Tone Language Teaching Assitant Model

Hanwen Yang

Department of Information and Computer Science, Jinan University, Jinan, China

hxy222@student.beam.ac.uk

Abstract. The study for the tonal languages has always been an important issue in the field of the linguistics. How to improve the efficiency of language learning and provide learners with appropriate feedback in time is worthy of attention. In this paper, the research aims to design an automatic system or application to help users learn the tones they pronounce in order to improve their learning efficiency relative to manual teaching. The user will first select the proficiency level, and the system will provide corresponding sentences for learners to read. Learners will have three buttons to choose from: read the sentence, ask for a prompt, or just skip. If the learner mispronounces the sentence, the app will automatically mark the word the learner mispronounces in red and provide the correct pronunciation. The general framework and the corresponding example results were also provided in this study. Such a system has a potential to be applied in the real life.

Keywords: Language model, tonal language study, computer science.

1. Introduction

In the world, there are many different kinds of languages and some of them are called tone languages, which means that for the words in this kind of languages, different tones represent different meanings and also to convey grammatical distinctions. It has also been viewed by some researchers as a language in which pitch is used as an apart of speech changing the meaning of a word. In East Asia, many languages such as Chinese, Thai, Vietnamese and Lao are all tone languages. When using this kind of languages to communicate with others, it is always significant to use the correct tones, or it may lead to the wrong meaning that she want to convey. However, for a beginner who is studying tonal languages, it is not always easy to do so. Normally, it is more complicated to learn a tone language than a intonation language and people need a teacher to guide them to move on.

In this paper, this study aims to figure out an automatic system or an app, to help the user learn the tone they pronounce in order to improve their learning efficiency compared to the artificial teaching. The user will first choose the proficiency and the system will then provide corresponding sentences for learners to read. Learners will have three buttons to choose, read the sentence, ask for the hint or just skip. If learners read the sentence incorrectly, the app will automatically mark the word that learner mispronounced in red color and provide the correct pronunciation. On the other hand, if learners can read the sentence correctly, the app will move on to the next sentence. When learners finish the practice, the app will provide the performance feedback, tips and store the record to learners' linked account. As a result, next time learners use the app, they can choose whether to review the wrong word that they previously mispronounced. The whole system is divided into 4 parts, namely models, content generation, evaluation and embedded experiments. The model part contains 2 types of models that the system would need, and also how they are used by the system. The content generation part 2 concrete examples of content that this study generates automatically, and how the system uses them. The evaluation part contains at least 2 criteria for the researchers to judge the system and also for the users to estimate whether this system really helps. The embedded part contains 2 specific experiments that the system could perform, as well as what the results would tell me.

The rest of this paper is organized as follows: Section 2 describe the models considered in this study. The content generation is included in Section 3. Section 4 describes the evaluation and embedded experiments of the designed app. Finally, the conclusion is provided in the Section 5.

2. Methodology

2.1. The architecture of the model

Two models will be applied to the app. The first one is knowledge tracing. Knowledge tracing will update the tutor's estimate of a learner's mastery of specific skills. Since its prediction is based on the continuing observation of previous performance, it can predict how well the learner performs next and hence pick an appropriate problem [1]. In the app, the implementation of this model will allow the app to evaluate the learner's ability based on a series of answers and then choose corresponding sentences which match the learner's level for practice.

Here is a concrete example shown in Fig. 1 and Fig. 2. Bill is curious about Chinese and just started to learn it two weeks ago. When he uses the app, he chooses his language proficiency as "starter". The app hence provides a very simple sentence "where is her mom going? (in Chinese)" Since lots of new learners may mispronounce the "mom" as "horse" in Chinese, this is a perfect example to evaluate them. If Bill answers it incorrectly, the app will mark the wrong word with red color, provide a correct pronunciation and then choose a simpler sentence. Otherwise, if Bill can answer it correctly, the app will move on and provide a more complex sentence like the mom rides a horse to the market. In this sentence, "mom" and "horse" (in Chinese) are presented together and hence pose a larger challenge for the learner. Through a series of sentences, knowledge tracing model will have a better understanding of the learner's proficiency based on the observed performance and hence provide more complex or simpler sentences according to its prediction.

