Proceedings of the 2nd International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/27/20231778

Revitalizing Cantonese Proficiency: An Interactive ASR-Driven Approach

Yuanhang Xu[1,a,*]

[1]*Beijing Normal University, Zhuhai, No. 18, Jinfeng Road, Tangjiawan, Xiangzhou District, Zhuhai City, Guangdong Province, China*
a. 1877109236@qq.com
**corresponding author*

Abstract: In recent years, an increasing number of children in the Guangdong region found themselves unfamiliar with Cantonese which is their native language. This trend poses a significant threat to the preservation of the rich cultural heritage and traditions associated with the Cantonese language, meanwhile Cantonese is very different from Mandarin so it is very difficult to learn. Recognizing this challenge, the "Assistant for Cantonese Learning" app has been developed to bridge this gap. The app aids in using an ASR model to provide exercises tailored to a student's proficiency, allowing the user's mastery of Cantonese to be assessed based on the comparison between the translated text and the original text provided by the system, as well as the correctness of the user's pronunciation and intonation. Using OpenAI's advanced Whisper ASR model, the app offers an interactive and dynamic learning experience, ensuring that anyone who wants to learn Cantonese can connect with the linguistic roots efficiently and effectively.

Keywords: Cantonese, language study, Whisper, OpenAI

1. Introduction

The software is designed to guide learners in mastering Cantonese progressively based on their proficiency. First of all, when users embark on their Cantonese study within the software, they are presented with several simple, everyday transcripts in their native language. Users are prompted to read these texts aloud, after which the software captures the audio and processes it using OpenAI's Whisper ASR model. Recent studies, such as [1], have highlighted the inherent advantages of reading tutors in language education. Based on the resulting transcription and its comparison with the original text, the software gauges the learner's proficiency in Cantonese. Learners can be categorized into different proficiency levels, for instance, beginners, intermediate, advanced, and native speakers. As learners progress, the complexity of the manuscripts increases. intermediates are presented with lengthier and more specialized texts or some simple questions like "你觉得秋天点呀？" Which means " How do you like autumn ？", while advanced learners encounter unique Cantonese vocabulary Similarly, this stage will provide a Cantonese audio, after learners listen to this audio, learners need to answer questions about this audio and there is a time limit. In the native speaker stage, manuscripts and questions will be based on Cantonese news or famous Cantonese speeches which are usually more professional and official, there are also deeper questions. Simultaneously, the

DOI: 10.54254/2753-7064/27/20231778

software facilitates repeated practice for words and sentences identified as challenging, enabling learners to rectify language intonation problems. Finally, even though not all learners can reach the level of native speakers, at least the software can help them become qualified Cantonese speakers. Workflow of the application is shown in Figure 1.

