



Vocabulary model and User context-based Speech Recognition Assistant System for the elderly living alone

Mi-Hwa Song¹

¹School of Information and Communication Science, Semyung University, Jechon, South Korea

Abstract

Background/Objectives: As the fourth industrial age begins, research on speech recognition is actively being carried out. **Methods/Statistical analysis:** However, since speech recognition is developed based on the standard language, the recognition rate is lower for people using dialects and unusual tones. Because of these problems, it is more difficult to gather data because seniors and dialect users do not attempt to recognize speech. **Improvements/Applications:** Therefore, to solve this problem, we have added a specific database and configured a speech recognition system that can be used by people who use dialects and unusual tones.

Index Terms

IoT, Raspberry pi 3, Speech Recognition, Vocabulary Model, User Context

Corresponding author : Mi-Hwa Song
mhsong@semyung.ac.kr

- Manuscript received April 22, 2020.
- Revised May 19, 2020 ; Accepted June 20, 2020.
- Date of publication June 30, 2020.

© The Academic Society of Convergence Science Inc.
2619-8150 © 2019 IJASC. Personal use is permitted, but republication/redistribution requires IJASC permission.

I. INTRODUCTION

Recently, the era of the Fourth Industrial Revolution has begun, and the interest and research on technologies such as artificial intelligence, IoT (Internet of things), and BigData are constantly increasing. The production of products equipped with artificial intelligence is increasing not only in industrial fields but also in daily life, and in particular, AI speakers and smart homes that utilize voice recognition are smoothly commercialized. This is because the voice recognition service of the AI speaker not only provides a service close to that of a personal secretary, but also enables control through interaction with things in conjunction with the smart home. However, the conventional speech recognition technology has a higher recognition rate for standard language users based on the standard language, but has a low recognition rate that the use of dialect and tone is not common, and different language customs. There is a problem to be solved. Particularly for elderly people, such problems are more noticed and displayed, and at the same time, there is a feeling of refusal and difficulty with electronic devices, so it is practically difficult to use products such as smart homes and AI speakers. Due to the advanced issues that should make the vast range of data collection recognizing the tone used by seniors, there are fewer users and more difficult to collect material.

In this paper, instead of collecting a vast range of data, frequently used dialects are arbitrarily added to the database to improve the recognition rate, and functions necessary for seniors are built into furniture and things and controlled by voice. By configuring an automated system to provide a smart home of how to improve their environment without user interaction. It's an improvement over the traditional smart home for seniors, designed to provide easy and convenient features for practical use in life.

II. RELATED WORKS

A. Speech recognition system

The first speaker-type device equipped with a speech recognition AI assistant is Amazon's Echo. Soon, however, Google released an assistant that developed Now, and launched Home, as Echo's competitor. In addition, Microsoft is also competing with Invoke[1].

Currently, smart smart homes through speaker-type devices are on the market, and smart smart homes are re-lighting the voice recognition technology that emphasizes the value of smart homes while providing users with optimized services to support comfortable life. Also, with the advent of smartphones, the "voice" interface, which allows easier device control in the

"touch" interface, which has become most common, is gaining attention as a smart home-based technology [1].

Fig. 1 is a basic configuration diagram of voice recognition. Through the process of characteristic extraction after voice input, we make three models in the decoder. The language model is a model for probabilistically converting speech information into grammar. The acoustic model is a model that stochastically converts what a word sounds like. Vocabulary modeling is a method of mapping the pronunciation of a word to a standard word in advance using a typical pronunciation, mapping one pronunciation to one word, and considering the dialect/tone, mapping the vocabulary and multiple headwords in advance. There are multiple pronunciation models. Lexical models perform differently in proportion to database size and quality. After the modeling is finished, the recognition result is displayed.

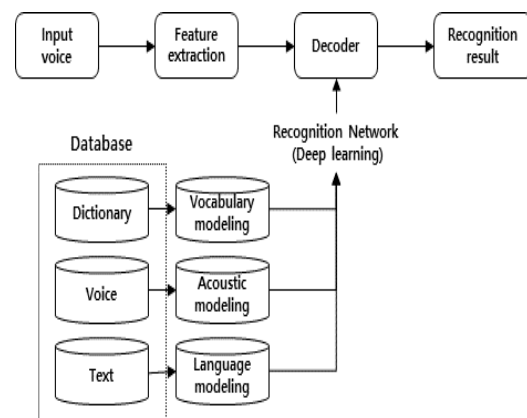


Fig. 1. Basic configuration of speech recognition system[2]

B. Smart home for the elderly

Currently, in the United States, a total of 11 types are focused as the main researches on smart homes. With fixed smartphones, smart doorbells, smart security alarms, smart fire alarms, smart leak detection, smart lighting, smart energy control, smart IP cameras, and real-time transmission system of activity data that can be used in everyday life, you can check your health through wearable devices. Advanced elements such as easy-to-use social media are incorporated as essential elements in smart home products for the elderly. General Smart Home If this function and convenience are focused on, smart home for the elderly will focus on management for the elderly, gentle adaptation of modern technology, environmental risk measurement and control. The research is actively underway.

