Haptic Guidance Benefits Musical Motor Learning

Graham Grindlay*
Media Laboratory
Massachusetts Institute of Technology

ABSTRACT

This paper presents the results of a pilot experiment looking at the effect of haptic guidance on musical training. A percussion performance task was used where subjects learned to play short rhythmic sequences on a device capable of recording drumstick movements with a high degree of spatiotemporal accuracy. Subjects learned to perform the sequences under three primary training paradigms: listening to the rhythm (audio), being guided through the motions involved in the rhythm's performance (haptic), and being guided through the required motions while listening to the resulting sound (audio+haptic). Performance was assessed in terms of both timing and loudness (velocity) accuracy using several different metrics.

Results indicate that haptic guidance can significantly benefit recall of both note timing and velocity. When subject performance was compared in terms of note velocity recall, the addition of haptic guidance to audio-based training produced a 17% reduction in final error when compared to audio training alone. When performance was evaluated in terms of timing recall, the combination of audio and haptic guidance led to an 18% reduction in early-stage error.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Haptic I/O; K.3.0 [Computers and Education]: General—

1 INTRODUCTION

Humans acquire new motor skills through a multi-stage learning process [7, 8], central to which is of course practice. Through trial and error, we continually refine our motor skills achieving better and more consistent performance. But in order for practice to be productive, a reliable means of evaluation is required, making feedback of paramount importance. Feedback can take many different forms in motor learning applications, including verbal communication (i.e. knowledge of results), visual and auditory signals [14], and vibrotactile stimulation [15]. Although different in sensory modality, these types of feedback are all indirect, meaning that the information they provide must be translated from a sensory coordinate system to the kinematic/proprioceptive coordinate system. For complex tasks this translation is likely to be difficult, particularly in the early stages of learning. A more direct form of feedback is *haptic guidance* where the learner is physically moved along the trajectory of the target motion [6]. Because the learner experiences the exact proprioceptive feedback that he/she would during a correct execution of the target task, no translation is necessary. The question of course is what impact haptic guidance has on learning and how it compares to other forms of feedback and training.

Previous research on haptic guidance has largely focused on comparing visual and haptic training paradigms. There is no existing work (that I am aware of) which examines the relationship between haptic guidance and auditory information for motor learning. In this paper, this area is explored through the use of musical performance tasks which naturally relate both sound and haptics. This approach is motivated both by the desire to utilize an ecologically valid task as well as by the potential practical implications of the research. Learning to play a musical instrument is no doubt one of the most complex and challenging human endeavors, but because of this complexity, music pedagogy suffers from a difficulty in effective task communication. Complex movements are hard to verbalize and it is therefore challenging for music teachers to communicate movements to their students. Historically, teachers have often resorted to a manual form of haptic guidance by moving their students' hands through the desired motions. However, this provides at best a rough approximation of the target movement and begs the question of whether technology might be leveraged to provide a more accurate form of haptic guidance. The success of such a system could lead to significant advancements in music pedagogy by speeding and easing the learning process and providing a more effective means of home instruction.

2 RELATED WORK

Haptic guidance, sometimes called mechanical guidance, manual guidance, as well as other names in the literature, refers to concurrent augmented feedback where the learner is moved, both temporally and spatially, through an ideal rendition of the task motion.

Early research in this area generally made use of fairly simple reaching or linear positioning tasks [12, 13]. One notable exception is the work of Armstrong who used a complex elbow movement task to compare haptic guidance, knowledge of performance (KP) delivered concurrently using a visual display, and KP delivered at the end of each trial [2]. He found that while the physical guidance and concurrent KP training conditions had superior performance during the trials, they were worse than terminal KP in a retention test. It should be noted, however, that in Armstrong's study, each of the training conditions used 100% relative feedback frequency.

Later research largely focused on comparing haptic guidance to different forms of visual guidance. Yokokohki et al. proposed several different combinations of haptic and visual guidance as part of a record-and-playback system that they called, "What You See Is What You Feel." [21] Although they did not conduct formal experiments, a very preliminary test using a virtual cube manipulation task did not yield any conclusive results. They speculate that this may have been due to the task being too easy.

Gillespie et al. developed a system called the *Virtual Teacher* to test haptic guidance in a crane-moving task [9]. This device consisted of a free-swinging pendulum attached to a cart which could be slid along a track. The task involved setting the pendulum into motion by moving the cart and then trying to stop the pendulum from swinging as quickly as possible. The optimal movement strategy, which involves first injecting energy into the system and then removing it after a carefully timed interval, was demonstrated to some subjects while others (the control group) simply tried to learn the system dynamics on their own. Although they did not observe any statistically significant advantage of the guidance-trained groups over the control group, guidance did seem to effectively communicate the basic components of the optimal strategy. The authors conjecture that the optimal strategy was probably too difficult to master and that better results might have been obtained if the *Virtual Teacher* had demonstrated the components of the optimal strategy individually.

---

*e-mail: grindlay@mit.edu

Several recent studies have compared the effects of haptic and visual guidance for learning. Feygin et al. looked at these types of guidance using complex sinusoidal task movements [6]. Subjects learned three-dimensional spatial trajectories under several different training conditions (haptic, visual, haptic and visual) and then had to manually reproduce them under two different unassisted recall conditions (with vision, without vision). The experiment contained 15 trials for each combination of training and recall conditions where each trial consisted of two training (presentation) runs followed by a test (recall) run. Performance was measured during each of the recall runs using several different error metrics, including position, shape, timing, and drift. They found that subjects significantly improved their performance in all training conditions under the position and shape metrics, but not under the drift or timing metrics. In terms of performance averaged over the last five trials, haptic training alone was less effective than visual training under the position and shape metrics, but more effective under the timing metric. Recall mode only affected timing and drift (marginally) metrics with the addition of vision benefiting performance. Training and recall mode were found to interact such that performance under haptic training modes decreased when vision was included in the recall condition. The authors suggest that this interaction may be because vision overpowers proprioception, degrading its effect. A separate analysis of haptic guidance and visual training indicated that while position and shape accuracy were predominantly affected by vision, timing accuracy was largely affected by haptic guidance. The finding that haptic guidance benefits timing accuracy irrespective of whether visual information is present, agrees with previous research on observational learning [4, 3].