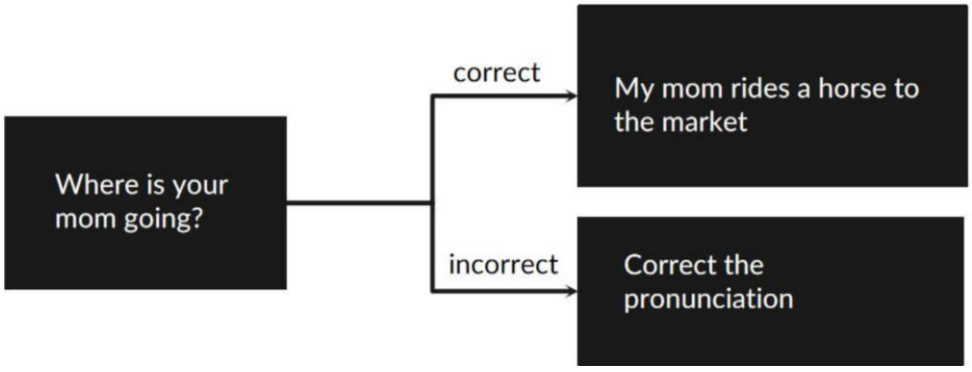


Figure 1. The example of the input-1

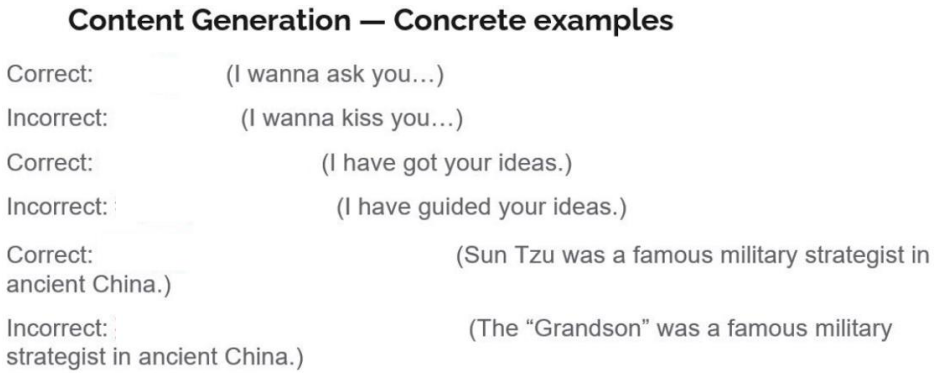


Figure 2. The example of the input-2

Another model that will be applied is learning decomposition. Learning decomposition can evaluate the relative impact of different types of practice based on learner's performance data [1]. As a case study explored students' performance with different types of practice by applying learning decomposition, it figured out that though rereading and massed practice tended to be less efficient, low proficiency students could benefit more than high proficiency students [2]. In the app, it will have a special study mode that allows learners to choose like learning the same word with different tones or learning different words with the same tone and so on. In this case, this app will treat their response time as their performance and assign 5s to practice like missing the word, asking for hints or

pronouncing incorrectly. By analyzing data from different learners with a lot of practice, learning decomposition can find out which practice is more efficient and what groups of people can benefit more from a certain type of practice.

For instance, Mary has learned Chinese for five years and she can now speak Chinese almost as fluently as her colleague Tom who is a native speaker. As a result, she decides to challenge herself by choosing "learning different words with the same tone", the app will then provide a tongue twister that is even challenging for native speakers because it is full of words with the same or similar tone. After Mary practiced for several days, the app will analyze her performance data alongside with other learners of various proficiency, and have a clear picture of whether these learners benefit more from this type of practice. As a result, the app can provide more personalized practice to learners by knowing which type of practice benefits more to this group of learners.

