



# Oral Practice Language Learning Companion Implementation on Voice Assistant Platform

Hsiang-Jen Chung<sup>(✉)</sup> 

Graduate Institute of Network Learning Technology, National Central University,  
Taoyuan, Taiwan  
ryanchung@ncu.edu.tw

**Abstract.** Oral practice for language learning usually needs to find partners or teachers to accompany, no matter how patient they are, it's not easy to stay for a long-time conversation practice. This study tries to implement an energetic oral practice language learning companion on voice assistant platform, it can always be with you to practice English conversation. It uses natural language processing and voice recognition technology to tell if you are repeating the same sentence, help students can practice conversation by themselves.

**Keywords:** Language learning · Learning companion · Virtual companion · Voice assistant · Artificial intelligence · Education technology

## 1 Introduction

### 1.1 Motivation

The best way to learn language is practicing with real people and makes you get into the whole foreign environment. But sometimes we do not have chances or budgets to do so. No matter gets a foreign friend or foreign place, it costs. So, if there comes a learning companion can be with learner to practice English every day, let learner know if he is doing right or not, that would be great helpful to every demand learner to practice by himself.

### 1.2 Literature Review

Learning companion is a type of Intelligent Tutor System [1]. It can simulate the teaching way like classmates, partners. It can help students to get more ideas, better than the experience of studying alone. Using machine learning to build the learning companion, there are still some details need to figure out [2]. When design the learning companion, we should take these in mind: Cognitive, Motivation, Sociability, Attitudes and Ethics [3].

Taking the voice assistant for education is just about the beginning stage. Students can use it as Q&A station, timer... We also believe it is also good for special education, because the feature of voice interface [4]. For knowledge query, mental support and recreation, the voice assistant platform can play a good role for learning companion [5].

## 2 Design & Implementation

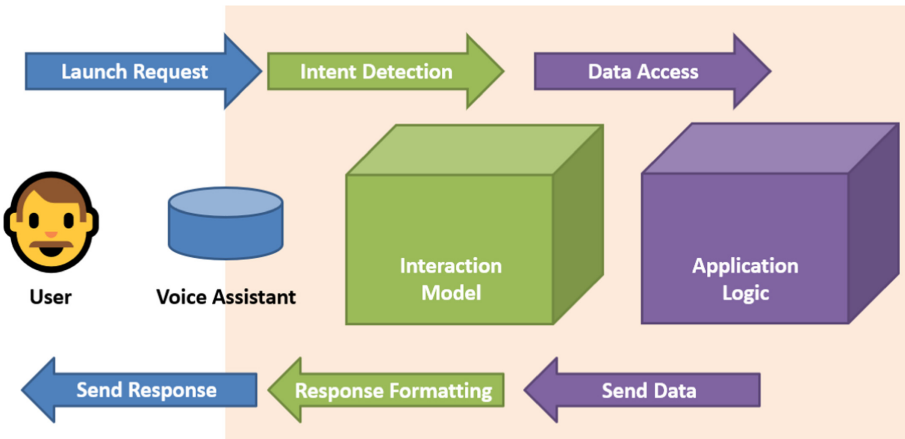
### 2.1 Voice Assistant

The voice assistant has several features just match the human beings; the following Table 1 gives the comparison.

**Table 1.** Voice assistant vs. Human beings

| Function | Voice assistant               | Human beings |
|----------|-------------------------------|--------------|
| Talk     | Speaker                       | Mouth        |
| Listen   | Microphone                    | Ears         |
| See      | Camera (only specific types)  | Eyes         |
| Show     | Monitor (only specific types) | –            |

Since the voice assistant can listen and talk basically, we are trying to make it as the conversation practice partner who never feels tired. At first, let us check how the voice assistant works (Fig. 1).



**Fig. 1.** How the voice assistant works

When the user reaches to voice assistant, he can wake up the voice assistant device with the invocation name like Alexa, OK Google, .... Then shoot the skill invocation name to enter specific skill. After the device get what user says, interaction model will recognize the sentence to matched intent, and get the keywords (also called slots or entities). The application logic usually designed by developer, to determine what kind of data to get and executes corresponding condition-based operation. After the operation, sends back the data to interaction model to transfer to the suitable data format (voice, text, image, video...) to user.