Recently, Liu et al. re-examined some aspects of the Feygin study, but altered the protocol to make it more similar to a rehabilitation context [17]. One of the more significant changes that they made was to the trial structure. Instead of each trial consisting of two practice runs through the task motion followed by a test run as in the Feygin et al. experiment, each trial in the Liu et al. experiment consisted of seven practice runs followed by seven test runs. This allowed for an examination of learning during repeated unguided practice. Other differences from the Feygin et al. study is that Liu et al. only considered recall with vision and they only looked at position error. Although they found that all training conditions produced a significant improvement between the first and last trials, they did not find a significant difference between training with and without haptic guidance (in fact vision alone was marginally better). Additionally, they found that subject performance degraded over the course of the test runs in each trial with movements gravitating towards an “attractor path”. Despite the fact that they did not measure timing error, making a comparison of the positive haptic guidance results found by Feygin et al. impossible, these results largely confirm those of Feygin et al.

3 METHODS

3.1 The Haptic Guidance System

The experimental apparatus constructed for this experiment is referred to as the *Haptic Guidance System* (HAGUS). The system was designed to record and playback rotational motions about a single axis (targeting wrist movement) with a high degree of spatiotemporal accuracy. Although percussion performance in general certainly isn’t restricted entirely to the wrist, I believe that this simplification provides a reasonable first-order approximation. Additionally, it should be noted that particularly for persons with no prior percussion experience, there are many non-trivial rhythmic tasks possible with a single degree of freedom.

3.1.1 System Design

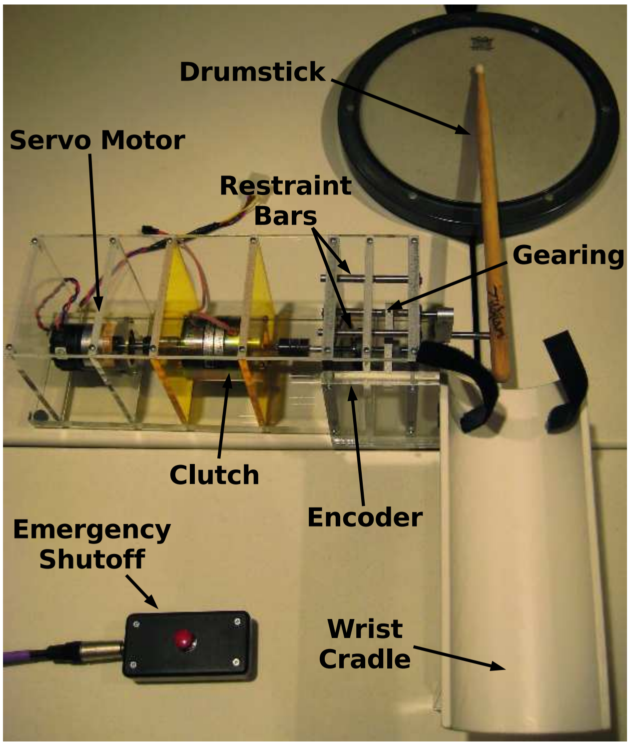


Figure 1: The Haptic Guidance System hardware

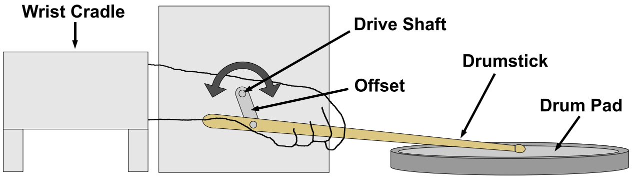


Figure 2: Diagram of HAGUS actuator.

HAGUS uses a combination of onboard electronics and PC-based computing power. The actuator (see Figure 1) consists of a 40 Watt servo motor[1] connected via an electromagnetic particle clutch[2] to the primary drive shaft. This drive shaft connects to a high resolution optical encoder[3] and is then geared-up (1:3 ratio) before connecting to the drumstick. Backlash in the system was minimal and corresponded to roughly 0.16 degrees of play.

The servo motor is run by motion control hardware[4] running a proportion-plus-derivative (PD) control filter. The filter was updated at a frequency of 1.953kHz and a small amount of deadband compensation was used to reduce jitter.

Position measurements were recorded and played back at a 60Hz sampling rate. To record a motion, the clutch is first disengaged to disconnect the servo motor from the rest of the drive train. This minimizes the amount of physical impedance presented to the user when he/she is freely playing. During recording, the motion control electronics stream encoder readings to the host PC over a USB con-

---

[1] Sanyo Denki Super V, model V404-011, 24V, 2.9Amp DC servo motor

[2] Placid Industries model C5-24 electromagnetic particle clutch; 80oz-in of holding torque and 1oz-in of drag torque.

[3] RENCO model RCML15 2000 line quadrature encoder; effectively giving 8000 counts/revolution or 0.045 degree accuracy.

[4] PIC-SERVO SC, Jeffrey Kerr, LLC., `http://www.jrkerr.com`

nection which then logs these data to disk. To playback a previously recorded motion, the direction of information is simply reversed; the host PC streams the position data from a file to the control electronics where they are used with the PD filter loop described above to reproduce the motion.

Along with a powersupply and servo control boards, the HAGUS system electronics include an *Arduino*[5] board for general I/O purposes. The *Arduino* was primarily used to control a set of LED lights which in turn were used during the experiment described in Section 3.5 to cue subjects as to when to expect motion playback to begin and when they should begin playing back a rhythm.

