

The *AHI*: An Audio And Haptic Interface For Contact Interactions

Derek DiFilippo and Dinesh K. Pai

Department of Computer Science
University of British Columbia
2366 Main Mall.

Vancouver, BC V6T 1Z4, Canada

Tel: +1-604-822-6625

{*difilip | pai*}@cs.ubc.ca

ABSTRACT

We have implemented a computer interface that renders synchronized auditory and haptic stimuli with very low (0.5ms) latency. The audio and haptic interface (AHI) includes a Pantograph haptic device that reads position input from a user and renders force output based on this input. We synthesize audio by convolving the force profile generated by user interaction with the impulse response of the virtual surface. Auditory and haptic modes are tightly coupled because we produce both stimuli from the same force profile. We have conducted a user study with the AHI to verify that the 0.5ms system latency lies below the perceptual threshold for detecting separation between auditory and haptic contact events. We discuss future applications of the AHI for further perceptual studies and for synthesizing continuous contact interactions in virtual environments.

KEYWORDS: User Interface, Haptics, Audio, Multimodal, Latency, Synchronization

INTRODUCTION

When two objects collide, we have a contact interaction. Everyday lives are full of them: pulling a coffee cup across a table, tapping our fingers on a computer keyboard, etc. These contact interactions can produce characteristic sounds and forces that communicate information about our relationship with rigid objects and the surrounding environment. By using our ears and hands we can tell if the coffee cup was placed safely on the table or if the table is made of glass or wood, for example. Depriving someone of this sensory feedback from his or her interactions could

limit their ability to navigate and control their environment.

An effective computer interface for interacting with a simulated environment would allow one to tap and scrape on virtual objects in the same way one can tap and scrape on real objects. In addition to providing visual feedback, the interface would also create realistic auditory and haptic cues. These cues would be synchronized so that they appear perceptually simultaneous. They would also be perceptually similar -- a rough surface would both sound and feel rough. This type of interface could improve the amount of control a user could exert on their virtual environment and also increase the overall aesthetic experience of using the interface. For example, a system designer might wish to use this interface to represent remote objects to discerning user groups varying from NASA scientists to e-commerce customers. In this paper, we present an experimental audio and haptic interface (AHI) for displaying sounds and forces with low latency and sufficient realism for interactive applications.



Figure 1: The AHI in a typical configuration.

Figure 1 shows the AHI in a typical configuration. The user grips the handle with their right hand and moves it in

the plane as they would move a computer mouse. The left side of the monitor screen shows a Java graphical user interface for loading sound models and controlling interaction parameters and the right side shows a graphical window for viewing haptic forces and audio signals in real-time.

The novelty of the AHI lies in the tight synchronization of the auditory mode and the haptic mode. User interaction with the simulated environment generates contact forces. These forces are rendered to the hand by a haptic force-feedback device, and to the ear as contact sounds. This is more than synchronizing two separate events. Rather than triggering a pre-recorded audio sample or tone, the audio and the haptics change together when the user applies different forces to the object.

Building a device to a certain technical specification provides no guarantee for how a user will perceive the effect of the device. However, an over-specified device can be used to help establish lower bounds for human perception that give system builders a reliable design target. An analogy would be to the design of computer monitor hardware to support refresh rates of 60Hz. This refresh rate is well known to be sufficient for comfortable viewing and sufficient to simulate continuous motion. Our goal is to use the AHI to help establish similar perceptual tolerances for synchronized audio and haptic contact interactions.

The first part of this paper describes the AHI hardware for user input, models for calculating haptic and audio contact forces, and the control structure for rendering these forces. The second part of this paper presents experimental results that suggest the 0.5ms system latency of the AHI lies below the perceptual tolerance for detecting synchronization between auditory and haptic contact events. Establishing that our interface works below perceptual tolerance enables us to use it for psychophysical experiments for multimodal perception.

HARDWARE

User input to the AHI comes from a 3 degree of freedom (DOF) Pantograph device (Figure 2). The 5-bar mechanism is based on a design by Hayward [14] but extended to 3 DOF to our specification. It reads 3 DOF of position as user input, and renders 3 DOF of forces as output. The user can move the handle in the plane as well as rotate the handle. There are two large Maxon motors attached to the base of the Pantograph which apply forces on the handle via the 5-bar linkage. A small motor in the handle can exert a torque on the handle as well. The device, therefore, is complete for rigid motions in the plane, i.e., it can render the forces and torque due to any contact with a rigid body attached to the handle in a planar virtual world ("flatland"). We do not currently use the third rotational DOF for our work with the AHI.

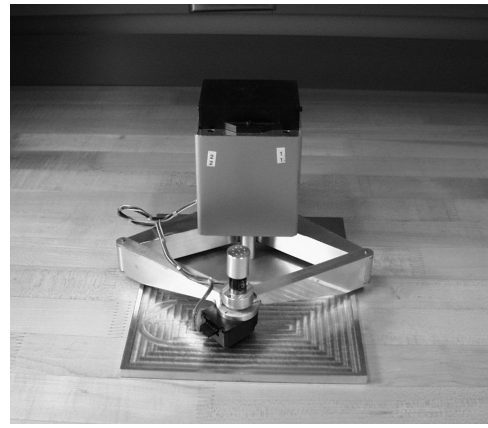


Figure 2: The Pantograph haptic device.

