Write Out Loud:
A Multimodal Approach to Learning Chinese Characters

Junru Ren <junruren@mit.edu>
Freya Tan <freya117@mit.edu>

April 11, 2025

Context:

The project was only ideated the night before Design Studio on April 8, 2025 as our original project idea ("City Moments") were deemed infeasible for the scope of this class project. We had to quickly brainstorm a new idea so as this "updated" proposal is due, we are still actively putting together important information such as software libraries.

Problem Statement and Basic Idea

Learning Chinese characters is challenging due to the complexity of multiple aspects including stroke order, pronunciation, and visual recognition. Traditional learning methods often fail to provide immediate multisensory feedback that could enhance memorization and calligraphic skills.

Figure 1: Copybook ("字帖"): a traditional paper-based learning tool

Our solution, "Write Out Loud," is a multimodal educational tool designed to aid learners of Chinese characters by integrating handwriting and speech modalities. As learners trace each character's strokes, they are required to simultaneously vocalize stroke names. The system provides real-time feedback to evaluate:

1. How accurately each stroke was traced?
2. Was stroke order followed?
3. How timely and accurately was the stroke name vocalized?

So that the learner masters the character through a combination of visual, auditory, and kinesthetic learning modalities.

Goal

The goal of this app is to improve **handwriting** of Chinese characters but not pronunciation (which can be a future feature addition). The deliberate vocalization of each stroke is meant to reinforce the learning of stroke order. As such, the app will not evaluate the pronunciation of each stroke name as long as the learner vocalizes the correct stroke name.

Concrete Example Scenario

A learner uses "Write Out Loud" on an iPad with a compatible writing stylus (e.g., Apple Pencile).

When the learner selects a character, the app displays, on the non-writing half of the screen [1] , that character overlaid with an animation showing how strokes are to be traced through.

On the writing half of the screen, the same character is displayed in less saturated grey color for the learner to trace through using their stylus.

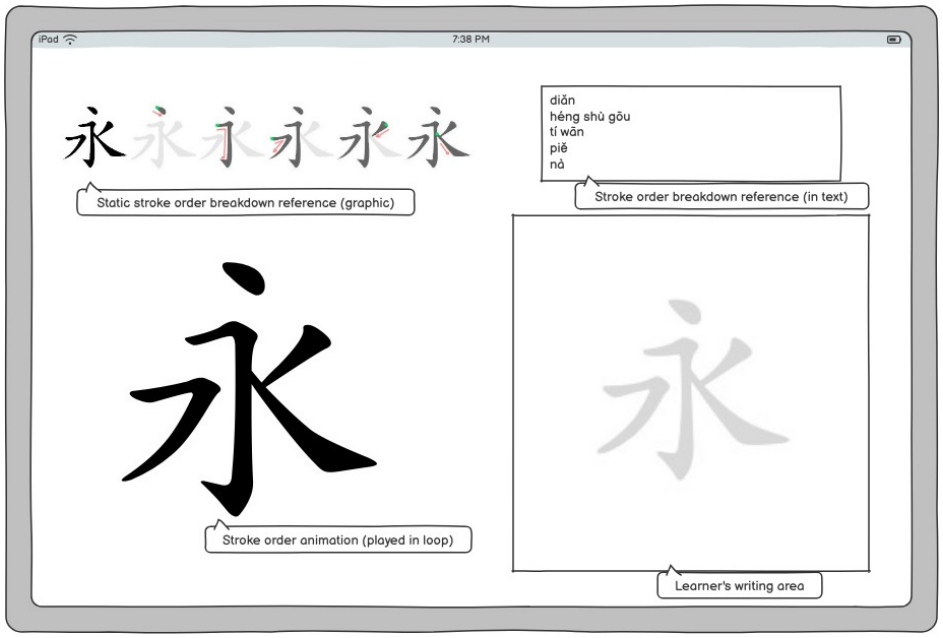

Figure 2: Mockup of app's writing GUI

The learner traces the first stroke, vocalizing its stroke name (e.g., "shù"). The app simultaneously captures the temporal and spatial handwriting input and timestamps the audio input. And so on. Once the character is detected to be fully traced. The app then assesses if the learner vocalized the correct stroke name concurrently with the stroke tracing. Feedback and a performance score are displayed after completing all strokes of the character, guiding learners to improve in subsequent attempts.

[1] We strive to use "the writing half" and "the non-writing half" to avoid being biased towards the right-handed or left-handed population. For for illustrative purpose, we assume that the left side of the screen is the non-writing half where necessary information is displayed.

