Evaluation of 3D Haptic Target Rendering to Support Timing in Music Tasks

Ricardo Pedrosa **Karon E. MacLean**
Department of Computer Science
The University of British Columbia
rpedrosa,maclean@cs.ubc.ca

Abstract

Haptic feedback is an important element that needs to be carefully designed in computer music interfaces. This paper presents an evaluation of several force renderings for target acquisition in space when used to support a music related task. The study presented here addresses only one musical aspect: the need to repeat elements accurately in time and in content. Several force scenarios will be rendered over a simple 3D target acquisition task and users' performance will be quantitatively and qualitatively evaluated. The results show how the users' subjective preference for a particular kind of force support does not always correlate to a quantitative measurement of performance enhancement. We describe a way in which a control mapping for a musical interface could be achieved without contradicting the users' preferences as obtained from the study.

Keywords: music interfaces, force feedback, tempo, comfort, target acquisition.

1. Introduction

The potential utility of including haptic (force and/or tactile) feedback in computer-music interfaces seems inarguable. Forces and vibrations are clearly an important element of playing a traditional acoustic instrument, and players of electronic music controllers often note inadequacies in this realm. Given the tight-coupled motor-auditory control loops involved in many aspects of music composition and performance, this is unsurprising; and numerous projects have incorporated haptic feedback into musical controllers in a variety of ways (e.g. [1][2][3][4][5]).

However, past efforts to design haptic feedback for computer music interfaces have focused on factors such as the interface's technical capabilities and the designer's personal experience and intuition. We are embarking on a project that attempts a more perceptually guided, user-centered evaluation for the inclusion of haptic feedback in a gesture-based computer music interface [6].

NIME09, June 3-6, 2009, Pittsburgh, PA

Traditional music instruments provide enough references to precisely generate a particular sound. These references go from visual location of targets (e.g. keys in a keyboard) to the amount of force needed to produce a sound (e.g. plucking a string). When using a gesture controller for musical tasks, the lack of haptic feedback constitutes one of the major problems. In these new controllers and especially in those called “open air” or non-contact controllers (e.g. [1]) the performer relies on proprioception and egolocation, receiving no haptic feedback from the media from where the sounds are generated [5].

We are interested in determining how force feedback could be of use when designing a gesture controller for computer music applications that will map the hand position in space to a sound generation process. The performer would need to know where in the space a particular sound could be found, i.e. generated from that location. Several methods have been proposed to locate targets in space by rendering forces through a 3D haptic device [7][8][9][10]. However, none of those methods has been tested in a musical task. In this paper we focus on one particular aspect of music: the need for accurately performing a sequence of notes in time. In the experiment described here, we propose a set of force-based interaction models and we objectively and subjectively measure their ability to support the selection of targets in space following a rhythmic time cue. A detailed description of the experiment is given in Sections 3 and 4.

Our approach for designing force feedback into computer music interfaces is to gain insight on the issues related to different aspects of music performance. In the results presented here only the temporal issues of music performance are analyzed. These results will be taken into consideration when designing the haptic feedback to support creation/control of other music elements, like dynamics or contour.

2. Related Work

2.1 Role of the Haptic Channels in Music Performance

Gillespie [6] provided a broad definition of a musical instrument as: “a device which transforms mechanical energy (especially that gathered from a human operator) into acoustical energy”. Thus, the musician-instrument relationship could be represented as a feedback control system [11][12] where the control loop is closed over two paths via the auditory and the haptic channels. The major feedback from the instrument to the musician occurs through the air in the form of sound waves. The mechanical contact between the musician and the instrument (fingertips, hands or mouth just to mention a few possibilities) acts as an additional bidirectional flow channel through which the musician sends information in the form of mechanical energy, and receives from the instrument the haptic information associated with the sound generated and the status of the sound control mechanism.

Through years of practice musicians develop what has come to be known as an "internal representation of music": an asset enforced with training and the development of motor programs [13]. The motor programs are believed to reside in the cerebellum and are part of a higher level of control, those in charge of triggering major events. At a lower level of control, spinal reflexes informed by sensory signals regulate the force and speed of finger or arm movements to perform an intended melody, whereas the motor program is triggered at a higher level, dictating the melody.

These and other past works strongly suggest that the haptic channels play an important role in both fine-tuning the performance and defining the instrument status. The development of low-level sensorimotor reflexes through practice and training depends heavily on the musician's apprehension of the relationship between force exerted and sound produced. Vibrations coming from the instrument's sound generation engine close the loop in the same mechanical channel to keep the control loop at the same low level that started the action, while reinforcing the feeling of using a musical instrument, or a "resonating and responding object" in the words of Askenfelt and Jansson [1].