The complete HAGUS setup is shown in Figure 1. This figure shows the forearm cradle and wrist strap which are mounted to the end of the actuator and ensure that all subjects are positioned ergonomically and consistently. Figure 2 shows a side view of the device and illustrates how users interact with it. The drumstick is attached to the drive shaft via an offset shaft which helps to better model natural drumstick rotation. Several safety features are also included in the HAGUS design (again, see Figure 1), such as an emergency shutoff switch and restraint bars used to physically prevent any drumstick motion outside of a safe range (40 degree).

High-level control of the HAGUS setup is the responsibility of a host PC. Software was developed to handle all aspects of streaming position data to and from HAGUS as well as organizing and running experimental sessions. It provides control over a number of PD filter loop settings as well as low-latency scheduling of position data. The experimental design features were intended to reduce the amount of experimenter intervention required and minimize the chance of human error. Automating control of the presentation and recall runs has the added benefit of making the experimental flow more consistent across subjects. Other features include error logging and the ability to schedule practice runs.

3.2 Experimental Design

The primary purpose of the experiment was to investigate the differences between haptic guidance and audio-based training on percussion learning. The hypothesis being tested was that haptic guidance combined with auditory feedback would result in participants being able to reproduce rhythms more accurately than either auditory feedback or haptic guidance alone. Differences in accuracy were measured in terms of note timing as well as note velocity.

Subjects were trained to perform four different rhythms under four different training conditions. During each trial, the subject was run through two training presentations of the task rhythm which was then followed by a recall run where the subject tried to reproduce the task rhythm with no assistance. Each training condition consisted of 15 consecutive trials followed by a five minute break.

The experiment compared four different training techniques, three primary and one ancillary. The first primary training technique was an audio-only (A) condition. During this training condition, subjects did not move their hands, but only listened to a recording of the task rhythm being played by the HAGUS device. This condition was designed to mimic a typical at-home self-instruction situation where a student may have an instructional book and CD with audio examples. The second primary training technique was a haptic guidance only (H) condition. During this condition, subjects were physically moved through the motions required to perform the task rhythm, but were unable to hear. Subjects wore -32dB earplugs as well as headphones which played white noise masking sound. A pilot experiment testing this setup confirmed its efficacy at preventing subjects from hearing drumpad sounds. The third primary training condition (A+H) was a combination of the first two where subjects were physically guided through the ideal task motion but were also able to hear the results.

Ideally, all conditions would have identical recall run setups. However, this was not possible as it would have meant that the subjects would need to remove the earplugs and headphones before each of the H condition's recall runs. Because the masking noise was necessary to effectively prevent hearing, subjects were allowed to leave the earplugs in and headphones on during the haptic guidance only (H) recall runs. Even though subjects reported being able to hear the drumpad fairly well with this setup, a fourth ancillary training condition was included to test for the effects of attenuated hearing with the presence of the earplugs and headphones (when not playing the masking noise). This condition (A+H(atten)) was similar in all ways to the A+H condition with the exception that subjects wore earplugs and headphones (without masking noise) throughout the condition.

Each of these four training conditions was considered a within-subjects factor in a repeated measures experiment. Subjects were randomly assigned to one of four groups and a balanced Latin square design was used to order the training conditions differently for each group. Four different rhythmic tasks were devised to prevent learning transfer between training conditions. The assignment of rhythmic task to training condition was varied across groups using a balanced Latin square design.

3.3 Subjects

Thirty-two right-handed subjects (20 females and 12 males) between the ages of 18 and 50 (the median age was 27) were recruited for the study. Most subjects were members of the MIT community. While none of the subjects had any percussion training or significant playing experience, some did have training and/or experience with other instruments.

3.4 Tasks

Four different rhythmic tasks were used in this experiment. Each was designed to be non-trivial yet learnable within the 15 trial period of each condition in the experiment. Each rhythm contains eight notes (one quarter note, three eighth notes, three sixteenth notes, and one dotted-eighth note). A small pilot study with three subjects suggested that these rhythms were of an appropriate level of difficulty (each subject's data showed reasonable learning curves).

Figure 3: Example rhythmic task.

Template audio and haptic performances of each of these rhythms (the target standard that the subjects were trained on and therefore compared against) were generated by the author. A tempo of 80 beats-per-minute was used, which meant that the shortest notes (sixteenth-notes) were 0.1875 seconds in duration and each rhythm was exactly 3.0 seconds total. The minimum note duration (0.1875 sec = 5.33Hz) was chosen so as to fit within known proprioceptive and motor system bandwith limits (5-10Hz) [5, 18]. The haptic guidance templates were produced by playing each rhythm on the HAGUS device while listening to a quantized audio rendition of that rhythm (to ensure accurate timing).[6]

---

[5] `http://www.arduino.cc`

[6] Although the use of artificially constructed motion sequences (i.e. sequences of single stroke motions stitched together) would be advantageous in terms of rhythmic precision, it is not clear *prima facie* how this could be done while ensuring that the sequences are ergonomically sensible. Therefore, it was decided that it was preferable to use non-quantized human performances rather than risk potentially awkward and unnatural task motions.

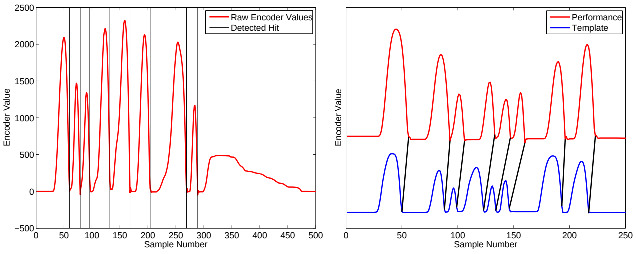
Several takes of each rhythm were recorded and the best one (to my ears) was retained as the template for that rhythm. An audio template for each rhythm was also produced by playing the haptic guidance template back on the HAGUS setup and recording the sound that was produced. This ensured that the audio used in the A training condition closely matched the audio that was produced during the A+H and A+H(atten) conditions.