Relevant Published Background

Multisensory learning, which engages multiple senses such as visual, auditory, kinesthetic, and tactile modalities, has been extensively studied for its benefits in enhancing educational outcomes. Studies indicate that this approach improves memory retention by engaging multiple sensory pathways, leading to stronger neural connections [2]. Engaging multiple senses caters to diverse learning styles, facilitating better understanding of complex concepts.

Planned Implementation and Basic Architecture

Our application will run on iPadOS, utilizing Apple Pencil for precise handwriting input and built-in microphone capabilities for capturing vocalizations.

System Components:

- **Pen Stroke Input Module**: Captures temporal and spatial data of stroke tracing.
- **Speech Input Module**: Records timestamped audio of stroke pronunciation.
- **Trace Analyzer**: Scores tracing accuracy based on predefined stroke patterns.
- **Speech Recognition Module**: Evaluates the correctness of stroke names vocalized by the learner.
- **Concurrency Analyzer**: Assesses the simultaneity of speech and handwriting inputs.

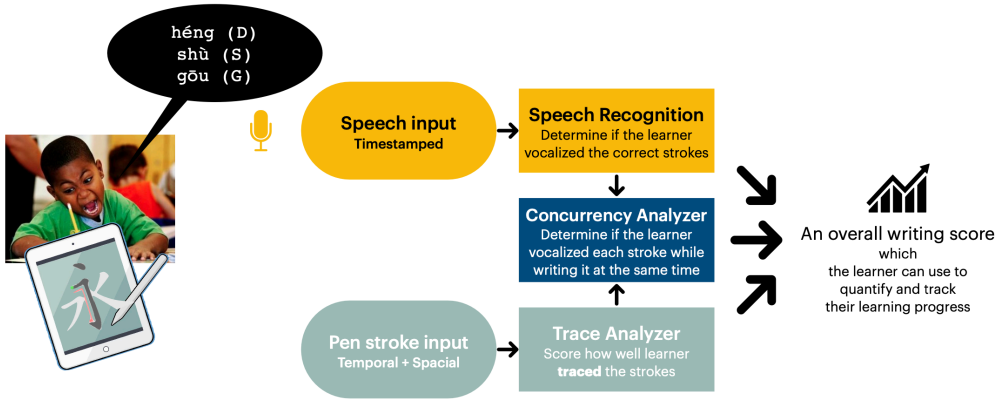

Figure 3: System Architecture Diagram of "Write Out Loud"

Required Resources

- iPad device with Apple Pencil
- Speech recognition software (built-in iOS Speech Framework)[3]
- Handwriting recognition and stroke analysis software (custom-built)[4]

[2] Okray, Z., Jacob, P.F., Stern, C. et al. Multisensory learning binds neurons into a cross-modal memory engram. Nature 617, 777–784 (2023). https://doi.org/10.1038/s41586-023-06013-8

[3] Apple's `SFSpeechRecognizer` library: https://developer.apple.com/documentation/speech

[4] Apple's `PencilKit` library: https://developer.apple.com/documentation/pencilkit

- A small hand-curated[5] Chinese character dataset that includes the following for each character:
  - Visual representation of the entire character
  - Visual representation of the entire character in grey for trace over
  - Animation of the stroke order
  - Stroke order (graphic and textual)
  - Broken down stroke coordinates within the writing pane

Experimental Evaluation

We will first evaluate the system using a baseline character, such as "口," which involves simple, well-defined strokes. Participants will complete tasks of writing and vocalizing characters, with performance metrics captured across both modalities. Metrics on the learner include accuracy of stroke tracing, correctness of vocalizations, and the timing alignment between writing and speaking. For the app itself, we will measure learner satisfaction and subjective ease of use. We also plan to invite a small group of people who doesn't know Chinese to try the fined app and then test whether they could write the same set of characters on paper later after using the app.

Evaluation Metrics:

- Accuracy of stroke tracing (shape, order, and orientation)
- Correctness and clarity of vocalized stroke names
- Temporal alignment between writing and vocalizing
- Feedback delivery effectiveness (e.g., text/audio corrections)
- Usability feedback through post-task surveys and interviews

Staged Development Plan

Version	Deadline	Description
V-0	April 13	Set up the iPadOS development environment and write a dummy iPadOS app via XCode.
V-1	April 27	Implement basic functionality to capture handwriting and vocalization for a simple set of characters, providing accuracy feedback.
V-2	May 2	Expand character set and refine feedback mechanisms.
V-3	May 5	Enhance user interface and integrate adaptive learning features based on user performance.

Table 1: Staged Development Plan

[5]In a class project scope, we decided to just manually create data for few character as our focus in on the interaction of learning a character but not production-ready coverage.