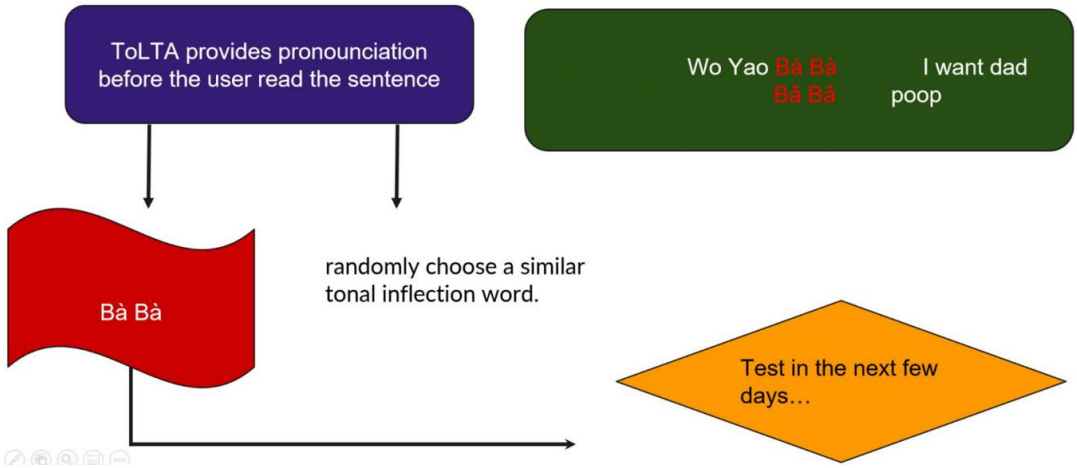


Figure 3. The example of the evaluation

2.2. Content Generation

The generation system is an automated generation system which can generate sentences for a user to practice either by randomly choosing some specific initial words or according to a certain word which the user has made a mistake in pronouncing it previously. When the user makes some mistakes in reading the tone of a certain word, the word will be recorded into a word list where the words are supposed to be reviewed. When there are no words left to be reviewed, the system uses the N-Gram model to generate some contents for practice by randomly choosing a certain initial word from the database. The database collects the words which have tones that may be easily confused. For example, when the user does not have any words left to be reviewed, the system may choose a word "horse (in Chinese)" from the database because the word "horse (in Chinese)", which means "horse", is often confused with the word "mom", which means "mother". Then, the system will generate some candidate sentences including the word "horse", like "He went on a journey on horseback", "The horse is white", and "Horses like to eat grass", for the user to practice.

When there are words left to be reviewed, the content generation system will generate sentences which include these words. Here this study list some concrete examples for this mechanism. For the first example, the user read the word "ask (in Chinese)" incorrectly to a "kiss (in Chinese)". In the previous practice, the user read the sentence "This person wanna ask she something (in Chinese)" to "This person kiss she (in Chinese)". This wrong pronunciation may be really embarrassing in real life situations. Then the system will fill the word "ask" into the word list, and when the user reviews this word later, the system will generate sentences including the word "ask", like "She go ask him" and "That's an interesting question". For the second example, the user read the word "know" incorrectly to a "guidance", and the user read incorrectly the sentence "This person got her ideas." to the sentence "This person has guided her ideas", which will be confusing. Then the system will fill the word "know" into the word list, and when the user reviews this word later, the system will generate sentences including the word "know". For the third example, the user read the word "grandson" incorrectly to

"grandson", although they seem to be the same, they have different meanings in these two pronunciations. The first "grandson" has the meaning of a famous person in ancient China, while the second "grandson" means the grandson. Then the system will fill the word "grandson" into the word list, and when the user reviews this word later, the system will generate sentences including the word "grandson".

3. Evaluation

3.1. Evaluation Process

There are many ways to evaluate whether an app works well and really helps the user, but this study just picked 2 of them that this study thought this study the most effective. The first one was the accuracy, and the second one was the reading speed. As the main factor that determines whether a person can communicate with a native foreigner correctly, without accuracy people probably don't understand what this study are actually saying and as soon as they get the wrong meaning of the words, it will be pretty hard to continue the communication. Also, the accuracy is a direct manifestation of the effect throughout the learning phase. Thus far, this study could firstly get a rough evaluation of it. When this study considered the further aspect of this app, reading speed might be an important part. This study all knew that when talking to other people in the daily life, if she couldn't express her core meaning at the first time, people would get impatient, especially in a foreign country. So as far as this study this study concerned, reading speed should be tested during the evaluation part. The researchers needed to collect the statistics two times. One from the beginning, when the learner first registered and used the app, and the other one should come from when the learner finished a stage of learning. And by this way this study could effectively test out to which degree the app helps the user.

As for how someone else could evaluate it, this study thought maybe this study could add a rating system in app. It could record the voice of the learner two times, the same as it did in the researchers' part when they tried to evaluate the reading speed. Of course, if the learner wanted to test their learning results half way, it is OK. So next, the app would integrate the statistics from different stage of the learning and give a feedback of the comprehensive evaluation to the user, and this feedback, this study thought, should contain the following aspects, the increasing rate of the accuracy, the reaction time the learner used to speak out a difficult word and which words this study still need to be improved. And next, the app would give the user some plans, which focused on the learning of the following stage. This would do further analysis to the data the app collected during the first stage, and it would offer a few options for the user to choose, this study meant, they could choose which aspect they wanted to improve first, the reading speed or something else. And as soon as the users made their decisions, the app would draw up a long-term plan, and it would help the user improve in the shortest time. This way, this study could see that throughout the process above, it was pretty easy for the learner to evaluate this app and get a better command of their learning effects. The data analysis may considered the use of the machine learning algorithms e.g. decision tree [3-5], random forest [6-8] and neural network [9, 10].