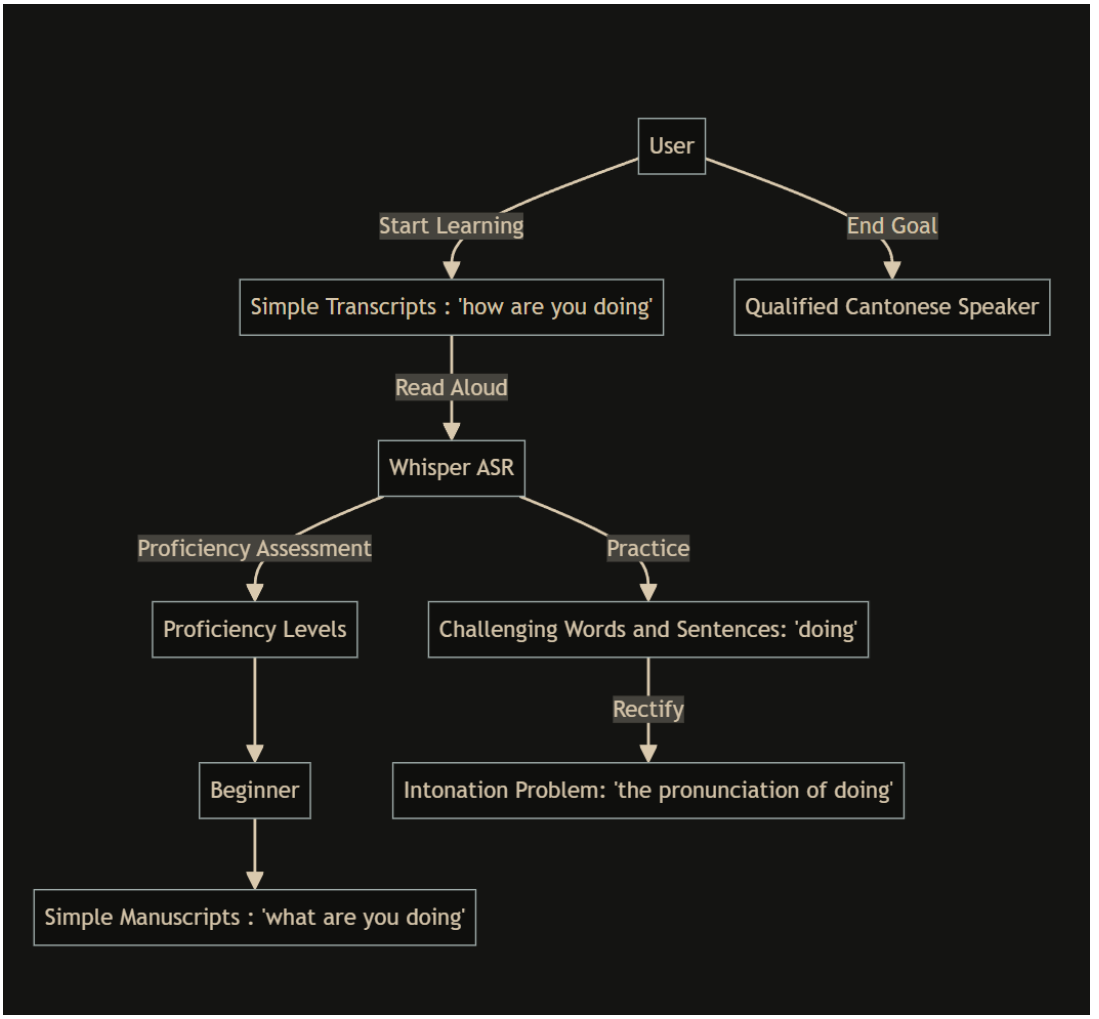

Figure 1: Workflow of the application

2. Applying ASR to Cantonese Learning

When students engage with the app by speaking in Cantonese, Whisper captures the learner's verbal inputs and swiftly transcribes them into the transcript. This real-time transcription serves multiple functions:

Immediate Feedback: Much like in Project LISTEN [2], Students can instantly see their spoken words translated into transcripts. This allows learners to identify any discrepancies between what they intended to say and what they actually verbalized, fostering the ability to think in Cantonese and self-correction.

Proficiency Assessment: By comparing the transcribed text to the pre-set text, the app assesses the accuracy of the learner's pronunciation and sentence structure. Based on the comparison, it provides them with a proficiency level that classifies them as beginners, intermediate, advanced, or native speakers.

DOI: 10.54254/2753-7064/27/20231778

Customized Exercises: "As noted in a previous study [3], previous research has demonstrated that practicing the reading of words in running text is more effective for transferring that skill to new text, as compared to practicing words in isolation. Building upon this principle, the app using the transcriptions, the software identifies common errors or patterns in a learner's verbal inputs. It then recommends personalized exercises to address these specific areas of improvement. For instance, if a learner consistently mispronounces a certain tone like “做(zo3)” which means doing, the system provides exercises focusing exclusively on that tone such as, In the next question or transcriptions, the learner will be asked to say “做(zo3)” again.

Vocabulary Expansion: By analyzing the transcript, the software can suggest new vocabulary or phrases related to the learner’s response to the question provided by the system. For instance, learners say “好远嘅地方” which means “It's so far away” and the app will suggest learner: you also can say “冇雷公咁远”.

Conversation Practice: Taking a page from Project LISTEN's book [2], For more advanced learners, the application might simulate conversations. After a learner speaks, the software, using Whisper's transcriptions and Text-to-Speech technology, crafts appropriate responses, pushing learners to think on their feet and engage in real-time Cantonese speaking.

Using ASR revolutionizes the process of learning Cantonese, making Cantonese learning more interactive, personalized, accurate and effective. The workflow of the ASR is shown in Figure 2

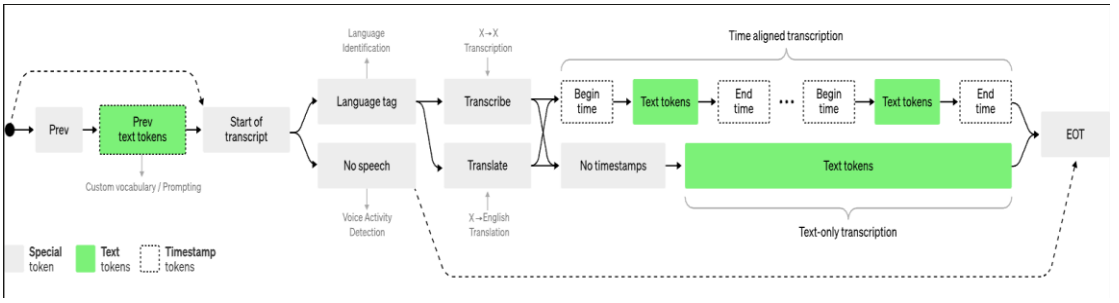

Figure 2: The workflow of the ASR

3. Models

Overall, the "Assistant for Cantonese Learning" encompasses four distinct proficiency levels. The higher the level, the greater the fluency in Cantonese.