A dedicated motion control board (MC8, Precision Microdynamics), hosted in a PC running Windows NT, controls the AHI. The board has 14 bit analog to digital converters (ADCs) for reading the potentiometers attached to the base motors as well as quadrature decoders for reading the optical encoder which measures the handle rotation.

A SHARC DSP (Analog Devices 21061 running at 40MHz) on the MC8 synthesizes sounds and forces. Output voltages for controlling the Pantograph motors and for rendering audio are sent out through 14 bit digital to analog converters (DACs). The audio waveforms can be input directly to a sound system; in our set up they are sent to the soundcard of the computer for ease of capture and playback.

By using this specialized hardware we bypass the complications that arise from balancing the needs of real-time, deterministic response and ease of access from user-level software on a widely available operating system such as NT. The AHI control code is compiled for the DSP and has exclusive control over its resources. This allows us to precisely time our control algorithms as well as accurately diagnose inefficiencies and bugs. In particular, the overall system latency is 0.5ms for synchronized changes in audio and haptics.

AUDIO SYNTHESIS

We wish to simulate the audio response of everyday objects made out of wood, metal, ceramic, etc. Contact with these objects can be characterized by impulsive excitation of relatively few exponentially decaying, weakly coupled sinusoidal modes. Modal synthesis and impulse generation techniques have been developed for these types of percussive sounds [2]. We use the modal audio synthesis algorithm described in [20]. This algorithm is based on vibration dynamics and can simulate effects of shape, location of contact, material, and contact force. Model parameters are determined by solving a partial differential equation, or by fitting the model to empirical data [15].

The sound model assumes that the surface deviation y obeys a wave equation. We add a material-dependent decay coefficient to the wave equation to damp the sounds. The exponential damping factor $d = f\pi\tan(\phi)$ depends on the frequency f and internal friction ϕ of the material, and causes higher frequencies to decay more rapidly. The internal friction parameter is material dependent and approximately invariant over object shape. Equation 1 represents the impulse response of a general object at a particular point as a sum of damped sinusoids.

$$y(t) = \sum_{n=1}^{\infty} a_n e^{-d_n t} \sin(\omega_n t) \quad (1)$$

The sound model of an object consists of a list of amplitudes a_n and complex frequencies $\Omega_n = \omega_n + id_n$. Equation 2 shows how one complex frequency is computed. At time 0, the signal is the product of the frequency-amplitude a , and the contact force $F(0)$. At each successive time step (determined by the sampling frequency F_s), the signal is the sum of a decayed version of the previous signal plus a new product of amplitude and contact force. The model responds linearly to input force $F(k)$. Once we have the model parameters, all we need to begin synthesizing sounds is a series of contact forces to plug into the right-hand side of the recursion. The output signal at time k is $Re(\sum y_n(k))$, with the sum taken over all computed frequencies.

$$\begin{aligned} y_n(0) &= a_n F(0) \\ y_n(k) &= e^{i\frac{\Omega_n}{F_s} k} y_n(k-1) + a_n F(k) \end{aligned} \quad (2)$$

This synthesis algorithm has two benefits. First, it is linear. The computed audio is the discrete convolution of the force history with the impulse response. There is a natural relationship between the input forces and the output signal; this relationship would not be as straightforward if we used sample-based synthesis. The linearity also makes it efficient. With a basis change, an audio signal can be computed with 2 multiplications and 3 additions per complex frequency, per sample. Second, the audio quality can degrade gracefully. If DSP time is running short we can compute fewer frequencies, resulting in a smooth loss of detail rather than sudden audio dropouts.

HAPTIC FORCE SYNTHESIS

As the user moves the Pantograph handle we need to compute the contact forces resulting from these interactions, and then render them as forces on the handle of the Pantograph by exerting torques on its base joints. These computations take place in two coordinate frames. One is the world frame of xy -coordinates and the other is the Pantograph frame of joint angles. The simulated environment uses the world frame, but the control code

only knows about joint angles. We need a forward kinematic mapping that gives the xy -position of the handle as a function of base joint angles, as well as a differential kinematic mapping that gives the base joint torques as a function of applied force to the handle.

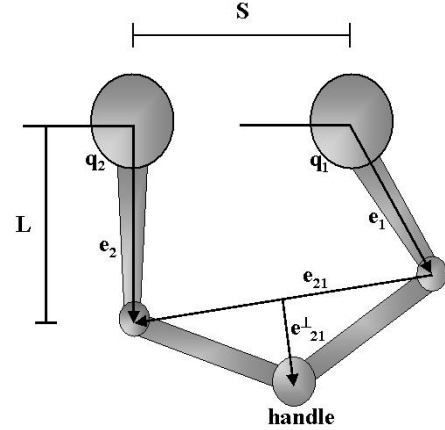


Figure 3: Pantograph kinematics.

For the forward kinematic mapping, we specify the base joint of motor 1 as the origin of the world frame. There is a geometric constraint that allows us to compute the position of the handle: the vector pointing from elbow 1 to elbow 2 ($e_{21} = e_2 - e_1$) in the world frame is always perpendicular to vector pointing to the handle from the midpoint of e_{21} . If q_1 and q_2 are the base joint angles, then the elbows become $e_1 = (L\cos(q_1), L\sin(q_1))$ and $e_2 = (L\cos(q_2), L\sin(q_2))$ where L is the length of the proximal arms, and S is the separation between the two base joints. Setting e_{21}^\perp as the vector pointing from the midpoint of e_{21} to the handle h , we have $h(q) = e_1 + 0.5e_{21} + e_{21}^\perp$. This expression for h in terms of joint angles q has a simple geometric interpretation, as shown in Figure 3.

