A Multimedia Handwriting Learning and Evaluation Tool

Mohammed Mansour, Mohamad Eid, and Abdulmotaleb El Saddik
Multimedia Communications Research Laboratory - MCRLab
School of Information Technology and Engineering - University of Ottawa
Ottawa, Ontario, K1N 6N5, Canada
{mansour, eid, abed} @ mcrlab.uottawa.ca

Abstract
In this paper, we present a multimedia system designed to facilitate learning of alphabetical handwriting of various languages by incorporating visual, auditory, and haptic feedback. An XML-based schema has been used to easily introduce new characters from other languages. The system supports two modes of operation: a learning mode and an evaluation mode. In the learning mode, a user can see, hear, and feel the letters and is guided (either partially or fully) to recreate their shapes. On the other hand, the evaluation mode tests the performance of students by comparing the trajectory/stroke and force information of the test character and a reference character. A pattern recognition algorithm, named Dynamic Time Warping (DTW), has been used to measure this similarity and give the student a pass/fail grade. We also verify the effectiveness of the DTW algorithm by trying several shapes and see how the difference value is changing.

1. Introduction
The development of handwriting skills is a fairly complex sensorimotor task that is best learned through practice and refinement. It needs attention, memory and cognition, and motor skills [1, 2]. A person can best learn how to carry out a sensorimotor skill task by seeing how an expert performs the task as well as by physically interacting with the trainer. Even though continuous visual feedback is an integral part of the learning process, transferring sensorimotor skills can be difficult to describe graphically or verbally and therefore the sense of touch will be a must for physical guidance. Examples of tasks that require physical guidance to improve learnability include learning handwriting for visually impaired people [3], medical procedures and rehabilitation training [4], and painting/sculpting techniques [5]. For instance, a teacher might need to hold the student’s hand to show him/her how to write a Japanese character or a letter from the Latin alphabet.

The haptic technology seems to have enormous potential for learning and training sensorimotor skills. Haptics refers to the emerging discipline that studies the communication of tactile and kinesthetic sensations between the human and the computer world [6]. Haptic sensations allow for multimedia applications, which utilize gesture recognition and force feedback. Furthermore, haptics – as a new media – plays a prominent role in making real-world objects physically palpable in a shared virtual environment and recognizing the physical properties of virtual objects such as shapes, textures, and stiffness. For instance, a system where haptic, visual, and audio information can be integrated to enhance user learning and whereby users can perform a specific learning exercise without the presence of a tutor can positively contribute to learning or training.

The present paper describes a multimedia system that combines visual, auditory, and haptic feedback to enrich the learnability of the students of handwriting alphabetical characters. The current version of the system supports five languages: Arabic, English, Chinese, Japanese, and French. The system not only enables students to learn handwriting skills, but also it allows a teacher to evaluate the performance of students by measuring how close the student-made character is to the reference character.

The remainder of this paper is as follows: in Section 2, we discuss related work and highlight the main distinguishing differences between our proposed systems and its competitors. Section 3 presents an overview of the system architecture and briefly discusses the two modes of operation (the learning and the evaluation modes). Section 4 presents implementation aspects of the proposed system and describes the experimental setup and analyzes the collected results. Finally, Section 6 summarizes the contents of the paper and provides an insight for possible future perspectives.

2. Related Work
Language characters are usually formed by sequential strokes that have to be performed in a certain order. Several researches have been conducted in the field of haptics to enhance or recover handwriting skills. For example, some researchers have demonstrated that haptic-visual systems are promising for learning [7, 2]. Other preliminary results using haptic-visual systems have shown their potential in helping kindergarten children to control handwriting movements [2]. As for handwriting recovery, a haptic
handwriting aid interface that stimulates re-learning the skill of writing after a stroke, or hand/eye coordination in writing, is presented in [4].

The literature includes several haptic-visual systems for teaching handwriting of various language alphabets such as for Japanese characters [8, 9], Chinese characters [10], and Latin alphabet letters [2]. All these systems are based on the concept of haptic playback. Haptic playback refers to the ability to display prerecorded haptic information (position and/or forces) using a haptic interface [11]. Because trajectory or position information is more useful for teaching handwriting skills, most of the systems have focused on displaying trajectories saved using a haptic device. Furthermore, some prototypes define only the full guidance playback mode. We believe that partial guidance could improve the learnability of students as it provides intermediate phase between completely dependent and completely independent performance.

The authors in [8] showed that there was some positive effect on learning one Japanese character from the kanji writing system when using force feedback. In [10], to guide users to the beginning of the stroke, five different models of assistance were distinguished. These models were defined according to different forces, stiffness, or damping. The initial results with six users were satisfactory and showed their improvement during the learning process.