3.5 Procedure

Subjects were first familiarized with the purpose of the study and equipment. Verbal instructions were given and informed consent was obtained. Each subject then practiced one trial of the A training condition and one trial of the H training condition (the A+H and A+H(atten) were judged to be similar enough that practice was unnecessary). Each trial consisted of two presentation (training) runs immediately followed by a recall (test) run. Subjects were instructed to "play along" with the HAGUS device during the training runs that included haptic guidance (H, A+H, and A+H(atten)) and to just listen during the A training condition. Each subject was also instructed to try to reproduce the task rhythm as accurately as possible in all respects during the recall runs. There were 4 training conditions, each of which consisted of 15 trials. A set of LED lights was used to cue subjects as to when each training and testing run was about to begin. This light sequence took 3 seconds to complete which, when added to the 3 second task duration and a 1 second delay after each task, produced a total run length of 7 seconds. Given 3 runs (2 training and 1 test) per trial, each trial lasted 21 seconds and each training condition lasted 5 minutes and 15 seconds (15 trials per condition). Subjects completed each training condition (all 15 trials) and then were given a short (5 minute) break. After all 4 training conditions had been completed, subjects filled out a brief questionnaire.

3.6 Preprocessing & Measurement

3.6.1 Onset Detection



(a) Detection (b) Alignment

Figure 4: Example of onsets detected in raw encoder data (a) and alignment of performance and template data (b).

The data from each recall run collected during the experiment was recorded by HAGUS and logged by a host PC. However this data consists of raw encoder readings which provide a very accurate history of the drumstick movements but are too low-level to be useful in assessing percussion performance. Therefore, we need to translate the encoder data into symbolic form by finding the times at which each drum stroke occurred. This was done using a relatively straightforward trough-picking algorithm, although in practice data pathologies necessitated several ad hoc modifications to prevent false onset detections. After the sequences of onset times were extracted, they were normalized to begin at time 0 (i.e. each element of each sequence had the first element of that sequence subtracted from it). Figure 4(a) shows a representative example of raw encoder data and the onsets detected in that data by the algorithm.

3.6.2 Stroke Velocity

Using the extracted onset times in conjunction with the raw encoder data, it is fairly simple to generate measurements of stick velocity. Pseudocode for this algorithm is given in Algorithm Listing 1.

Algorithm 1 Stroke velocity detection algorithm

```
Require: $onsetTimes[1:T], encoderData[1:N]$
  $velocities[1:T] \leftarrow []$
  for $t \leftarrow 1:T$ do
    if $t == 1$ then
      $lastIndex = 1$
    else
      $lastIndex = onsetTimes[t-1]$
    end if
    $window = encoderData[lastIndex : onsetTimes[t]]$
    $maxIndex = indexOfMaxValue(window)$
    $\Delta encoderData = encoderData[lastIndex + maxIndex - 1] -$
                    $encoderData[onsetTimes[t]]$
    $velocities[t] = \frac{\Delta encoderData}{onsetTimes[t] - (lastIndex + maxIndex - 1)}$
  end for
  return $velocities$
```

3.7 Scoring

Once we have obtained the symbolic representation of a rhythmic performance we need to find a way to compare it to its template rhythm. However, assessing the total similarity or difference between two performances of a rhythm is a fairly difficult task. There are several dimensions (number of notes, length in seconds, loudness, etc.) in which the rhythms may differ and it is unclear how these differences should be combined into a single metric. Instead of trying to devise a universal comparison metric, several different evaluation metrics were used; one to assess velocity accuracy and two to examine different aspects of timing precision.[7]

3.7.1 Timing Metrics

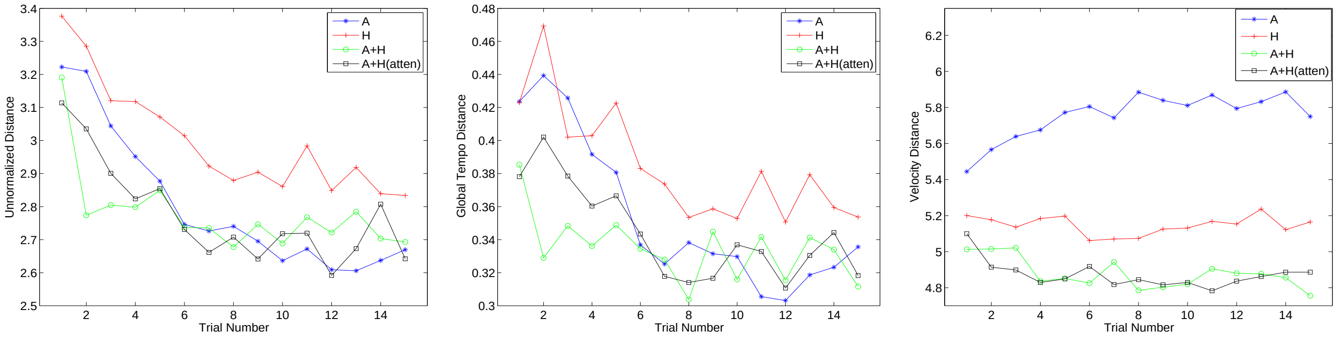
Algorithm 2 Unnormalized rhythmic distance algorithm

```
Require: $template[1:N], perf[1:M]$
  $DTW[1:N, 1:M] \leftarrow 0$
  $DTW[1, 2:M] \leftarrow \infty$
  $DTW[2:N, 1] \leftarrow \infty$
  $cost_{ins} \leftarrow \frac{1}{2} max(template[N], perf[M])$
  $cost_{del} \leftarrow \frac{1}{2} max(template[N], perf[M])$
  for $i \leftarrow 2 \ldots N$ do
    for $j \leftarrow 2 \ldots M$ do
      $cost_{dist} = |template[i] - perf[j]|$
      $DTW[i,j] = min(DTW[i-1,j] + cost_{dist} + cost_{ins},$
                    $DTW[i,j-1] + cost_{dist} + cost_{del},$
                    $DTW[i-1,j-1] + cost_{dist})$
    end for
  end for
  return $DTW[N,M]$
```