Once we have the handle coordinates, and compute a contact force F , we need to transform this force into base joint torques τ for rendering. The Jacobian $J = \partial h(q) / \partial q$ of the forward kinematic mapping relates forces to torques by $J^T F = \tau$. The details of constructing the Jacobian for the Pantograph are quite general and are covered in basic robotics texts [11]. In our particular implementation, we can avoid the expense of computing the partials of $h(q)$ by exploiting the structure of the Jacobian. Details are removed here for the sake of brevity.

NORMAL FORCES

For interactions normal to the surface of a plane, a spring/damper/impulse combination constrains the user to the surface by applying a penalty force. If the normal displacement past the surface is x_n , and the current normal velocity is v_n , then the haptic constraint force is $F = Kx_n + Dv_n$ where K and D are spring and damping constants. For 10ms after a new contact, we add a unilateral impulse Pv_n

to the spring/damper combination. This technique is known to increase the perception of haptic stiffness without introducing closed-loop instabilities that can occur with large spring coefficients [16].

TANGENTIAL FORCES

For interactions tangential to the surface of a plane, we have implemented a stick-slip friction model to provide force feedback. Stick-slip models exhibit two states: *slipping* and *sticking*. Specifying a stick-slip model requires defining state transition conditions. In general, a virtual proxy point connects the real contact point to the surface. In the sticking state, the real contact point separates from the proxy and frictional force opposes further motion proportional to the separation. When a contact point is in a sliding state, the virtual proxy slides along with it.

Hayward’s stick-slip model only uses displacements to determine state transitions [6]. If we define $z = x_k - x_{proxy}$ as the displacement between the real contact point and the proxy and z_{max} as the maximum displacement then the update for the next proxy point is $x_{proxy} = x_k \pm z_{max}$ if $\alpha(z)/|z| > 1$ (slipping), or $x_{proxy} = x_{proxy} + |x_k - x_{k-1}|/\alpha(z)/|z|$ otherwise (sticking). Once the displacement between the proxy and real contact point passes a maximum the contact becomes fully tense and enters the slipping state. The proxy point and the real contact point move together, separated by z_{max} . For displacements less than the maximum the proxy point does not move much; this is the sticking state. The non-linear adhesion map $\alpha(z)$ allows the proxy point to creep between these two regimes.

We have only selected two of many possible haptic force models that could be used by the AHI to represent surface properties. Stochastic models for haptic textures based on filtered noise and fractional Brownian motion are well known in the haptics community [5,18]. For example, perturbing the normal force by sampling from a Gaussian distribution generates haptic sandpaper. Friction models have a long history [1]. The algorithms and techniques we implement for planar normal and tangential forces are intended to be incorporated with more sophisticated applications that manage their own collision detection between complex polygonal geometries. After collision detection, these applications could use any suitable model for computing local normal and tangential forces.

AUDIO FORCE SYNTHESIS

Naively using the raw normal forces to synthesize audio produces unsatisfying results. There are two properties of our synthesized normal forces that cause trouble. This section will describe how we filter out these two properties from the haptic force. The filtered result is the *audio force* that we convolve with the stored impulse response in Equation 2. The two properties of the haptic force that we wish to filter are as follows: (1) a spurious impulse results when the user breaks contact with the surface and the haptic

force discontinuously drops to zero, and (2) high frequency position jitter.

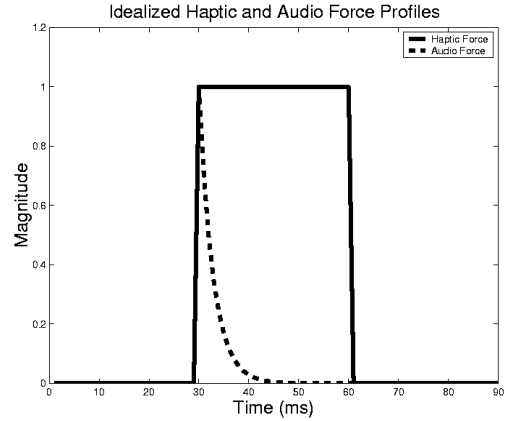


Figure 4(a): Idealized haptic and audio force profiles.

Figure 4(a) plots an idealized haptic contact force. At 30ms the user comes into contact with the surface and stays in contact for another 30ms. Convolution of this square wave profile with the impulse response of the surface will produce a spurious second “hit” when the user breaks contact. We introduce an attenuation constant β to allow the audio force to smoothly move to zero during sustained contact. If t is the elapsed time since contact, then the current audio force is the current haptic force attenuated by β^t . We have found that attenuating the audio force starting 10ms after a new contact with $\beta = 0.85$ (half-life of 5ms) produces good results. Waiting 10ms before decaying improves the quality of impulsive contacts, whereas decaying the audio force immediately upon contact excessively reduces the overall amplitude and dynamic range of the resulting audio signal.