The state of the art systems are limited by three main factors. First, they support the learning of the alphabets of only one language. Second, they do not incorporate the three major senses that are related to learning (vision, auditory, and haptics). Finally, many of these systems do not support a computational way to evaluate the student performance and progress. The proposed system provides a Graphical User Interface (GUI) where the user can load the alphabets of the Arabic, French, English, Japanese, or Spanish languages. Eventually, the user selects a character and learns its shape through graphic and/or haptic playback, in addition to the letter’s phonetics. The haptic interface includes full playback and constrained guidance with associated visual and auditory feedback also available. Visual feedback is given to the student by echoing the haptic stimulus onto the computer screen. In the testing mode, the DTW pattern recognition algorithm has been used to beautify model alphabets and computationally evaluate the student progress.

### 3. Multimedia Learning Tool

The haptic learning system comprises two components: the learning component and the evaluation component. The learning component enables students to learn the alphabets of a language using audio, visual, and haptic guidance. The details of this component are elaborated in our previous work [12]. On the other hand, the evaluation mode acts as a tool to computationally evaluate the student performance and progress of handwriting. In this section, we briefly describe the learning component and elaborate explain the evaluation component.

#### 3.1. Learning Mode

The learning component is divided functionally into four blocks: the language repository, the haptic player, the audio-visual player, and the graphical user interface. In the followings, we briefly describe each of these components. For more details about the learning part, the reader is referred to [13].

- **Language Repository**: The language repository contains, for each language, the symbol images, the pronunciation audio files, and the haptic stimuli for the alphabets. The repository contents are made accessible using an XML-based description that is structured into three main schemes: the metadata, the alphabets, and the haptic device. The metadata scheme describes the language name, audio files extension, images extension, the author’s information (such as the author's name, address, company, etc.), and intellectual property. The alphabet scheme contains a list of items, each representing a character in the language alphabets. The haptic device scheme lists the featuring characteristics of the haptic device used to capture the haptic stimuli for the alphabet characters. An instance of the XML file for the Japanese katakana alphabets is depicted in Figure 1.

```
<xml version="1.0" encoding="UTF-8" ?>
<Language Name="Japanese_Katakana" Flow="left">
  <MetaData>
    <AuthorName>Steve</AuthorName>
    <ContactInfo>steve@msn.com</ContactInfo>
    <Copywrite>LGPL</Copywrite>
    <CreationDate>06-05-2007</CreationDate>
  </MetaData>
  <HapticDevice>
    <DeviceName>Phantom</DeviceName>
    <Workspace>160X120X0</Workspace>
    <MaxForce>3.0</MaxForce>
    <DOF>3</DOF>
  </HapticDevice>
  <Alphabets>
    ....
    <Item Name="ka"/>
    <Item Name="ki"/>
    <Item Name="ku"/>
    <Item Name="ke"/>
    <Item Name="ko"/>
    <Item Name="sa"/>
    ...
  </Alphabets>
</Language>
```

Figure 1. A snapshot of XML language description

- **The Haptic Player**: There are three modes of operation in the haptic player: the no guidance mode, full guidance mode, and partial guidance mode. The full guidance playback mode can be used with real beginners to lead them through the writing of the alphabet where the system fully guides the student. In the partial guidance mode, the player applies partial forces
whenever the user movement diverges significantly from the desired path to bring him/her back to the correct path. As long as the student is following the correct stroke path, the device will not intervene with the subject’s hand movement.

- **The Audi-Visual Player:** When the student selects a particular character, the core application will automatically invoke the graphic player to plot the character symbol in a review area. Moreover, the user can hear the pronunciation of the character independently by clicking on a speaker icon. The student can review the character shape and sound as many times as he/she wants.

- **The Graphical User Interface:** The main GUI shows a drawing area, a keyboard panel that lists the language alphabets, and a control panel. The control panel defines the playback modalities (graphical, audio, or haptics) and the level of guidance (full, partial, and no guidance). Additionally, the GUI enables the student to load language packs (currently the application supports Arabic, English, French, Spanish, and Japanese Katakana) by loading the XML description file for the corresponding language. Afterward, the student can practice the handwriting of the alphabets and optionally plot what they have drawn and the desired shapes to get a feeling of how well they are performing. A snapshot of the learning component GUI is shown in Figure 2.

### 3.2. Evaluation Mode

The evaluation mode is developed to automate the process of testing the performance of the students after practicing with the tool. The major challenge here was to quantitatively measure the semantic distance between the user handwriting and the reference pattern. To compare the two time series, we have implemented a revised version of the Dynamic Time Warping (DTW) algorithm. The algorithm is adapted to increase the robustness against translation, rotation, and scaling since the student can write at any part of the drawing area and with varying size.