The first and most complete timing metric, referred to as the *unnormalized distance* (UD), uses a variant of the well-known *dynamic time warping* algorithm [19]. This technique allows for the alignment and comparison of two sequences (see Figure 4(b)) of possibly different length and produces a scalar number representing their distance/similarity. The cost of matching any two elements (onset times) between sequences is taken to be the absolute value of

---

[7] In fact two other timing metrics were analyzed as well, but there is not room to report them here. Interested readers are referred to my thesis [10].



(a) Recall under the UD timing metric

(b) Recall under the GT timing metric

(c) Recall under the VD velocity metric

Figure 5: Comparison of recall performance under the four training conditions as measured under the three different distance metrics. Each plot shows the recall curve for its training condition averaged across subjects.

their difference and an additional cost to insert or delete elements is also included to reflect the severity of this type of error. An extra cost of 50% of the total length (in seconds) of the longer of the two sequences was used for both insertions and deletions. Pseudocode for the algorithm is given in Algorithm Listing 2.

The second comparison metric, referred to as the *global tempo distance* (GT), provides an isolated measure of the global tempo similarity between two performances of a rhythmic sequence. This is done by comparing the final onset times in both sequences using the following symmetric function:

$$GT(template[1:N], perf[1:M]) = \left| log_2 \frac{perf[M]}{template[N]} \right| \qquad (1)$$

3.7.2 Velocity Metric

Because the velocity measurements do not represent temporal measurements like the onset data do, the distance metrics described in the previous section are not appropriate. Instead, the sum of the absolute values of the differences between performance velocities and template velocities was used. However, we still face the difficulty of sometimes having to deal with sequences of differing lengths. Although this suggests a dynamic programming approach, the nature of the velocity data prevents the direct use of dynamic time warping. The solution that was chosen was to save the performance-to-template mapping obtained during the alignment of timing data under the UD metric. This information can then be used to determine which elements of the performed timing data (and therefore the velocity data as well) were missing or are extra.

Algorithm 3 Velocity distance function

```
Require: $template[1:N], perf[1:M],$
         $tMatch[1:max(N,M)], pMatch[1:max(N,M)]$
  $distance = 0$
  for $i \leftarrow 1 : max(N,M)$ do
    $distance = distance + |template[tMatch[i]] - perf[pMatch[i]]|$
  end for
  return $distance$
```

Now we can perform the straightforward sum of absolute differences on the velocity data that has been matched using the timing data and we can also add in extra cost for insertions or deletions. In practice, additional insertion/deletion penalties were not included as they did not appear to significantly affect the results. The pseudocode for this distance metric, which is referred to as the *velocity distance* (VD), is given in Algorithm Listing 3.

4 RESULTS

Differences in training conditions were assessed by both looking at learning curves as well as differences between conditions at different points in the trial sequence. Learning was evaluated using paired t-tests while early and late trial performance was evaluated using repeated-measures analysis of variance (ANOVA) tests with training condition (A, H, A+H, A+H(atten)) as the within-subjects factor. Separate ANOVAs were run for each type of distance metric and then Bonferroni-corrected pair-wise t-tests were used to compare A+H and A+H(atten) as well as each of the possible pairings of primary training conditions ({A,H}, {A,A+H} and{H,A+H}) [1].

4.1 Learning Across Trials

Figure 5 shows the average (arithmetic mean across subjects) recall curves for the three distance metrics. These curves give an overall sense of differences between training conditions and distance metrics. For the timing metrics, we see general trends in error reduction across trials, providing evidence that, on average, subjects learned to improve their performance. The VD metric on the other hand, shows the somewhat puzzling trend towards *worse* performance across trials when subjects were trained using audio-only (A) guidance.

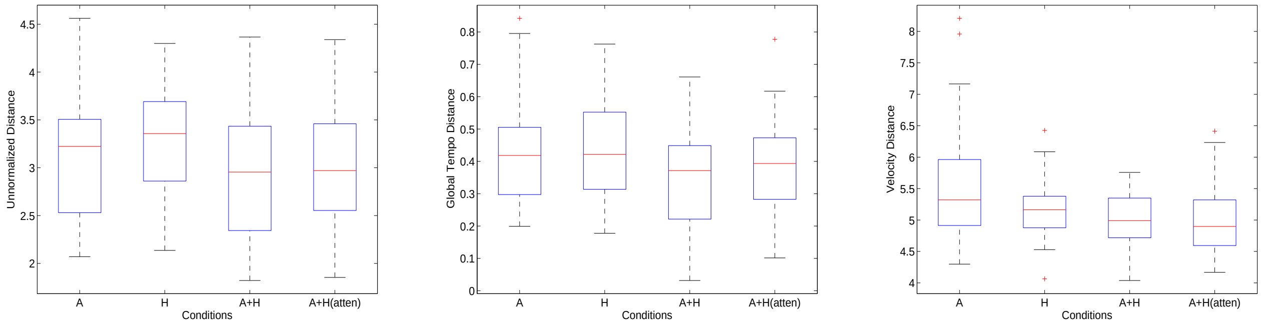
	Unnormalized	*Global Tempo*	*Velocity*
A	$p < 0.001$ (Y)	$p < 0.069$ (M)	$p < 0.260$ (N)
H	$p < 0.002$ (Y)	$p < 0.159$ (N)	$p < 0.798$ (N)
A+H	$p < 0.028$ (Y)	$p < 0.159$ (N)	$p < 0.046$ (Y)
A+H(atten)	$p < 0.006$ (Y)	$p < 0.158$ (N)	$p < 0.153$ (N)

Table 1: Pair-wise t-test results for comparisons between the first and last recall runs. p-values for each combination of training condition and distance metric are given along with whether the value is significant (Y), not significant (N), or marginally significant (M).