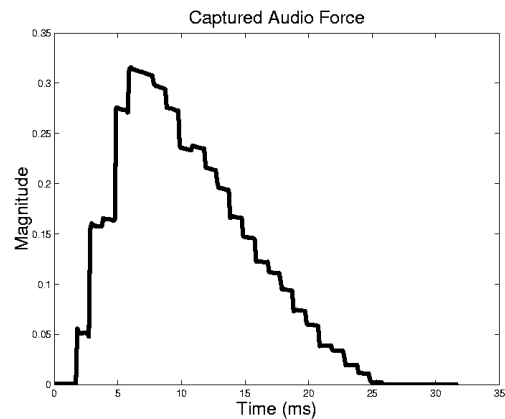


Figure 4(b): A captured audio force profile.

Haptic instabilities and signal noise generate sustained low amplitude, high frequency jitter in the position readings. These high frequencies are passed into the haptic force profile by the linear spring constant. Without filtering, this

noise becomes audible as crackling and popping while the user maintains a static contact with the surface. This low amplitude noise can be removed by truncation. Typically, we remove the 8 lowest order bits. Figure 4(b) plots a typical audio force profile. (The signal is not perfectly constant during each millisecond interval because it was captured as the input signal to a soundcard.) Using discrete optical encoders instead of potentiometers to read joint angles removes the need to truncate the position signals – the finite resolution of the encoders effectively truncates the signal for us.

REAL-TIME SIMULATION

The basic control structure for the AHI real-time synthesis and simulation is interrupt-driven. There is a haptic interrupt service routine (HISR) that generates haptic and audio forces and an audio interrupt service routine (AISR) that convolves the audio force with the impulse response of the modelled object. Using these two separate interrupts we can synthesize the audio signal at a much higher rate than we generate haptic feedback. This section will describe the two interrupt routines shown in Figure 5.

The AISR and all DAC/ADC latches are synchronized to trigger at the audio control rate by using a programmable interval timer that counts at half the bus clock rate of 8.33 MHz. The AISR reads the Pantograph joint angles from

the ADCs and stores them in an array that contains a history of joint angle readings. Converting the DAC input to an equivalent floating point number requires 1 comparison, 2 multiplications, and 2 additions. The current audio force is then clipped to lie between 0.0 and 1.0 and truncated to remove low amplitude noise. This requires 2 comparisons and 2 multiplications. A discrete convolution step using this filtered audio force $F(k)$ produces the output audio signal $y_n(k)$. This signal is placed in the DAC out. Computing the audio signal requires 2 multiplications and 3 additions per complex frequency, per sample. If the DSP is short on cycles, we can decrease the number of active frequencies. In our current scheme this isn't necessary -- there are no other competing processes for DSP time. Once a number of complex frequencies are selected the total amount of processing time is fixed and does not need to be adaptively adjusted. This would change if the DSP was also managing a complicated environment with graphics, collision detection, and rigid body dynamics.

Haptic interrupts trigger at an integer fraction of the audio control rate. The current joint angles are the mean of the array of joint angles captured during the AISR. From these filtered values, we use the forward kinematics of the Pantograph to compute the handle position. Since we only consider interactions with a plane, determining contact between the handle and the plane takes a sign check. If