2.2 Use of Forces to Enhance Computer Music Interfaces

O'Modhrain [12] established that the presence of force feedback in music interfaces informs players of the consequence of their actions. Her experimental work shows that as long as the force feedback is consistent with the instrument's auditory response, it could be learned, whether or not this behavior is correlated with the auditory feedback (e.g. either increasing the force with an increasing pitch or decreasing the force with an increase of pitch). This opens many possibilities. However, her experiments were limited to a 2D environment and forces were used as global references only (e.g. general instrument status).

The interface could potentially be improved if some local reference were provided in order to enhance the perception of certain musical attributes. Previous research on the use of haptic feedback for target acquisition and localization in virtual environments have shown the benefits of collocated visual and haptic cues [9], the use of "virtual magnets" or force fields attached to each target [8][10] or having depth perception enhanced with audio feedback combined with the haptic perception of the targets [7]. However, these findings have been obtained using generic target selection tasks, and it is not clear how closely they will apply to music-related tasks. Musically-focused tasks have certain distinguishing attributes that might translate to special needs; for example, the need for both expressiveness and precision.

We will then try to add some simple elements to put the task of three-dimensional (3D) target acquisition into a music context. We will then evaluate the inclusion of references for local targets (notes or musical events location in space) on top of global references to support temporal and spatial control.

3. Approach

To further our understanding of the potential ways in which forces can augment musical performance, we carried out a study where performance measurements were informed by observation and followed by interviews. Our principal **research questions** are:

a) In a musically focused target acquisition task, will rendering haptic cues of possible target locations support, or conversely disrupt acquisition following a time cue?

b) How does subjective preference for a particular kind of force support relate to performance enhancement?

3.1 Musically Focused Experiment Task

Participants were required to repeatedly acquire 4 spatial targets in a particular sequence. To musically focus this task, the targets had to be acquired in synchrony with a temporal cue from a metronome. Each target could represent a note or a musical phrase that needs to be triggered at a specific time. Participants should be able to a) precisely acquire each target and b) maintain a fluid and rhythmic movement while going from one target to another.

3.2 Force Supports

In choosing the types of supporting force behaviors to test, we chose a metaphor based on an early electronic music instrument called "Ondes Martenot" [15], controlled either by depressing keys on a six-octave keyboard or by sliding a metal ring worn on the right-hand index finger in front of the keyboard. The ring was attached to a string that could be pulled in an attempt to provide the same degree of expressivity associated with a bowed instrument like a cello or violin.

We used this last interaction method as global reference. We haptically and graphically rendered an elastic 3D string attached to two anchors, located at the left and right extremes of the workspace. Attached to the center of the string we placed a cursor that could be moved around the workspace. To explore the goal of providing additional directional context to the user, we tested an additional variation: one of the anchors (the leftmost) moved up and down following the height of the cursor. These were tested as the fixed anchor and mobile anchor variations in the force condition levels.

A series of targets were graphically rendered in space. A local haptic reference was rendered as gravity wells limited to a spatial region near each target.

4. Methods

Participants were presented visually with a screen with four blue small spheres (the targets) and a cursor linked to two other white spheres (the anchors) through a graphically rendered red line as depicted in Figure 1. Using this background, 8 levels of the Force condition were presented, while users where asked to acquire each target in a fixed sequence following two auditory temporal cues from a metronome (60 and 80 beats per minute, or bpm).

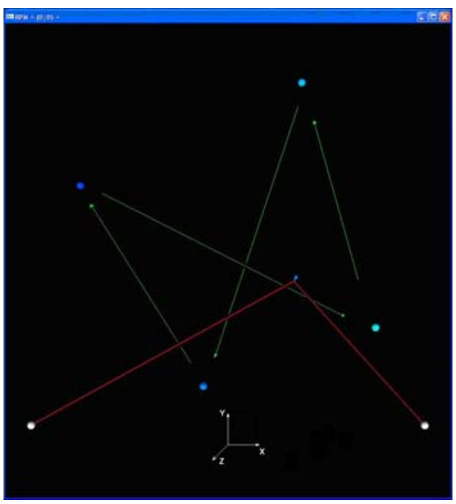

Figure 1. Visual representation used for all trials. The green arrows, which indicate the target sequence, were not part of the graphical rendering. The red line (showing the elastic string connected to its anchors) was present, in all conditions.