Paired t-tests were used to examine performance improvement between the first and last trial of each training condition and under each distance metric. These tests, which are summarized in Table 1, show significant improvement for all training conditions under the UD metric. Under the VD metric, the A+H training condition showed a significant difference between first and last trials, while under the GT metric a marginally significant improvement was found for the A training condition.

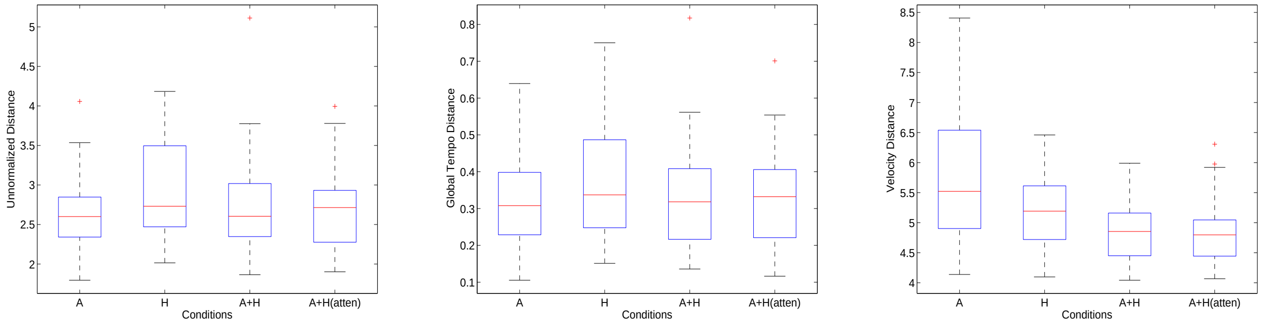
4.2 Early Trials

Although one might expect that subject performance on the first trial would not vary significantly between training conditions, the



(a) Recall under the UD timing metric (b) Recall under the GT timing metric (c) Recall under the VD velocity metric

Figure 6: Boxplot summaries of early trial data distributions. Boxes have lines at the lower quartile, median, and upper quartile, whiskers indicate the extent of the data, and the '+' symbols represent outliers.



(a) Recall under the UD timing metric (b) Recall under the GT timing metric (c) Recall under the VD velocity metric

Figure 7: Boxplot summaries of final trial data distributions.

performance curves in Figure 5 suggest a possible difference in early performance. They also show that there is some instability in the earlier sections of some of the recall curves. Therefore, to give a representation of the "early" section of training performance (as opposed to the potentially noisy and misleading first trial), the mean of the first three trials worth of data was used. Next, a repeated-measures ANOVA (again with training condition as the within-subjects factor) was performed using this data and a separate ANOVA was run for each of the three distance metrics. The results are given in Table 2 and the distributions are summarized in Figure 6.

	Unnormalized	*Global Tempo*	*Velocity*
p	< 0.0195	< 0.0479	< 0.0012
Significant?	Y	Y	Y

Table 2: Summary of results of a repeated-measures ANOVA with training condition as the independent variable and early trial recall performance as the dependent variable.

There was a significant main effect of training condition for all distance metrics. Bonferroni-corrected t-tests were used to both confirm that the A+H and A+H(atten) were statistically similar and allow for further analysis of the ANOVA results.

Under the UD metric, there was a statistically significant ($p < 0.003$) difference between the H and A+H conditions which corresponded to a 10% error reduction for the A+H versus H training condition.

Marginally significant differences were found between the A and A+H conditions ($p < 0.0198$, 18% error reduction for the A+H condition) and between the H and A+H conditions ($p < 0.0170$, 18% error reduction for the A+H condition) under the GT metric.

Under the VD metric, the A and A+H conditions differed significantly ($p < 0.0009$) which corresponded to a 10% reduction in error under the A+H training condition. The difference between the A and H conditions was of marginal significance ($p < 0.0164$) with 7% less error under the H training condition than the A training condition.

4.3 Final Trials

Next, the differences between training conditions at the end of the trial sequences were examined. As with the first trials, the recall curves were somewhat noisy and so the mean of the final five trials worth of data was used. Repeated-measures ANOVAs with training condition as the within-subjects factor were run (one per distance metric) to check whether there were statistically significant differences between the final sections of trials. Table 3 summarizes the results and Figure 7 summarizes the data distributions.

	Unnormalized	*Global Tempo*	*Velocity*
p	< 0.0373	< 0.2813	< 0.0001
Significant?	Y	N	Y

Table 3: Summary of results of a repeated-measures ANOVA with training condition as the independent variable and the final trials' recall performance as the dependent variable.

While both the UD and VD distance metric showed a statistically significant ($p < 0.037$ and $p < 0.001$, respectively) effect of training condition, no significant main effect was found for the GT metric. Pair-wise Bonferroni *t*-tests again confirmed the statistical similarity between the A+H and A+H(atten) conditions under all metrics.

Under the UD metric, there was a significant difference ($p < 0.006$) between the A and H conditions with the A condition leading to a 13% improvement in error over the H condition.

Under the VD metric, there was a significant difference ($p < 0.0003$) between the A and H conditions as well as between the A and A+H conditions ($p < 0.0003$). Figure 7(c) contains a boxplot summary of the data distributions under the VD metric. It shows a fairly clear trend towards lower error when haptic guidance is part of the training condition and lowest error when both audio and haptic guidance were used. In fact, the H training condition showed an 11% reduction in error over the A condition, while the A+H condition showed a 17% reduction in error over the A condition.