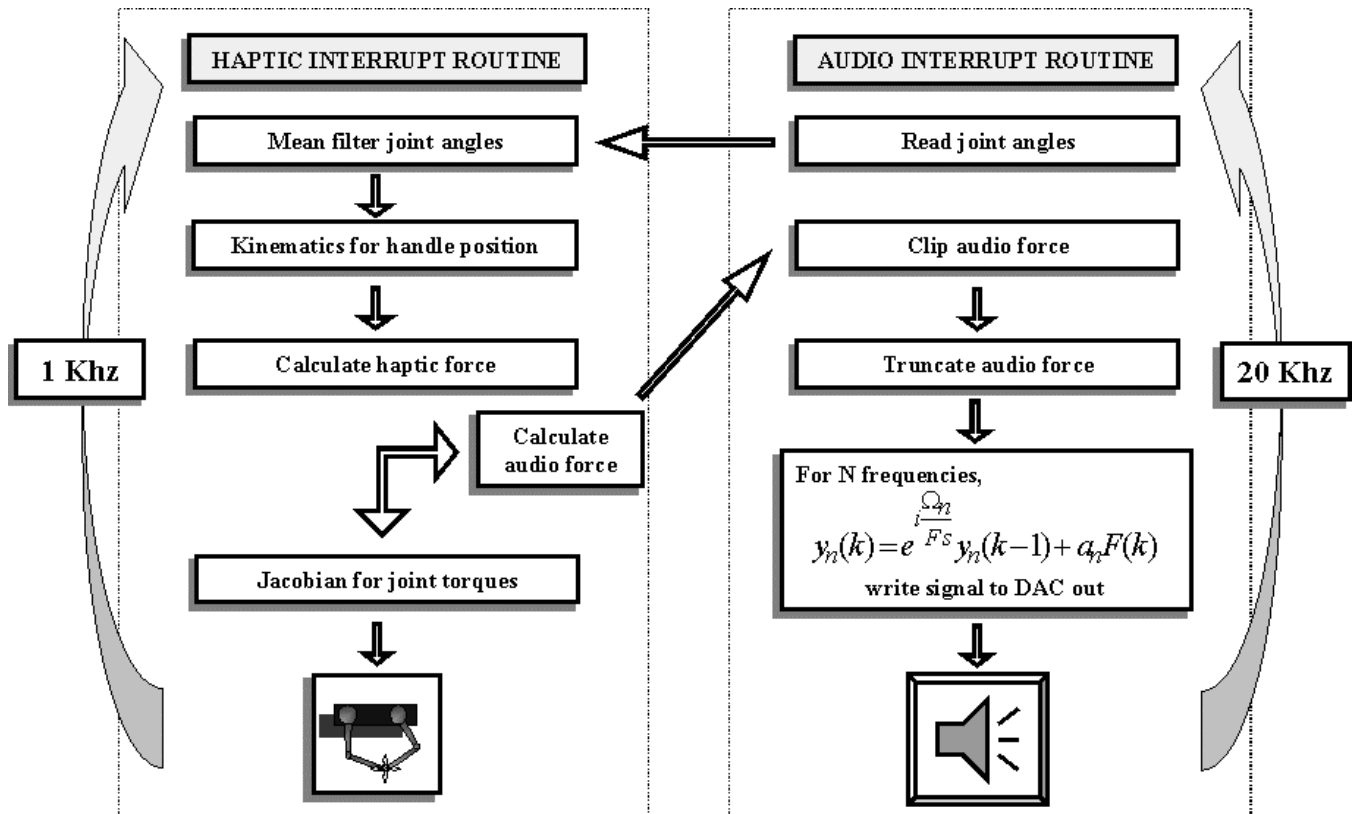


Figure 5: Flow of control for real-time synthesis and simulation.

there is contact, we compute the resulting normal force using a spring/damper/impulse model and a tangential force using a stick-slip friction model. Updating the current Jacobian takes 22 multiplications, 4 trigonometric calls, and 2 square roots. The Jacobian of the Pantograph translates the haptic force into motor torques. The voltages to generate these torques are written to the DACs. If there has been contact for $(10+t)$ milliseconds, then β^t times the normal force plus the tangential force becomes the current audio force. The HISR writes the current audio force to a global variable shared with the AISR.

CONTINUOUS CONTACT INTERACTIONS

We have experimented with some examples using the basic control structure just described to demonstrate how the AHI can generate continuous audio and haptic contact interactions. In the first example, the user scrapes the AHI handle across a sinusoidally modulated surface profile. In the second example, the user slides the handle across a surface with stick-slip friction. In both examples, we convolve the resulting audio force with the impulse response of a brass vase acquired from the University of British Columbia Active Measurement facility [15]. These examples have been informally tested in our laboratory. Haptic interrupts trigger at 1kHz and audio interrupts at 20kHz. Thus, there is a 1ms latency for changes in force and audio. Informally, the auditory and haptic stimuli are perceptually simultaneous.

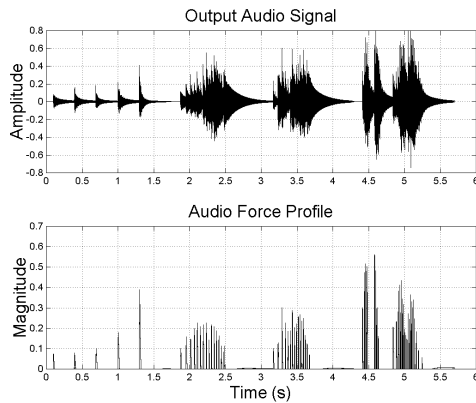


Figure 6: Output audio signal and audio force magnitude.

Figure 6 plots a captured audio force and the convolution of this force with the measured impulse response of the vase. We compute 20 modes for each audio sample. The user interaction in this example was five single strikes of increasing force normal to the surface, then tangential motion across the surface. The middle two bursts are slower scrapes back and forth, and the final two bursts are faster scrapes. The auditory signals produced in this fashion are satisfying. “Zipper” audio effects can be created by rapidly scraping on the surface. These synthesized audio signals compared favorably to live signals of tapping and

scraping along the ribbed surface of the vase with the tip of a pen.

Figure 7 shows captured audio signal and audio force magnitudes when interacting with the AHI and Hayward’s stick-slip friction model. Audio force decay was enabled for this captured signal, but only applied to the normal force component. The user interaction in this example was to slide the handle tangentially across a flat plane. A force hysteresis is apparent. The force increases as the displacement between the real and proxy contact point increases (sticking phase) and then discontinuously drops to zero during the slip phase. These discontinuities in the audio force create impulses in the audio signal. We used a similar model of a brass vase in this example to the one in Figure 6.

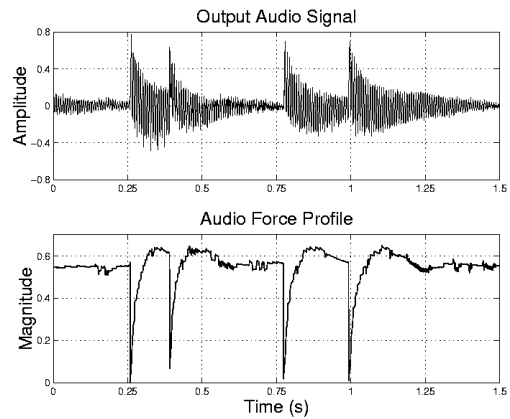


Figure 7: Output audio signal and audio force magnitude for interacting with the AHI and Hayward’s stick-slip friction model.