4.1 Apparatus

The experimental apparatus consisted of an Intel Core2-Duo based computer running Windows XP and controlling a SensAble PHANTOM Premium. The PHANTOM was placed on a table in front of the user. Visual information was presented on a 20” LCD monitor placed 60cm away from the user and raised 24cm above the table where the PHANTOM was placed. The software used for the experiment was developed in C++ using the OpenHaptic Toolkit from SensAble.

4.2 Force Conditions and Tasks

The four targets to be acquired by the users were rendered throughout a single plane perpendicular to ground and facing the participant. The coordinates for each target were fixed at: (-60,138,40), (-10,56,40), (30,180,40) and (60,80,40) where the axis at X=0 was placed at the center of the rendered area. The PHANTOM was placed in such a way that the vertical target plane coincided with the edge of the table in front of the user in order to provide a stronger spatial reference. Also, the pointer on the graphical interface changed its size according to how distant it was from the plane where the targets were placed (the size diminished as the pointer was moved towards the PHANTOM away from the user and was enlarged as it was pulled towards the user).

The anchors were placed in the same vertical plane as the targets. There were two variations for the anchors: Fixed Anchors (**FIX**) where the anchors were kept fixed in space at (80,40,40) for the right anchor and (-80,40,40) for the left anchor; and Mobile Anchor (**MOB**) where the left anchor tried to follow the pointer movement in the Y axis but always remained at a lower vertical level than the pointer and at the same (X,Z) coordinate. For each FIX or MOB, four force supports were presented to give 8 levels of the Forces condition (see below). These levels were:

$\mathbf{ctrl_{FIX}}$ (or $\mathbf{ctrl_{MOB}}$): No forces level. The visual representation was the only cue to acquire the targets.

$\mathbf{M_{FIX}}$ (or $\mathbf{M_{MOB}}$): Magnet level. A magnetic field attracting the pointer towards each target was haptically rendered on top of the visual representation. This field was only active in a region close to each target.

$\mathbf{S_{FIX}}$ (or $\mathbf{S_{MOB}}$): String level. On top of the visual representation, an elastic string was haptically rendered to attach the pointer to each anchor point.

$\mathbf{M+S_{FIX}}$ (or $\mathbf{M+S_{MOB}}$): Magnet plus String level. Both the magnetic fields around each target and the elastic string attaching the pointer to the anchors were haptically rendered on top of the visual representation.

There were two major task involved in the experiment:

Acquisition Task: Users were presented with these 8 force conditions and asked to use the PHANTOM stylus to acquire each target following a fixed sequence (starting at the right-most target and ending at the left-most), by clicking on the target with the stylus’ button. Users were not allowed to rest their arm on the table while acquiring the targets. Users were asked to click in synchrony with an auditory temporal cue from a metronome. Two metronome speeds were used: 60bpm (one tick per second) and 80bpm (one tick every 0.75s). Users were asked to repeat the target acquisition from right to left as long as the auditory cue was present (14-18 repetitions).

The audio tracks with the metronome ticks consisted of 4 bars of four ticks each in both the slow (60bpm) and fast (80bpm) tempos. The duration in time of these tracks were 20 and 15 seconds respectively.

Cognitive Load Task: After a training session for each condition, users' performance was measured when there was a voice recorded over the metronome reading a sequence of letters and numbers in synchronism with the ticks. This was intended as an additional cognitive load to the task of target acquisition. A total of 16 letters/numbers were read in random order for every trial. The letters read were B, C, D and E and the number 3. They were selected for their similar phonetics and because the letters also serve as the names of musical notes. Users were asked to count the occurrences of either the second or third item read. For these voice plus metronome audio tracks two time lengths were used. The slow tempo track consisted of one bar (four ticks) without voice-over and then 4 bars where a letter or number was read over each tick. The duration in time of each of these slow tracks was 20 seconds. The fast tempo consisted of two bars (eight ticks) without voice-over followed by 8 bars where a letter or number was read over every other tick. The duration in time of each of these fast tracks was 30 seconds.

4.3 Procedures

Each participant received a description of the whole session. They were specifically instructed to make sure to click (acquire each target) following the tempo dictated by the temporal cue, but to try to get as close as possible to the targets as they could.

For each scenario, the user was presented first with the simple metronome tick as the temporal audio cue (training session) and then with the cue with the voice recorded over the metronome tick. The sequence of the scenarios, the audio tracks with the voice-over and the sequence in which for each scenario the low and high tempo were presented, were all randomized between users. Each user completed 16 trials (8 scenarios x 2 tempos) and their corresponding 16 training sessions.