5 DISCUSSION

5.1 Effects of Attenuated Hearing

Recall that the reason for including the A+H(atten) condition was to test whether the presence of earplugs and headphones (not playing any masking sound) affected recall results since subjects wore these during the H test condition but not the A or A+H test conditions. If it did not affect the recall results, then there should not be a significant difference between the A+H and A+H(atten) training conditions. The results presented in Section 4 confirm this hypothesis as none of the *t*-tests run for any of the analyses found a statistically significant difference between the A+H and A+H(atten) conditions. This, therefore, validates the experimental design difference between the H and A/A+H test conditions.

5.2 Timing

In general, subjects improved their performance between early and later trials. Although performance improved significantly for all training conditions under the UD metric, no conditions showed significant improvement under the GT metric (although A improved marginally). Although the curves in Figure 5(b) suggest reasonable improvement, particularly for the H and A+H conditions, the early and final trial distributions overlap substantially under the GT metric (see Figure 6(b) and Figure 7(b)) which helps explain the lack of significance.

In terms of early trial performance, the combination of haptic and audio guidance appears to provide some advantage over audio or haptic guidance alone in terms of timing accuracy. The UD and GT metrics both showed some kind of trend in the first few trials towards lower error rates under the A+H training condition as compared to the A or H training conditions alone. From looking at the subject means of the first few trials in Figure 5(a), it appears that under the UD metric, the hybrid training conditions (A+H and A+H(atten)) incur less error than the A and H conditions. This was indeed the case; there was a 10% reduction in early trial error when subjects trained with the A+H condition versus the H training condition.

Under the GT metric, there is some amount of statistical support for this type of difference as well (see Section 4.2).[8] In this case, the A+H training condition led to 18% less error than both the A and H training conditions. This suggests that subjects were able to take advantage of both sources of information to learn more effectively. The fact that this difference occurred for early learning and for the global tempo metric is also interesting. Tempo is, by definition, a very different musical property than note duration or velocity as it does not apply to individual notes, but to entire sequences. From these results, it appears that combined audio-haptic guidance is particularly effective at quickly communicating this type of global property.

Although it may seem surprising at first that early trial results should show significant differences between training conditions, it is important to keep in mind that subjects had practiced the task rhythm twice before even the first test run (each trial contained two practice runs followed by a test run). Also, recall that early trial analyses used the mean of the first three trials' data as discussed in Section 4.2. It is therefore not unreasonable that performance could differ between training conditions during the early stages of learning.

The results of analysis of the final trials show that training with haptic guidance alone led to greater timing error at the end of the trial sequence than did the other training conditions. The results also show no statistical difference between the final error values when subjects were trained using audio alone versus when they were trained using a combination of audio and haptic guidance. Together, this suggests that the presence of audio information, either alone or in conjunction with haptic guidance, is responsible for the lower final error.

This finding is in line with related research on haptic guidance and vision. Both the Feygin et al. [6] and Liu et al. [17] studies found haptic guidance training to be inferior to visual training and visual with haptic guidance training in terms of recall performance. Although in some respects audition and vision are very similar (e.g. both provide exteroceptive feedback), there are clearly significant differences as well (e.g. information bandwidth), making this confirmation of earlier results an interesting finding.

Information bandwidth may also help to explain differences in timing error between training conditions with and without audio feedback. Although the task rhythms were designed with motor system bandwidth limits in mind (see Section 3.4), it may be that at the bandwidths used (1-5.33Hz), discrete temporal information is more easily processed by the auditory system. Another possible explanation is that there exist differences between short-term echoic (auditory) and haptic memory that favor auditory encoding and storage of timing information.

5.3 Velocity

One interesting aspect of the velocity results is the lack of change in performance over time under the A and H training conditions. Although Figure 5(c) suggests a trend of *worse* performance across trials for the audio training condition, the difference between first and last trial was not statistically significant (providing a good example of why one should never blindly trust plots). The lack of a significant difference between first and last trials under the A condition is also evident from the boxplot summaries in Figure 6(c) and Figure 7(c). They show a larger amount of variance for the A condition (versus the other training conditions) and that the early and final trial distributions for the A condition overlap substantially.

Early trial performance differences under the VD metric were significant, particularly between A and the other training conditions (see Figure 5(c)). Since the H and A+H conditions did not differ significantly while the A and A+H conditions did (and the A and H pairing differed marginally), it appears that the presence of haptic guidance information was primarily responsible for the better performance levels of the H and A+H conditions. Compared to the early trial error levels of the audio training condition (A), the haptic guidance only (H) condition showed 7% less error while the combined haptic guidance with audio (A+H) condition showed 10% less error. These results suggest that the presence of haptic guidance was particularly effective at reducing velocity error. The results also show a smaller amount of variance (see Figure 6(c)) in the conditions which included haptic guidance than in the audio (A) condition. This implies that the inclusion of haptic guidance in training can produce more consistent performance as well as more accurate performance in the early stages of learning.

---

[8] In fact, given the conservative nature of the Bonferroni correction, it is not unreasonable to pay some attention to effects with marginal significance.

The recall results for the final trials under the VD metric show perhaps the most striking differences of all of the experimental analyses. While there was no difference between H and A+H conditions, there were significant differences between the A training condition and the H and A+H training conditions. When subjects trained with haptic guidance only, the average error was reduced by 11% as compared to when they trained with audio only. The difference was even more pronounced for combined haptic and audio training where error was reduced by 17% when compared to audio training alone. The conclusions made above about the role of haptic guidance in determining recall performance apply here as well, as does the observation that subjects were able to make more effective use of both sources of sensory information when they were available concurrently. Figure 7(c) shows a sizable difference in variance between training conditions that use haptic guidance and those that don't (i.e. the A condition) which again indicates that the use of haptic guidance leads to more consistent performance than audio only training. Interestingly, the variance for the H condition is larger than the A+H condition, which suggests that not only does the combination of audio and haptic guidance lead to superior performance in terms of measured error, but it also leads to more consistent performance as well.