The software gauges a learner's proficiency by contrasting the provided Cantonese manuscript and question with the transcription obtained from Whisper after processing the learner's verbal inputs. This comparison hinges on several key factors:

3.1. Vocabulary Accuracy

It checks for the usage of appropriate words in the learner's response against the original manuscript. For instance, if the software expects "早上好" (Good morning) but receives "晚上好" (Good evening), it indicates a vocabulary misalignment. also, Learners may not realize that they are using the wrong word, and they may bring word habits from their native language into Cantonese, when in fact native speakers do not speak the same way, and this phenomenon will be regarded as vocabulary inaccurate, For instance, in Mandarin the phrase ‘一边吃饭，一边看电视’(eat while watching TV), In Mandarin, you can use "一边...一边..." ("side... One side..." / while) to represent simultaneous actions but in Cantonese a similar sentence format is not used; instead, the phrase ‘食饭嘅时候，睇电

DOI: 10.54254/2753-7064/27/20231778

视。’. The system will help the learner to correct vocabulary inaccurate by offering the correct statement formatting and appropriate words.

3.2. Pronunciation Analysis

Due to Whisper's high tolerance for different accents within the same language and its accurate recognition rate, the system can accommodate variations in learners' accents. Consequently, when learners truly mispronounce words, the system can effectively pinpoint the errors and assist learners in making corrections. For instance, if a learner says "狗" (dog) but Whisper transcribes it as "高" (high), (the only difference between these two words in Cantonese is intonation) more like the pronunciation of bare and the pronunciation of bear in English.

3.3. Fluidity and Structure

After the learner's audio is transcribed by Whisper, the app examines the overall fluidity and grammatical structure. Smooth, coherent sentences suggest a higher proficiency.

3.4. Tone and Intonation

Since Cantonese is a tonal language, the correct tone is prime importance. The software analyzes if the learner's tonal variations align with the expected tones of words or phrases.

3.5. Results

Taking the above factors into account, if a learner consistently exhibits accurate vocabulary usage, correct pronunciation, fluent sentence structures, and appropriate tones, their proficiency score might be at 0.9,

The system identifies the learner as a native speaker and system will offer the manuscripts and questions based on Cantonese news or famous Cantonese speeches to better determine the level of learners also aligning with their demonstrated proficiency.

Furthermore, the application dynamically calibrates the difficulty of the manuscripts or questions learners are exposed to, ensuring it matches their proficiency, thereby providing an optimal learning curve.

4. Educational Data Mining

Collect learner background data: Upon registration, the system gathers details about the learners, including their region, native language, current proficiency in Cantonese, and specific learning objectives. Based on this data, and informed by the past performances of learners with similar backgrounds, the system personalizes manuscript sets and question sets to align with the learner's proficiency.

Mine data during Cantonese practice sessions: As learners engage with the content, especially during practice sessions, the system collates various metrics indicative of proficiency. These include Vocabulary Accuracy, Pronunciation Analysis, Fluidity and Structure, Tone and Intonation, the time taken to respond, and the number of exercises completed.

Randomize content decisions: Leveraging insights from the mined data, the system gauges learners' mastery over specific Cantonese concepts. Based on this evaluation, it determines the suitable difficulty level for their next practice session. For instance, if a learner opts to practice Cantonese greetings and is presented with content at level beginner, but their performance suggests they're an intermediates, the subsequent session will challenge them with content that truly matches their proficiency. suitable difficulty level for their next practice session.

DOI: 10.54254/2753-7064/27/20231778

Adjust content difficulty in real-time: “The resulting data was mined to pinpoint the specific common syntactic and lexical features of text that children scored best and worst on. These features predict their fluency and comprehension test scores and gains better than previous models.” from[4]. Based on the experience gained in [4], The system continually assesses the aptness of the difficulty levels assigned to various exercises. If, for instance, a manuscript or a question is categorized as advanced, but the majority of learners navigate it effortlessly, the system infers the exercise is easier than initially thought. Consequently, it might reclassify this exercise to a more fitting level, like a beginner.