For example, for a given user we randomly selected one force condition (e.g. $\mathbf{M+S_{MOB}}$), one sequence of tempo (e.g. low tempo first) and for each tempo, one audio track with the voice-over (e.g. track 5 for low tempo and track 2 for high tempo). The user was then presented with the following sequence of trials:

a) $\mathbf{M+S_{MOB}}$, low tempo, metronome only audio track.
b) $\mathbf{M+S_{MOB}}$, low tempo, track 5.
c) $\mathbf{M+S_{MOB}}$, high tempo, metronome only audio track.
d) $\mathbf{M+S_{MOB}}$, high tempo, track 2.

Then another force condition, tempo sequence and audio track were selected and presented in the same manner.

4.4 Metrics

Throughout the study we measured the 3D spatial coordinates of the users' clicks, the acquisition timestamp for each click and the cognitive-load task letter count as reported by the user.

At the end of the study, participants were asked to rank the best and worst force conditions for each the fixed and mobile anchor categories, and to select from those the best and worst force conditions overall (across both fixed and mobile anchors). At this time, users were allowed to briefly try again each force condition as needed. They were also interviewed for less structured subjective responses.

4.5 Design

Our analysis took the form of an 8x4x2 factorial experiment (8 haptic scenarios, 4 targets, 2 tempos). Dependent variables were coordinate accuracy (X, Y and Z separately) and a single variable T for tempo accuracy. All of these were computed with repeated measures (average of all clicks in a trial).

Each trial was scored individually for tempo and target acquisition accuracy. Letter count accuracy was also scored.

The data measured directly from the user was filtered for involuntary double clicks while acquiring the targets (defined as consecutive clicks that were less than 200ms one from another). For each trial, a graph of the consecutive timestamp for each click was created and the slope of the best linear fit crossing the origin was taken as a measure of the overall tempo. In order to characterize more subtle variations in rhythm, the inter-onset intervals (IOI) between clicks were also analyzed.

For the target acquisition accuracy we took two measurements: the target's coordinate for each axis and the target acquisition sequence. Instead of defining a valid target size as a limited spatial region around each target, we retained more information by considering each user's click as the acquisition action for the target closer to the click's spatial location; and computed error as distance from that target. Spatial error was measured independently for X, Y and Z coordinates. The correctness of acquisition sequence (right to left) was measured by counting the amount of targets per trial that were out of sequence.

For each trial (one setting of each condition), the errors on acquiring each target on each coordinate axis and the error on following the tempo were computed and these results were used in statistical analysis.

5. Results

For this study we employed four independent variables: force condition (8 levels), tempo (2) and target location (4). We suspected that target location would be a factor affecting the acquisition performance because of the PHANTOM physical structure.

5.1 Participants

Nine participants took part in the study. One participant's data was excluded from analysis due to inability to maintain the required spatio-temporal coordination. From the eight valid users (four male and four female, 1 left handed) only 3 had previous exposure to haptic devices. Participants were compensated with $10 for their performance.

5.2 Data and Statistical Analysis

Graphical figures are referred to in the Discussion. We performed an 8x4x2 factor ANOVA on X, Y, Z and T. The statistical analysis performed on the data arranged in this manner showed significant differences for the target's coordinates in the X and Z axis (F(3,448)=7.43, p<.0001 and F(3,448)=6.78, p<.001, respectively) and on the force conditions for the Z axis (F(7,448)=8.64, p<.0001).

No significant differences were found among force condition levels with respect to letter count recall, the sequential accuracy or the temporal accuracy (both overall and IOI).

Users' subjective statements of the best and worst Force Condition level for each of the FIX, MOB categories and the best and worst overall are shown in Figure 2.

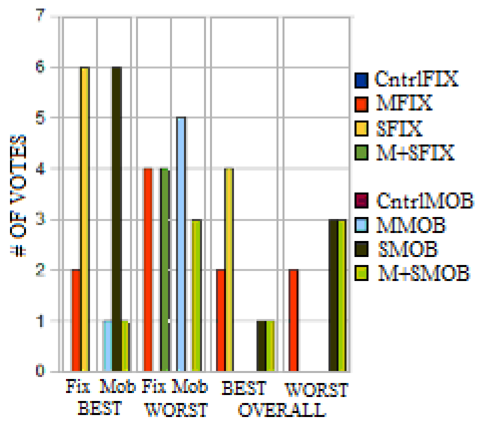

Figure 2. Users' qualitative ranking of the haptic scenarios.