USER STUDY

The remainder of the paper will describe a user study we conducted with the AHI. In this user study we tested the hypothesis that a 2ms latency lies below the perceptual tolerance for detecting synchronization between auditory and haptic contact events. We selected 2ms because it is larger than our 0.5 system latency, but a small enough interval to establish a lower bound. Briefly, a subject tapped on a virtual wall and received appropriate auditory and haptic stimuli, except that one of the two was delayed by 2ms. We tested the hypothesis that all subjects would perform at chance when asked to choose which stimulus came first.

Participants

Twelve members of our department (2 females, 10 males) participated in the user study. Their mean age was 32 years, with a minimum age of 21 and a maximum age of 55. There was one left-handed participant. All twelve reported normal hearing and touch. The participants were not paid for their time.

Apparatus and Stimuli

With a few software modifications, we used the AHI as described in the first half of this paper. The stimuli consisted of 24 contact events where the audio signal preceded the haptic by 2ms, and 24 contact events where the haptic signal preceded the audio signal by 2ms. The haptic control rate was 2kHz and the audio control rate was 20kHz, resulting in a 0.5ms system latency. We created precedence by delaying one of the two signals with a short circular buffer. Delaying the haptic signal did not cause any instability. We disabled tangential forces.

A vertical “wall” was positioned in the right half of the workspace. One set of 48 random locations of the wall within $\pm 1\text{cm}$ were generated and used in the same order for all subjects. Haptic force-feedback was provided using the spring/damper/impulse combination. For audio, striking the wall was treated as striking a single point on an ideal bar, clamped at both ends. The contact point was at 0.61 of the length of the bar. For a given fundamental frequency, this simple geometry has an analytical solution for the frequencies and relative amplitudes of the higher partials. We used fundamental frequencies ω_1 of 1000Hz (High) and 316Hz (Low) and four higher partials.

As shown in equation 1, the decay of these frequencies are determined by a damping coefficient $d = f\pi \tan(\phi)$. We used two values of $\tau_d = 1/(\pi \tan(\phi))$ that correspond to slow decay and fast decay: 300 (Fast), and 3 (Slow). These particular auditory stimuli were selected because they are a subset of those used for a study that connects variation of the coefficient τ_d to auditory material perception [7].

Experimental Design

The experiment used a two-alternative forced choice design. The subjects were asked to decide whether the audio signal preceded the haptic signal, or the haptic signal preceded the audio signal.

A three within-subject design was used with audio/haptic precedence, frequency, and damping as the within-subject variables. In total, there were 8 different stimuli presented to the user. Audio precedence coupled with one of four frequency/damping combinations (High + Fast, Low + Fast, High + Slow, Low + Slow), and haptic precedence coupled with the same four sound combinations. Six of each of these 8 different types were permuted once and used in the same order for all subjects for a total of 48 stimuli.

Experimental Procedure

Subjects sat on a chair with the Pantograph on a desk in front of them. The Pantograph base was affixed to the desktop with a rubber sheet to minimize sliding and rotating. Subjects wore closed headphones (AKG K-240) for the audio signals and to minimize external sounds including those from the Pantograph device. We told the subjects that they would be striking a virtual wall, and that this would produce both haptic forces and audio signals.

They were told that, in each case, one of the stimuli preceded the other. We demonstrated how to hold the handle of the pantograph and how to make an impulsive strike to the wall. Then, the subjects were allowed to practice a few strikes with the headphones on. Finally, the experiment began.

Subjects were not told that there were equal numbers of stimuli types, nor were they told the number of repetitions in the experiment. No requirement on striking force was suggested -- subjects were free to strike the wall as hard or soft as they wished, as long as it was a single strike. There were no visual cues; however, the subjects were not blindfolded. After being read the the subjects were allowed to ask questions about the purpose of the experiment. If they expressed some concerns about their ability to discriminate between the two alternatives they were told that the discrimination task was designed to be difficult and to expect some ambiguity.

RESULTS

Table 1 shows the mean number of correct responses, along with the maximum, minimum, and standard deviation of correct responses out of 48. Table 1 also shows the number of audio responses selected. Four different subjects were responsible for each of the maxima and minima.

	# Of Correct Responses	# Of Audio Responses Selected
Mean	21.08	23.58
Max	26	31
Min	17	16
Std	2.57	4.52

Table 1: Number of correct responses, and number of audio responses selected out of 48, for all 12 subjects.

We test the null hypothesis that the subjects perform at the chance level (each response is a pure guess), for each of the 12 subjects. By hypothesis, the mean number of correct responses $\mu = 24$ and the standard deviation $\sigma = 3.45$. Using the normal approximation to the binomial distribution, we see that we can reject the hypothesis with a two-tailed test at the significance level $\alpha = 0.05$ only if the sample mean is outside the interval $\mu \pm 1.96\sigma = [17.21, 30.78]$. Except for the lone subject with 17 correct responses, we cannot reject the hypothesis that the subjects are performing at chance. We note that a one-tailed test may be more appropriate since we want to know if the subjects can detect the precedence *better* than chance. With a one tailed test, we can not reject the hypothesis for any of the subjects.

DISCUSSION

The results indicate that 2ms is a valid lower bound for the perceptual tolerance of synchronization latency of a contact interface like the AHI.