C. Core technologies of smart home

Table 1. SMART HOME REQUIREMENTS FOR SENIORS

| Component | Function |
|---------------|---|
| sensor | Detect changes in residents and the environment |
| Processor | Data analysis and judgement |
| Communication | Communication between object |
| Interface | I/O of information |
| Security | Authentication for security |

Table 1 is the core technology of the smart home. The sensor is used to recognize changes in the user or the environment and convert them into data. Analyze the converted data, communicate with things through judgment, and provide appropriate services. The user can check the changes and circumstances of the surrounding environment in real time through the interface.

III. SPEECH RECOGNITION SYSTEM DESIGN

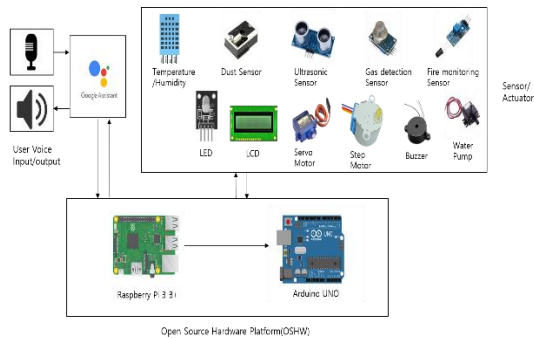


Fig. 2. Overall System diagram

Figure 2 shows the conceptual structure of the entire system. GPIO control and user-configured response through voice recognition in the Raspberry Pi 3 B+ are covered in detail in Figure 5.

In the case of the Arduino UNO, the Arduino sketch code stored in the Raspberry Pi 3 B+ is compiled and received. The compiled Arduino UNO provides the user with the results measured by the temperature/humidity sensor, fine dust sensor, ultrasonic sensor, gas detection sensor, and fire monitoring sensor through the LCD, servo motor, and buzzer. The temperature/humidity and fine dust sensor sends the result value to the LCD screen, and the ultrasonic sensor calculates the distance and recognizes when the user lies on the bed, and automatically closes the door. In addition, the gas detection and smoke detection sensors are a system

that notifies the user with a buzzer when the measured value exceeds a certain amount.

A. Vocabulary model

First, collect the data that is used frequently in dialect/uncommon tone, extract common parts and create a new database. After that, the process of lexical modeling is not performed in the existing database, but the common dialect database added randomly is performed first. If dialect data exists, the token processing other than mapping the token is transferred to the existing database, the remaining tokens are processed, and then the recognition network processing process is performed. If the dialect data does not exist, the existing database is used for the recognition network processing. Fig. 3 shows the overall flow of the dialect recognition improvement algorithm.

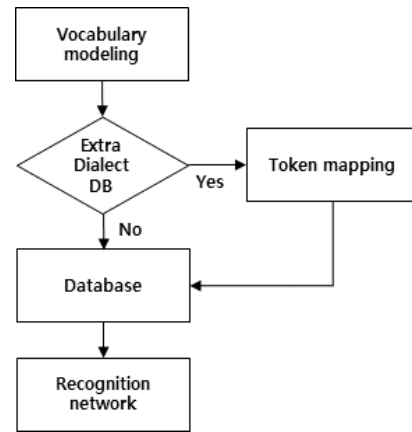


Fig. 3. Flowchart of dialect recognition improvement Algorithm

B. Speech recognition device service flow

Fig. 4 is a flowchart of the voice recognition device. Receive voice input to the Raspberry Pi using the microphone from the user. The Raspberry Pi that received the voice input performs voice processing via the Google Assistant SDK and Fig. After the process of 2, respond with GPIO Control and speaker.