6. Discussion and Conclusions

Figure 3 shows the root mean squared errors among target coordinates for the X and Z axis. This confirms our assumption that the spatial location of the target is a factor that needs to be taken into consideration in these tasks. We can also see that the errors on the Z coordinate are higher than in the X coordinate. This should come as no surprise since Z positioning was more difficult. The XY positioning benefited from good graphical cues, whereas the depth perception relied on secondary cues like the size of the pointer.

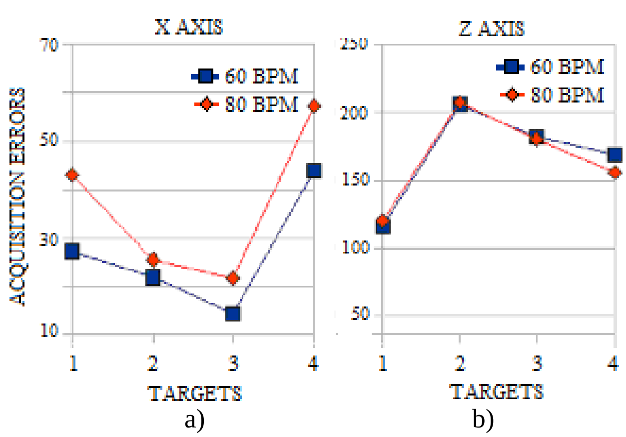

Figure 3. Errors among targets coordinates for a) the X axis and b) the Z axis.

The errors for the X coordinate are another artifact of the interface we are using. The PHANTOM structure is more suitable for deploying curved trajectories when no “hard limit” (like a solid plane rendering) is provided. As the targets were placed symmetrically in the X axis with respect to the center of the PHANTOM, the effect in this coordinate is higher than in the Y coordinate, making the acquisition of the targets close to the center less prone to errors.

Figure 5 shows the root mean squared errors for the Z coordinate by force condition for each tempo. As expected, the Z errors are smaller for those force condition levels where the targets are rendered without a reference support (MFix, M+SFix, MMob, M+SMob), at both 60 and 80BPM. If we combine this result with the lack of significance among force conditions in X and Y accuracy, we might conclude that the best renderings were those that represent the targets in space.

However, when looking at the subjective rankings, we see that users didn’t like those force conditions where the targets were rendered (magnet support). This result links to the comfort level while using the interface for the task at hand. The preferred force condition for both the fixed anchors and the mobile anchors groups were those where an elastic string was rendered (6 out of 8 participants). From those, the force condition favored overall was the one where the anchors were fixed. The least-favored force condition levels included all conditions where a magnet force was rendered; and these also reflected the highest rate of “double-clicks”. These double-clicks were caused by not holding the PHANTOM stylus tightly enough, resulting in movement jerk upon reaching the target, and involuntary button presses. User's verbal analysis revealed that the magnetic field seemed useful to reach a target but at the same time it made difficult to move away from it when going for another target. The resulting movement was not fluid at all and the users found hard to reach the following target while following the rhythm of the metronome under the conditions with magnet support.

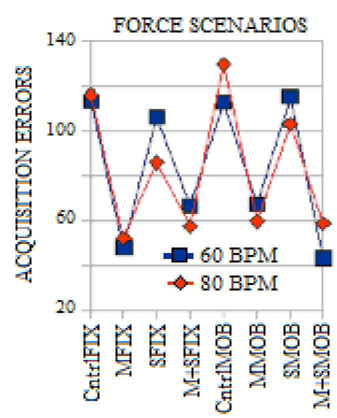

Figure 5. Performance under haptic scenarios for both tempos in the Z axis.

Several scenarios were not mentioned in the subjective preferences. This is the case for the control conditions where no forces were displayed. This seems to indicate that no force might be perceived as in between a bad force and a good force.

6.1 Conclusions

Summarizing the results, we can conclude that the presence of a fixed reference is preferred over fluctuating environments (mobile anchors, magnetic fields distributed in space or a combination of both) while acquiring several targets sequentially in time. More important is the fact that the results of a subjective perceptual evaluation could counter objective results by disregarding those force supports with a similar or better quantitative performance in favor of those that simply “feel” better.