#### 3.2.1 Test GUI

Once the training is over, the student is now ready to perform an evaluation test. First, the student is asked to enter his name, select the language he/she wants to be tested for, and select the test mode (Figure 3). Then the student is forwarded to the test GUI (Figure 4). The test GUI presents the student with five randomly selected characters from the alphabets and asks him/her to write them sequentially. Also, the test GUI includes a timer that computes the test time. At the end of the test, the student will be presented with a test report that shows the characters and the grade for each (pass or fail).

#### 3.2.2 Dynamic Time Warping

In order to measure the similarity between the test and the reference handwritings of a character, several algorithms have been proposed. For instance, Euclidean...
distance is an efficient distance measurement that can be used. However, if one of the time series is shifted slightly along the time axis (which is very likely to happen in case of handwriting), then the Euclidean distance may consider them very different from each other (since it computes the sum of the squared distances from the \( n^{th} \) point in one series to the \( n^{th} \) point in the other).

The DTW algorithm was introduced in [14] to overcome this limitation by ignoring both global and local shifts in the time dimension. The DTW algorithm is used to compute the optimal alignment warp between two time series. It has been used for several applications including speech recognition, data mining, gesture recognition, robotics, manufacturing, and medicine [13].

For our particular application, we have shifted the startup position of the test pattern to that of the reference pattern in order to minimize the errors in computing the separation distance between the two patterns. The pseudo code of the algorithm is shown in Figure 5 and has been implemented as in [11].

Given two time series

\[
// x_i \text{ is a point in the test trajectory } \\
X = x_0, x_1, \ldots, x_N \\
// y_i \text{ is a point in the ref. trajectory } \\
Y = y_0, y_1, \ldots, y_T \\
\]

\[
D(x_i, y_i) = \sqrt{(\text{abs}(x_i) - \text{abs}(y_i))^2 + (\text{ord}(x_i) - \text{ord}(y_i))^2)} \\
\]

\[
\text{DTW}[0,0] = D(x_0, y_0) \\
\]

For \( i=1 \) to \( T-1 \)
\[
\text{DTW}[0,i] = D(x_0, y_i) + \text{DTW}[0,i-1] \\
\]

For \( i=1 \) to \( N-1 \)
\[
\text{DTW}[i,0] = D(x_i, y_0) + \text{DTW}[i-1,0] \\
\]

For \( i=1 \) to \( N-1 \)
\[
\text{For } j=1 \text{ to } T-1 \\
\text{DTW}[i,j] = D(x_i, y_j) + \min\{\text{DTW}[i,j-1], \text{DTW}[i-1,j], \text{DTW}[i-1,j-1]\} \\
\]

Return \( \text{DTW}[N-1,T-1] \)

4. Performance Evaluation

In this section, we present the performance of the DTW algorithm in the context of trajectory matching. The experimental setup of the application includes a Pentium 4 with 1 Gb Ram computer and the Phantom Omni haptic device, developed and marketed by SensAble Inc. The application is developed based on the .NET 2.0 Framework and is programmed using C#.

The grading scheme that we have used for this application is bipolar (pass/fail). The decision of whether the user passes writing a specific character depends on a threshold distance between the test character and the reference character. Unfortunately, there is no quantitative procedure to find these thresholds. However, they can be determined subjectively by measuring the distance for a large set of possible trials. Figures 6 and 7 demonstrate how the distance varies with deterioration in the character shape. Eventually, the thresholds for the three modes (easy, medium, hard) can be defined.

Therefore we used the ’trial and error’ method to estimate the threshold values for the three difficulty levels (easy, medium, difficult). After the user handwrites the test characters, the application uses the DTW algorithm to compute the separation distance between what the user writes and the reference character. Then this distance is compared to the easy threshold for the specific character if the user is performing the test in the easy mode. The same concept is applied in case of medium and difficult modes. Tables 1 and 2 show the threshold DTW distance that correspond to selected five characters in the three modes (easy, medium, and hard) for the Japanese and the Arabic languages, respectively.
5. Conclusion and Future Work

This paper presents a multimedia handwriting learning tool that incorporates audio, visual, and haptic modalities in the learning process. The system is capable of testing the student performance by quantitatively measure the similarities between the test characters and the reference characters. The preliminary results showed that incorporating the haptic modality increases the learning ability and fluency of handwriting alphabets in various languages. Furthermore, the performance of the DTW algorithm gave satisfactory results.

There are many possibilities as for the future work of this project. For instance, a collaborative environment where the trainer and trainee work closely together has potential benefits. Future studies will examine the effectiveness of close collaboration in the learning process. Additionally, we plan to add the playback capability to guide students about not only handwriting characters, but also complete words.

6. Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Alexander von Humboldt Foundation.

References