In general, the results of the velocity analyses might be explained by differences in sensory processing. In contrast to haptic guidance where direct proprioceptive information is provided, subjects had to translate perceived loudness into velocity information during the audio-only trials. This level of indirection presumably made audio-only training more difficult and may help explain the differences in velocity error between training conditions with and without haptics guidance.

6 CONCLUSIONS

This paper presented the results of an experiment designed to compare the effects of haptic and auditory guidance on learning in a musical performance context. Although the results indicate that audio guidance is important for learning timing information, the hybrid combination of audio and haptic guidance appears to provide additional benefit, particularly in the early stages of learning. The results also show that haptic guidance is substantially more effective at communicating velocity information than auditory guidance alone.

One major question that was not addressed by the current work is long-term retention. There is evidence that short-term (within a few minutes) and long-term (24 hours or more) retention can vary greatly, particularly when forms of augmented feedback have been used [11, 20, 2]. Dynamic or adaptive training schedules, where guidance or augmented feedback is provided less and less frequently with time, have been used to counter this effect [16] in other experimental contexts. This will be considered in a future set of experiments.

It would also be interesting to test how much of the observed advantage of the A+H condition is due to arousal effects. One could argue that the confluence of auditory and proprioceptive sensory information could have a stimulating effect on subjects which in turn could boost performance during multimodal training conditions. Because this possibility cannot be ruled out with the current data, a new set of experiments would need to be conducted. One possibility is to train subjects using the A+H condition and once their performance has stabilized, change to the H (or A) training condition. If the error level returns to the stable value reached under the A+H condition, it would suggest that multimodal arousal played a role in performance.

Given the encouraging findings presented here, the intersection of haptics and audio appears to be a fertile area for continued research. Although the implications of the current work are most clear for music pedagogy, it seems likely that these techniques could generalize to other application areas involving complex physical skill such as dance, sports, and remote medicine.

ACKNOWLEDGEMENTS

The author wishes to thank Sile O'Modhrain and the three anonymous reviewers for many insightful comments and suggestions during the development of this work.

REFERENCES

[1] H. Abdi. *Encyclopedia of Measurement and Statistics*, chapter Bonferroni and Sidak Corrections for Multiple Comparisons. Sage, Thousand Oaks, CA, 2007.

[2] T. Armstrong. Feedback and perceptual-motor skill learning: A review of information feedback and manual guidance. Technical Report 25, University of Michigan, Department of Psychology, Ann Arbor, MI, 1970.

[3] C. Black and D. Wright. Can observational practice facilitate error recognition and movement production? *Research Quarterly for Exercise and Sport*, 71(4):331–339, 2000.

[4] Y. Blandin, L. Lhuisset, and L. Proteau. Cognitive processes underlying observational learning of motor skills. *Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology*, 52(4):957–979, 1999.

[5] G. Burdea. *Force and Touch Feedback for Virtual Reality*. John Wiley and Sons, 1996.

[6] D. Feygin, M. Keehner, and R. Tendick. Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 40–47, 2002.

[7] P. Fitts. *Categories of Human Learning*, chapter Perceptual-Motor Skill Learning. New York: Academic Press, 1964.

[8] P. Fitts and M. Posner. *Human Performance*. Greenwood Press, 1967.

[9] R. Gillespie, M. O'Modhrain, P. Tang, C. Pham, and D. Zaretsky. The virtual teacher. In *ASME International Mechanical Engineering Conference and Exposition*, Anaheim, CA, November 1998.

[10] G. Grindlay. The impact of haptic guidance on musical motor learning. Master's thesis, Massachusetts Institute of Technology, September 2007. `http://www.media.mit.edu/hyperins/articles/Grindlay_MSthesis_final.pdf`.

[11] J. Hagman. Presentation and test-trial effects on acquisition and retention of distance and location. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9:334–345, 1983.

[12] D. Holding and A. Macrae. Guidance, restriction, and knowledge of results. *Ergonomics*, pages 289–296, 1964.

[13] J. Kelso. Planning and efferent components in the coding of movement. *Journal of Motor Behavior*, 9:33–47, 1977.

[14] T. Lee, M. White, and H. Carnahan. On the role of knowledge of results in motor learning: Exporing the guidance hypothesis. *Journal of Motor Behavior*, 22(2):191–208, 1990.

[15] J. Lieberman. Accelerated and improved motor learning and rehabilitation using kinesthetic feedback. Master's thesis, Massachusetts Institute of Technology, Auguest 2006.

[16] G. Lintern and D. Gopher. Adaptive training of perceptual motor skills: Issues, results, and future directions. *Journal of Man-Machine Studies*, 10:521–551, 1978.

[17] J. Liu, S. Cramer, and D. Reinkensmeyer. Learning to perform a new movement with robotic assistance: Comparison of haptic guidance and visual demonstration. *Journal of Neurorehabilitation*, 3(20), August 2006.

[18] K. Shimoga. Finger force and touch feedback issues in dextrous telemanipulation. In *Proceedings of NASA-CIRSSE International Conference on Intelligent Robotics Systems for Space Exploration*, 1992.

[19] K. Wang and G. Gasser. Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3):1251–1276, 1997.

[20] C. Winstein, P. Pohl, and R. Lewthwaite. Effects of physical guidance and knowledge of results on motor learning: Support for the guidance hypothesis. *Research Quarterly for Exercise and Sport*, 65:316–323, 1994.

[21] Y. Yokokohji, R. L. Hollis, T. Kanade, K. Henmi, and T. Yoshikawa. Toward machine mediated training of motor skills - skill transfer from human to human via virtual environment. In *5th IEEE International Workshop on Robot and Human Communication*, pages 32–37, 1996.