The lack of accuracy on the Z plane could be improved by choosing some other visual representation to improve depth cues. A rigid force rendering to limit the movements on that plane is something we wouldn’t recommend, since this may be an important performance utility. Instead we could choose to map in this coordinate some parameters of relatively small significance in the interface at hand, and leave the XY plane for the most important parameters.

For instance, for the given scenarios, and taking into consideration that the quantitative performances are similar enough between the best cases according to the subjective evaluation, we will pursue to use the elastic string rendering as the final approach. Due to the lack of accuracy on the Z, main parameters should be mapped to the X and Y axis. As such, we note that velocity (intensity of the sound) and pitch could be assigned to either X or Y coordinates while the Z axis could be used to inflect several effects like pitch bending or reverb - but only if the mapping is set in a way that small displacements in Z are not taken into account.

We could generalize these ideas beyond music interfaces to any design of expressive computer interfaces. The best approach in terms of quantitative results may not always be the one that feels better in the hands of the performer. Some consideration in the design must be accorded to how natural the interface feels and how comfortable the performer is going to be. For expressive interfaces, the designer should find a way to maximize comfort and provide the means to achieve a sufficient level of control.

It will be interesting to follow this study with a larger set of subjects. This study has approached only one musical aspect: the need to repeat elements accurately in time and in “content” (read as notes played or phrases executed following a predetermined time and pattern). However, this study did serve as a first approach to remove from future examinations those force renderings that are of no relevance for the task, either because of outright disruptiveness or because they provide neither assistance or harm.

References

[1] Bongers, B. “The use of active tactile and force feedback in timbre controlling electronic instruments.” In Proceedings of the 1994 International Computer Music Conference, pp. 171-174, 1994.

[2] Chu, L. “Haptic Feedback in Computer Music Performance”. Proceedings of ICMC, 1996, pp. 57-58.

[3] Gillespie, B. “The Virtual Piano Action: Design and Implementation” Proceedings of the International Computer Music Conference, Aahus, Denmark, Sept 12--17, 1994. pp. 167--170.

[4] Nichols, C. “The vBow: Development of a Virtual Violin Bow Haptic Human-Computer Interface”. Proceedings of the 2002 Conference on New Interfaces for Musical Expression (NIME-02), pages 29-32, Dublin, Ireland, May 24-26 2002.

[5] Rovan, J. Hayward, V. “Typology of Tactile Sounds and their Synthesis in Gesture-Driven Computer Music Performance”. In Trends in Gestural Control of Music. Wanderley, M., Battier, M. (eds). Editions IRCAM, Paris, 2000.

[6] Pedrosa, R., MacLean, K. “Perceptually Informed Roles for Haptic Feedback in Expressive Music Controllers.” In A. Pirhonen and S. Brewster (Eds.): HAID 2008, LNCS 5270, pp. 21–29, 2008.

[7] Cockburn, A., Brewster, S.: Multimodal feedback for the acquisition of small targets. Ergonomics 48 (2005) 1129–1150.

[8] Magnusson, C., Rassmus-Gröhn, K. “Audio haptic tools for navigation in non visual environments.” In: ENACTIVE 2005, the 2nd International Conference on Enactive Interfaces. (2005)

[9] Swapp, D., Pawar, V., Loscos, C. “Interaction with co-located haptic feedback in virtual reality”, Virtual Reality, Springer, 1-7, May 2006, pp. 24-30.

[10] Wall, S. A., Paynter, K., Shillito, A. M., Wright, M., Scali, S. “The effect of haptic feedback and stereo graphics in a 3D target acquisition,” in Proc. EuroHaptics 2002, pp. 23-29, July 2002.

[11] Gillespie, B. “Haptic display of systems with changing kinematic constraints: The Virtual Piano Action.” PhD Dissertation. Stanford University, 1995. http://www-personal.umich.edu/~brentg/Publications/Thesis/thesis.html

[12] O’Modhrain, S. “Playing by Feel: Incorporating Haptic Feedback into Computer-Based Musical Instruments”. PhD Dissertation. Stanford University, 2000. http://www-ccrma.stanford.edu/~sile/thesis.html.

[13] Lederman, S.J., Klatzky, R.L. “Haptic aspects of motor control”. In F. Boller and J. Grafman, eds. Handbook of neuropsychology, Vol.11, 131-148. New York. Elsevier. 1997.

[14] Askenfelt, A., Jansson, E.V. “On vibration sensation and finger touch in stringed instrument playing.” Music Perception. Spring 1992, Vol. 9, No. 3. pp. 311-350.

[15] http://en.wikipedia.org/wiki/Ondes_ Martenot