3.2. Embedded Experiment

What Embedded Experiment means is that relevant variables are systematically manipulated in real learning contexts, carrying strong benefits for making causal inferences. There are two Embedded Experiments this study can do based on the system using the method Mining Data from Randomized Within-Subject Experiments.

Research Question: Will previewing the pronounce of the word before the user reads be helpful?

Purpose: To find out what method is more practical and helpful for the user to learn tonal language.

Process: At the beginning, the teaching Assistant will randomly find a sentence from the database, and randomly pick a similar tonal inflection word in the sentence, pronounce it before the user reads. As an example shown in Fig. 3, in sentence "(This child wants dad), "Dad (in Chinese)" has a similar

tonal inflection with “poop (in Chinese)”. People who learn Chinese very likely to pronounce it wrong from “Dad (in Chinese)” to “poop (in Chinese)”, which will make the sentence inappropriate. After showing and previewing several sentences, this study will do a quiz on how good this method works for the users.

Data: Collecting all the performance the user does on the test. Differentiating them in percentage grade, then comparing the proportion in each percentage range.

Research Question: What type of correction is the most effective?

Purpose: To find out which type of reminding or correction works the best.

Process: The teaching Assistant will randomly find a sentence from the database, and show it on the screen. As the user reads, the system records the words and pronunciation the user said and randomly picks a type of correction. Some corrections this study used are, first, playing the correct pronunciation directly to the user as when the system detected a wrong pronunciation. Second, the system would still record from the user, but after the user is done speaking, this study will print the sentence that the user said, and mark the word this study detected wrong on pronunciation. Third way is to play the correct reading records after the user is done speaking, this study will let the user make corrections by themselves. After practicing, this study will do quizzes to find out which method is better.

Data: Performance on the quizzes, category them in different methods. Analyzing the distribution, this study will find out the best type of correction.

4. Conclusion

In this paper, this study aims to figure out an automatic system or an app, to help the user learn the tone they pronounce in order to improve their learning efficiency compared to the artificial teaching. The user will first choose the proficiency and the system will then provide corresponding sentences for learners to read. The experiments demonstrated the effectiveness of the proposed system. In the future, more real experiments related to coding or system development should be considered to further improve the performance of the system.

References

[1] J. Mostow, and X. Zhang. "Analytic Comparison of Three Methods to Evaluate Tutorial Behaviors." Educational Data Mining 2008. 2008.

[2] J. E. Beck, & J. Mostow, "Mining Data from Randomized Within-Subject Experiments in an Automated Reading Tutor (poster in session 34.080, "Logging Students' Learning in Complex Domains: Empirical Considerations and Technological Solutions"). American Educational Research Association 2005 Annual Meeting: Demography and Democracy in the Era of Accountability, Montreal, Canada, 2005.

[3] A. J. Myles, et al. "An introduction to decision tree modeling." Journal of Chemometrics: A Journal of the Chemometrics Society 18.6, 2004, 275-285.

[4] Y. Song, and L. U. Ying. "Decision tree methods: applications for classification and prediction." Shanghai archives of psychiatry 27.2, 2015, 130.

[5] S. R. Safavian, and D. Landgrebe. "A survey of decision tree classifier methodology." IEEE transactions on systems, man, and cybernetics 21.3, 1991, 660-674.

[6] M. Belgiu, and D. Lucian. "Random forest in remote sensing: A review of applications and future directions." ISPRS journal of photogrammetry and remote sensing 114, 2016, 24-31.

[7] Y. Qi, "Random forest for bioinformatics." Ensemble machine learning: Methods and applications. Boston, MA: Springer US, 2012. 307-323.

[8] T. M. Oshiro, S. P. Pedro, and A. B. José. "How many trees in a random forest?" Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8. Springer Berlin Heidelberg, 2012.

[9] C. M. Bishop. "Neural networks and their applications." Review of scientific instruments 65.6, 1994, 1803-1832.

[10] H. Adeli. "Neural networks in civil engineering: 1989–2000." Computer-Aided Civil and Infrastructure Engineering 16.2, 2001, 126-142.