Other data suggests that the rate of decay of the auditory stimulus is a factor in the user responses, independent of the actual order of stimulus presentation. Table 2 contains the six most extreme number of correct responses, listed across stimuli. Sounds that decay more slowly are perceived as preceding the haptic stimulus and sounds that decay quickly are perceived as lagging the haptic stimulus, independent of the actual order of stimulus presentation. An increase in the audio decay rate (sounds decaying more quickly) reduces the total energy and total duration of the audio signal. It is possible that the subjects are using decay rate or total duration or total energy as a criterion for their response. The subject could be choosing the “loudest” signal (in terms of duration or energy) as the precedent stimulus. More studies are required to describe (and eventually understand) this effect.

Stimulus #	# Of Correct Responses	Stimulus Type
4	1	Audio + High + Fast
6	2	Haptic + High + Slow
11	10	Haptic + High + Fast
30	10	Haptic + Low + Fast
33	1	Haptic + Low + Slow
45	2	Haptic + Low + Slow

Table 2: The six most extreme values for number of correct responses, listed across stimuli. The number of correct responses is out of 12. Stimulus type is listed in order of precedence, frequency, and decay.

FUTURE WORK

In this last section we consider further opportunities for using the AHI in perceptual studies of integrated audio and haptics, specifically for exploring multimodal texture and friction. One of the basic questions about cross-modal synchronization has been addressed recently, but many questions about similarity and synchronization between dynamic vibration stimuli remain.

Levitin, et al., have helped to identify the perceptual tolerance for synchronization between auditory and haptic contact events [10]. Our original plan was to conduct a very similar study. Nevertheless, the AHI is well suited to help establish similar perceptual tolerances for continuous audio and haptic contact interactions such as scraping and sliding over textured surfaces.

In the Levitin study, subjects manipulated a baton. They would strike a horizontal surface (containing a capacitor) with this baton. By tracking position, velocity, and acceleration, the experimenters could predict the time of actual impact. Using these predictions, a digitized sample of a stick striking a drum was played back to the subject at random temporal offsets varying between -200ms and 200ms. Subjects were asked to judge whether the baton strike and the audio sample occurred at the same or different times. The threshold where subjects considered the stimuli to be synchronous 75% of the time

corresponded to -19 and 38ms; that is, the interval between the audio preceding the haptics by 19ms, and the audio lagging the haptics by 38ms. When adjusted for response bias (using confidence ratings) the corrected thresholds for detecting synchrony are -25ms and 66ms. Levitin’s practical motivation for this study was to determine an upper limit for reliable perceptual synchronization that gives system designers a well-defined performance target. The 66ms 75% performance threshold is higher than what we expected partially based on our own experience with the AHI.

There may be an important difference in synchronization tolerances between active haptic devices with motors such as the AHI and passive devices such as the baton used in the Levitin study. The AHI’s motors do not operate quietly. Motor torques excite structural vibrations which produce sounds. It is known that the time resolution for successive audio clicks is on the order of 2ms and that this value is largely independent of frequency [13]. An intra-modal judgement would use both audio signals (one from the speakers, one from the motors) over the cross-modal judgement that compares the audio signal to the haptic signal. If this intra-modal judgement dominates the cross-modal judgements then it will be necessary to decrease the asynchrony to well below Levitin’s reported figure. Identifying and controlling asynchronies for active devices like the AHI does not directly address cross-modal perception, but given the preponderance of active haptic devices entering the market it would still be a valuable result to derive.

Our current simulation generates audio forces from haptic forces normal and tangential to a locally flat patch. Large scale surface features can consist of a collection of polygons which use our flat patch algorithms after collision detection. The spring/damper/impulse penalty method effectively parameterizes the normal component as surface hardness. Limited versions of small-scale surface features — friction and texture — have been implemented and described in this paper. We want simple and effective parameterizations of the force components as friction and roughness variables that remain relevant for auditory perception.

Hayward’s stick-slip model has been applied to the real-time physical modeling of violin bow-string interactions [17]. Bow-string interactions are some of the oldest studied examples of stick-slip friction. A good test for adding friction to our audio synthesis routine would be to simulate bowing a violin string and forcing it to resonate. Our AHI simulations of this sort of behaviour are not convincing yet — the audio signal in Figure 7 sounds more like a series of impacts than like the squeaking we associate with stick-slip phenomena. We might also use this model for synthesizing rough textures by imposing Gaussian noise perturbation on either the separation between the proxy and real contact, or on the spring coefficient that produces friction forces. Our

experience is that we will need another stage of force prefiltering to control the auditory roughness signal. Perhaps we could leverage work done in computer graphics on synthesizing fractional Brownian motion and fractals to gain more control over our signals [4].

Previous studies on the influence of auditory stimuli on haptic perception have used audio samples or tones triggered by contact events. The study by Miner, et al, used pitched tones and attack envelopes to simulate hard and soft sounds [12]. They found that “the auditory stimulus did not significantly influence the haptic perception”. The study by DiFranco, et al, triggered audio samples of contact events they recorded by hand [3]. They found that “sound cues that are typically associated with tapping harder surfaces were generally perceived as stiffer”. Both of these studies focus on the perception of hardness, which is a function of force on the user’s hand. These studies do not explore the perception of surface roughness, which in addition to being a function of force, can also be a non-trivial function of interaction speed and surface geometry. The spatial characteristics of the surface dominate roughness perception when using a bare finger. The speed of interaction in this case does not affect roughness perception. However, dynamic vibration effects as a function of speed are reported when interaction is mediated by a rigid probe [9]. These effects are complex and not fully understood. Increasing speed tends to render surfaces as smoother; however, unlike perception with the bare finger, the current effect tended to reverse itself as the interelement spacing increased.