Based on this infrastructure, the researcher wants to design an oral practice language learning companion, when user needs to practice, he can wake up the machine and do the conversation practice, it never tired. The system design diagram below (Fig. 2).

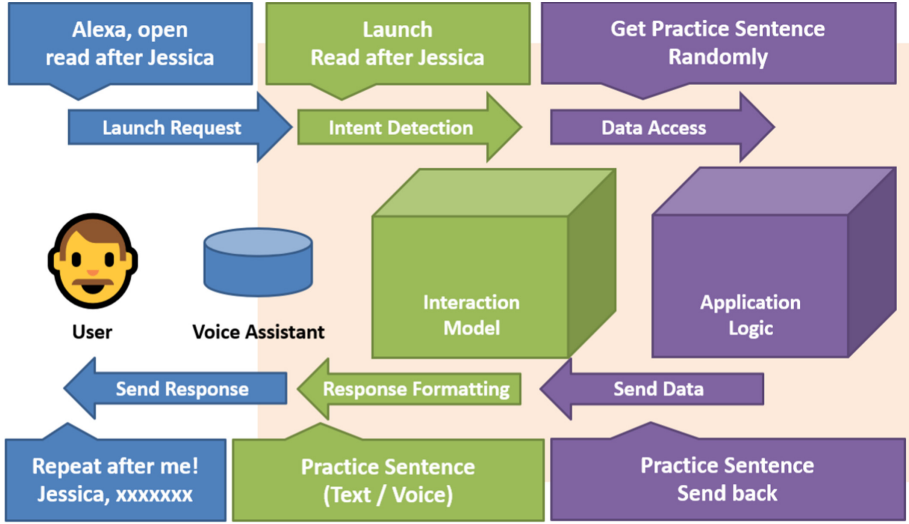


Fig. 2. How the oral practice language learning companion works

Let us go to the details for the interaction model and application logic. About the interaction model, we need to clarify what user says, and match to the right intent to get the next step. In this skill, we design 5 custom intents in the interaction model (Table 2).

Table 2. Intents in the read after Jessica

| Intent name       | Trigger                                     | Purpose             |
|-------------------|---|---------------------|
| HelloWorldIntent  | When user says something like hi, hello,... | Launch the practice |
| RepeatIntent      | When user request to say it again           | Repeat the sentence |
| AnswerIntent      | When user says "Jessica, xxxxx"             | Check user's answer |
| AMAZON.YESIntent  | When user gives positive response           | Go to next sentence |
| AMAZON.NextIntent | When user ask for next question             | Go to next sentence |

The core operation logic is in the handler for AnswerIntent, we will talk about it later. Let us review the user – voice assistant interaction flow first (Fig. 3).



Fig. 3. The interaction flow between user & voice assistant

In our design, the user needs to repeat exactly what the voice assistant says. So, it makes user to say the sentence very carefully. But sometimes the problem may incur from the voice recognition system or the microphone hardware. There are still lot of spaces to improve in this design scheme. Now let us go to see the details in the AnswerIntent (Fig. 4).

