliminary Observations from Temporal and Kinematic Analyses of Percussionists." *Journal of New Music Research No. 3, pp. 225-233*.
- Enkelaar, K. A website with information about the Theremin used for the test: (Visited 1.3.2004)  
<http://www.our.net.au/~cytronic/theremin/INDEX.HTM>
- Finney, S.A. (1997). "Auditory Feedback and Musical Keyboard Performance." *Music Perception, vol. 15, no. 2, pp 153-174*.
- Freed, A. Chaudhary, A. Davila, B. (1997) "Operating Systems Latency Measurement and Analysis for Sound Synthesis and Processing Applications" *Proceedings of the International Computer Music Conference*.
- Goldstein, B. E. (1999). *Sensation & Perception. Fifth Edition* Brooks/Cole Publishing Company.
- Hunt, A. Wanderley, M. Paradis, M. (2002) "The Importance of Parameter Mapping in Electronic Instrument Design." *In Proceedings of the Conference on New Instruments for Musical Expression (NIME-02)*.
- Hunt, A. Wanderley, M. Kirk, R.(2000) "Towards a Model for Instrumental Mapping in Expert Musical Interaction." *In Proceedings of the International Computer Music Conference*.
- ITU-R. (1997) Recommendation BS. 1116-1, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems". *International Telecommunication Union Radiocommunication Assembly*.
- MacKenzie, I. Ware, C. (1993) "Lag as a Determinant of Human Performance in Interactive Systems." *Proceedings of Conference on Human Factors in Computing Systems, p.488-493*.
- Mulder, A. (1998) "Design of Virtual Three-Dimensional Instruments for Sound Control." *PhD Thesis*, Simon Fraser University.
- O'Modhrain, M. S. (2000) "Playing by Feel: Incorporating Haptic Feedback Into Computer-Based Musical Instruments" *PhD Thesis, University of Stanford*.
- Paradiso, J. (1997) "Electronic Music Interfaces: New Ways to Play," *IEEE Spectrum, 34(12), 18-30*. Later expanded as an online article (1998) (visited 1.4.2003):  
<http://web.media.mit.edu/~joep/SpectrumWeb/SpectrumX.html>
- Rovan, J. Hayward V. (2000) "Typology of Tactile Sounds and Their Synthesis in Gesture-Driven Computer Music Performance." *Trends in Gestural Control of Music*. Ircam - Centre Pompidou - 2000.
- Sawchuk, A. A. Chew, E. Zimmermann, R. Papadopoulos, C. Kyriakakis, C. (2003) "From Remote Media Immersion to Distributed Immersive Performance." *Proceedings of the 2003 ACM SIGMM workshop on Experiential Telepresence*.
- Vertegaal, R. Eaglestone, B. (1996). "Comparison of Input Devices in an ISEE Direct Timbre Manipulation Task." *Interacting with Computers 8, 1, pp.113-30*.
- 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*.
- Wanderley, M. (2001). "Performer-Instrument Interaction: applications to Gestural Control of Music." *PhD Thesis. Paris, France: University Pierre et Marie Curie – Paris VI*.
- Wanderley, M. Battier, M. (2000) Eds. *Trends in Gestural Control of Music*. Ircam - Centre Pompidou - 2000.
- Ware, C. Balakrishnan, R. (1994) "Reaching for Objects in VR Displays: Lag and Frame Rate." *ACM Transactions on Computer-Human Interaction, vol. 1, no. 4, 331-356, 1994*.
- Watson, B. Walker, N. Ribarsky, B. Spaulding, V. (1998) "The Effects of Variation of System Responsiveness on User Performance in Virtual Environments." *Human Factors 40(3):403-414*.
- Watson, B. Walker, N. Ribarsky, B. Spaulding, V. (1999) "Managing Temporal Delay in Virtual Environments: Relating System Responsiveness to Feedback" *In proceedings of the ACM CHI 1999 Conference on Human Factors in Computing Systems*.
- Wright, J. Brandt, E. (2001) "System-Level MIDI Performance Testing" *Proceedings of the International Computer Music Conference*.