A texture model for any haptic interface (not just the AHI) could use these perceptual results to inform their implementation. However, because the AHI tightly couples the auditory stimulus with the underlying physical process of collision, a simple grid produces haptic and auditory textures that vary as a function of both force and speed (Figure 6). Previous studies on perceiving auditory and haptic textures with the bare hand suggest that the subject will use the haptic texture before the auditory texture for roughness discrimination [8]. The similar and synchronized stimuli rendered by the AHI could be used to extend these results for multimodal vibration effects when a rigid probe mediates interaction.

Devising and verifying new multimodal friction and texture models will require several psychophysical studies. We believe the AHI in its current state provides an excellent platform for experimenting with and understanding these models.

CONCLUSION

Designing compelling simulated environments is the high-level goal of this research. The representation and rendering of contact interactions comprises an essential component of any such simulation. Atomically representing the contact event as something that produces both sound

and force helps integrate auditory and haptic stimuli. We believe this is a natural way to think of representing and simulating contact. We have implemented a new interface that can render audio and haptic interaction atomically.

Our experimental results suggest that the AHI’s overall latency and synchronization between the auditory and haptic modes lies below the perceptible threshold. In the future, we will use the AHI to help explore perceptual synchronization and similarity between continuous auditory and haptic contact interactions.

ACKNOWLEDGMENTS

Thanks to Susan Lederman for valuable discussions. Thanks to Paul Kry for timely debugging wisdom. This work was supported in part by grants from the IRIS NCE, and Natural Sciences and Engineering Research Council of Canada. Derek DiFilippo is supported by an NSERC scholarship and in part by the Advanced Systems Institute of British Columbia.

REFERENCES

1. Armstrong, B., Dupont P. and Canudas de Wit C. A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction. *Automatica*. 30,7 (1994), 1083—1138.
2. Cook, P.R. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*. 21,3 (1997), 38—49.
3. DiFranco, D.E. and Beauregard, G.L. and Srinivasan, M.A. The Effect of Auditory Cues on the Haptic Perception of Stiffness in Virtual Environments. *Proc. ASME Dynamic Systems and Control Division*, (1997).
4. Ebert, D.S., Musgrave, F.K., Peachey, D., Perlin, K. and Worley, K. *Texturing and Modelling*. AP Professional, New York NY, 1998.
5. Fritz, J.P. and Barner K.E. Stochastic models for haptic texture. *Proc. SPIE Int. Symp. on Intelligent Systems and Advanced Manufacturing*, Boston MA, November 1996.
6. Hayward, V. and Armstrong, B. A new computational model of friction applied to haptic rendering. *Experimental Robotics VI*, Lecture Notes in Control and Information Sciences, Springer-Verlag, NY, Vol. 250, 403—412.
7. Klatzky, R.L., Pai, D.K. and Krotkov, E.P. Hearing Material: Perception of Material from Contact Sounds. To appear in *Presence*, October 2000.
8. Lederman, S.J. Auditory Texture Perception. *Perception*. 8 (1979), 93—103.

9. Lederman, S.J., Klatzky, R.L., Hamilton, C.L. and Ramsay, G.I. Perceiving Surface Roughness via a Rigid Probe: Effects of Exploration Speed and Mode of Touch. *The Electronic Journal of Haptics Research*, 1, 1, (1999).
10. Levitin, D.J., MacLean, K. and Mathews, M. The Perception of Cross-Modal Simultaneity. To appear in *Int. Journal of Computing Anticipatory Systems*, 2000.
11. Murray, R.M., Li, Z. and Sastry, S.S. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Ann Arbor MI, 1994.
12. Miner, N., Gillespie, B. and Caudell, T. Examining the Influence of Audio and Visual Stimuli on a Haptic Display. *Proc. of the 1996 IMAGE Conf.*, Phoenix AZ, June 1996.
13. Pierce, J. Hearing in Time and Space. Chapter 8 of *Music, Cognition, and Computerized Sound*. The MIT Press, Cambridge MA, 1999.
14. Ramstein, C. and Hayward, V. The Pantograph: a large workspace haptic device for a multi-modal Human-computer interaction. *Conf. on Human Factors in Computing Systems ACM/SIGCHI*, Boston MA, April 1994.
15. Richmond, J.L. and Pai, D. K. Active Measurement of Contact Sounds. *Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation*, San Francisco, April 2000.
16. Salcudean, S.E., and Vlaar, T. On the Emulation of Stiff Walls and Static Friction with a Magnetically Levitated Input-Output Device. *ASME J. Dynamic Systems, Meas., Control.* 119 (1997), 127—132.
17. Serafin, S., Vergez, C. and Rodet, X. Friction and Application to Real-time Physical Modeling of a Violin. *Int. Computer Music Conf.*, Beijing, October 1999.
18. Siira J. and Pai D.K. Haptic Textures — A Stochastic Approach. *IEEE International Conference on Robotics and Automation*, Minneapolis MN, April 1996.
19. Srinivasan, M.A. and Basdogan C., Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges. *Comput. & Graphics.* 4, 21 (1997), 393—404.
20. van den Doel, K. and Pai, D.K. The Sounds of Physical Shapes. *Presence* 7,4 (1998), 382—395.