By continually adjusting and personalizing content based on real-time learner feedback and performance, the system ensures an optimal learning journey for individuals endeavoring to master Cantonese.

5. Embedded experiments and Evaluation

To guarantee the usefulness of the "Assistant for Cantonese learning" app, it is crucial to assess its efficacy despite the already reliable whisper model. In order to avoid inaccurate results from incomplete data, it is essential to construct lab tests that evaluate the model meticulously. Various measures have been employed to enhance the trustworthiness of the model. To make the model more widely applicable, a diverse collection of Cantonese audio and learners from different ages, linguistic backgrounds, and educational experiences was gathered. The dataset was divided into two parts - training set and testing set. With the training set, the Cantonese learning model was trained. The performance of the model was then evaluated with the test set.

One of the methods used to assess the model's performance is detailed below:

Cross-validation: In this study, Cross-validation was applied to assess the impact of different student models on the decision quality of Assistant for Cantonese learning, as described in [5]. Multiple folds are used to split the dataset during cross-validation. The model undergoes training and testing on each fold repeatedly. This allows for the assessment of its generalization ability and stability as well as calculation of its average performance on different subsets through various iterations.

Error analysis: paid to the model's performance on varying levels of difficulty and Cantonese problems, analyzing its prediction results on the test set. By conducting an error analysis, potential areas of weakness in the model were identified, and utilize this information to enhance its performance further.

Human evaluation: Bring in some professionals within the Cantonese language or education field to give their evaluation of the model and its performance. Obtaining insight from experts on the effectiveness of the model can be highly valuable.

Benchmark Model Comparison: The potential for the Cantonese language learning model to be enhanced can be evaluated by pitting it against other benchmark models in the domain.

By assessing the outcomes of these assessment techniques in conjunction, it can be determined if the model can accurately measure the intensity of hardship experienced by Cantonese learners and whether there are issues related to overfitting. Once the model aligns with the prospects and if there are no prominent overfitting issues, confidence can be placed in the model's ability to deliver the difficulty level of adjacent learning exercises. Alternatively, by making further adjustments to the parameters, the model's efficiency can be heightened.

6. Conclusion

A ray of hope arises as the younger generation loses touch with the Cantonese dialect, with the advent of the "Assistant for Cantonese Learning" app. Through the use of OpenAI's Whisper ASR model,

DOI: 10.54254/2753-7064/27/20231778

the software provides learners with a unique, vibrant and interactive educational experience, customized to individual proficiencies. By means of instantaneous feedback, skill assessments and individualized exercises, the app ensures learners not only master Cantonese's intricate structure, but also develop a profound connection with its cultural heritage. Looking to the future, there are high aspirations for using advanced technology to preserve and promote linguistic diversity and cultural heritage. Through continuous data mining and adaptive content adjustments, every single user's journey can be optimized for their one-of-a-kind needs, making the learning curve even more powerful. It is believed that these technological interventions will be the key to maintaining the rich tapestry of Cantonese alive and thriving for generations.

References

[1] *[JECR 2013] Mostow, J., Nelson, J., & Beck, J. E. (2013). Computer-Guided Oral Reading versus Independent Practice: Comparison of Sustained Silent Reading to an Automated Reading Tutor that Listens. Journal of Educational Computing Research, 49(2): 249-276.*

[2] *Mostow, J. (1971) Potential Applications of Artificial Intelligence to Education*

[3] *[[SSSR 2012] Mostow, J., Nelson, J., Kantorzyk, M., Gates, D., & Valeri, J. (2012, July 11-14). How does the amount of context in which words are practiced affect fluency growth? Experimental results. Talk presented at the Nineteenth Annual Meeting of the Society for the Scientific Study of Reading, Montreal, Canada.*

[4] *[FLAIRS 2012 prosody] Sitaram, S., & Mostow, J. (2012, May 23-25). Mining Data from Project LISTEN's Reading Tutor to Analyze Development of Children's Oral Reading Prosody [Best Paper Award]. In Proceedings of the 25th Florida Artificial Intelligence Research Society Conference (FLAIRS-25), 478-483. Marco Island, Florida.*

[5] *[[BKT20y 2014] Xu, Y. & Mostow, J. (2014, July 4). A Unified 5-Dimensional Framework for Student Models. In Proceedings of the EDM2014 Workshop on Approaching Twenty Years of Knowledge Tracing: Lessons Learned, Open Challenges, and Promising Developments, 122-129. Institute of Education, London, UK.*