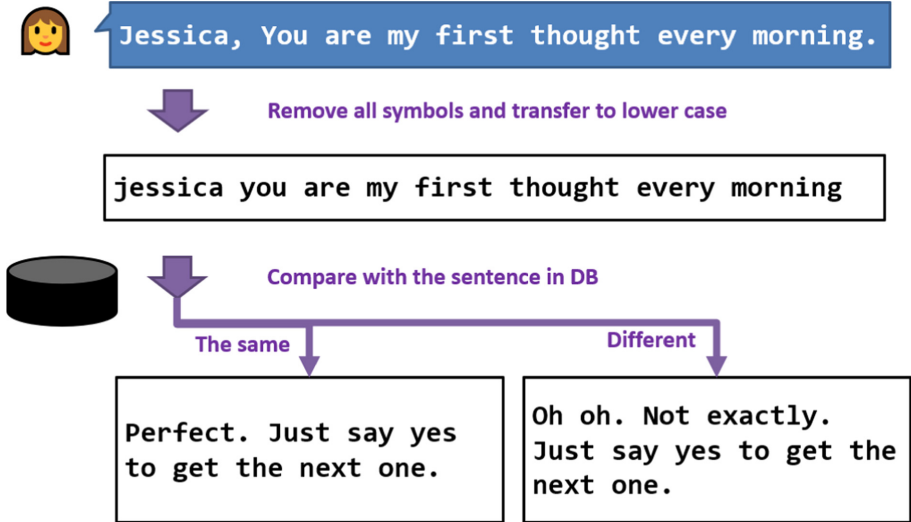


Fig. 4. The logic operation in AnswerIntent

When the voice assistant gets the practice sentence from user, the voice recognition system transfers it to text. Then the system determines which intent it belongs to. When it goes to AnswerIntent, it will firstly remove all symbols and change it all to lower case. The purpose for this step is for comparison later. Then the system will compare the sentence user spoken and the sentence in database, to see if it is the same or different. When it is the same, user will get the positive response, to know he is speaking right completely. Or user will get negative response, in those devices which have monitor, we also show what the system got from user, let user sees how far it is between with the right answer. This adding visual feature part is useful for language learner.

### 3 Results and Discussion

In this very early version of “Read after Jessica”, the researcher made it public to Amazon Alexa skill store.

Here is the result for the first month (Table 3):

**Table 3.** Users testing result

| Test number | Test record (5 times/unit) | Correct rate |
|-------------|----------------------------|--------------|
| 1           | XXXXX                      | 0%           |
| 2           | XOOOO                      | 35%          |
| 3           | XXOXX                      | 20%          |
| 4           | XXOXX                      | 20%          |
| 5           | OXXXX                      | 40%          |
| 6           | XXOOO                      | 60%          |
| 7           | XXOXO                      | 40%          |
| 8           | XXXXX                      | 0%           |
| 9           | XXXXX                      | 0%           |
| 10          | XXXXO                      | 20%          |
| 11          | XOOOO                      | 80%          |
| 12          | XXOXX                      | 20%          |
| 13          | XXOXX                      | 20%          |
| 14          | XXXXX                      | 0%           |
|             | Average correct rate       | 28.6%        |

We let them to interact with the voice assistant, and take 5 sentences as unit, no matter what the English level of user, for 5 sentences, user get the exactly right sentence are only around 1 ~ 3. According to the user response shows in the monitor, The common difference factors we list below (Table 4):

**Table 4.** The common difference from users

| Factors           | System   |
|-------------------|--|
| Spell error       | Can vs. can't, mine vs. mind...                |
| System error      | Amazon will transfer one to 1 from user        |
| Recognition error | Due to the volume or the quality of microphone |
| Speak too slow    | The system will stop listening to your input   |

## 4 Conclusion

It is the very beginning version of “Read after Jessica”, but we found the voice platform is a good way to simulate the oral practice learning companion. You can use it in any time you want, to get practice by your own. Although it just like a very strict teacher,

you can only get encourage when you get 100% right. But it can show you what you just said on the screen to check and do the oral practice repeatedly and continuously.

There are still some types of errors are nothing to do with users, so we can still improve the voice assistant platform, and add guidance to users who almost speak the same sentence.

## References

1. Chan, T.W., Baskin, A.B.: Studying with the prince: the computer as a learning companion. In: *The Proceedings of International Conference of Intelligent Tutoring System, ITS 1988*, pp. 194–200. Montreal, Canada (1988)
2. Chan, T.W., Baskin, A.B.: Learning companion systems. In: Frasson, C., Gauthier, G. (eds.) *Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education*, Chapter 1. Ablex Publishing Corporation, New Jersey (1990)
3. Chan, T.W.: Learning companion systems, social learning systems, and the global social learning club. *J. Artif. Intell. Educ.* **7**(2), 125–159 (1996)
4. George, T., Maya, S.: Voice assistants and smart speakers in everyday life and in education. *Inf. Educ.* **19**(3), 473–490 (2020). <https://doi.org/10.15388/infedu.2020.21>
5. Chung, H.J., Chan, T.W.: Virtual Learning Companion Feature Practicability on AI Voice Assistant Platform. *Taiwan E-Learning Forum* (2021)