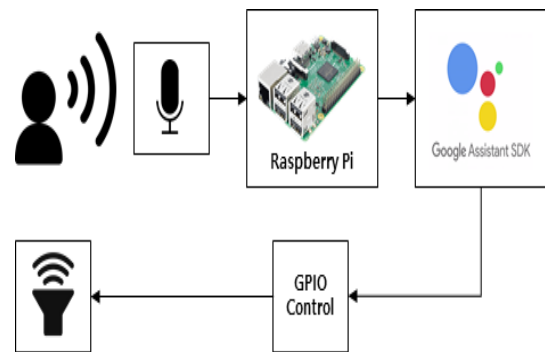


Fig. 4. Speech recognition device flowchart

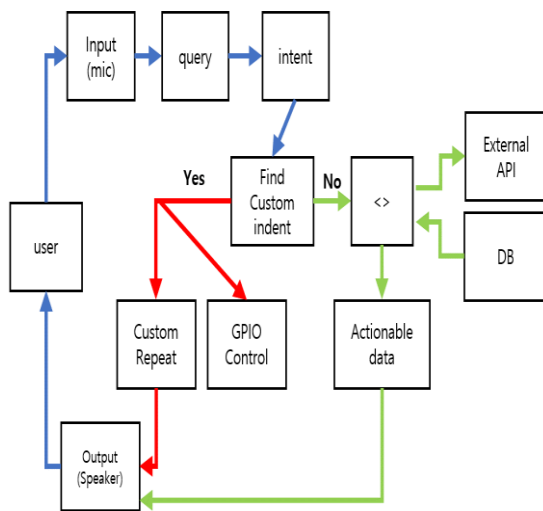


Fig.5. Speech recognition algorithm

Figure 5 shows the speech recognition algorithm using the Google Assistant API. When the user requests a request from the input (mic), the voice is decomposed into an intent through a speech-to-text (STT) process. If a custom intent is found in the decomposed intent, it exits from the existing algorithm and moves to the produced algorithm. The produced algorithm will perform GPIO control suitable for the user's situation. In addition, the response method is output to the speaker through the text-to-speech (STT) process. In addition, the user is designed to be able to communicate with the smart home service by allowing the user to continue the conversation by moving to the algorithm of the Google Assistant API after being output to continue the conversation.

C. Automation system

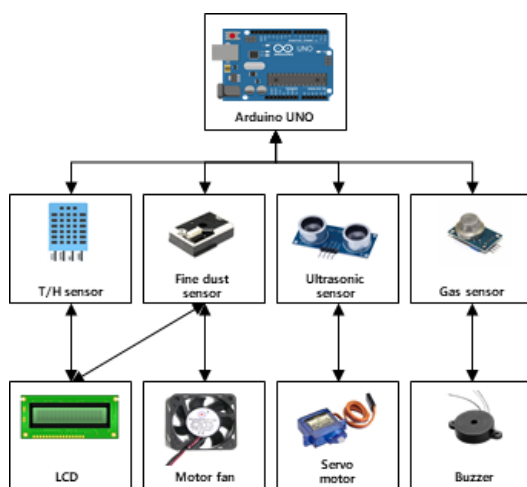


Fig. 6. Automatic system configuration diagram

Fig. 6 is a flow chart of the automation system. Arduino UNO measures the value every 0.1 seconds via the sensor. The values of T/M Sensor and Fine dust Sensor are displayed on the LCD for the user to confirm. When the value measured by each Sensor exceeds a certain fixed value, the Motor fan, Servo motor, and Buzzer connected to things are driven, and after executing the command entered in Arduino UNO, it is set to the initial value. To be done.

IV. IMPLEMENTATION

The smart home in Figure 7 is composed of a system that automatically controls using voice recognition control and Arduino. Raspberry pi 3 was used for the voice recognition development environment, and Python was used for the development language.

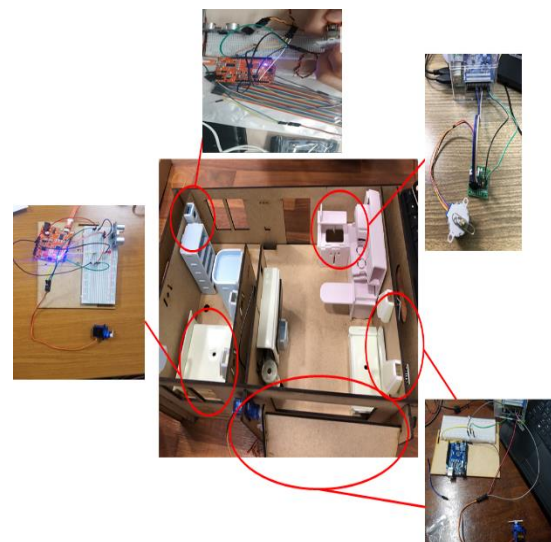


Fig. 7. Prototype of Smart Home for Elderly People

Basically, things are controlled through voice commands. The command can be executed even if a different tone is used, considering that the tone used by each person is different. In order to control things by voice, commands are stored in 'hotword'. Set each GPIO port number not to overlap. Through this, lighting using LED, washing machine and curtain using step motor, window using servo motor, and water purifier using water pump are controlled.

It is configured to operate by combining several sensors and motors in case a control is needed at the same time. In addition, a wake-up call sounds at the time specified by the user by adding the functions provided by the Google API. When it is time to take

the medicine, an alarm sounds, and the user can be reminded to manage the schedule to take the medicine without forgetting it. You can also get the information you want, such as weather and location. It also made it possible to play radio and music. A microphone and a speaker are configured to realize voice recognition.

The development environment of automatic systems that are not controlled by voice uses Arduino sketch, and the development language uses C language. When the user lies on the bed using the ultrasonic sensor, the door of the room is automatically closed, and the gas is measured and the buzzer sounds by using the smoke detection sensor to inform the gas leakage alarm and fire alarm. And the temperature, humidity and fine dust concentration inside the house are measured and displayed on the LCD.

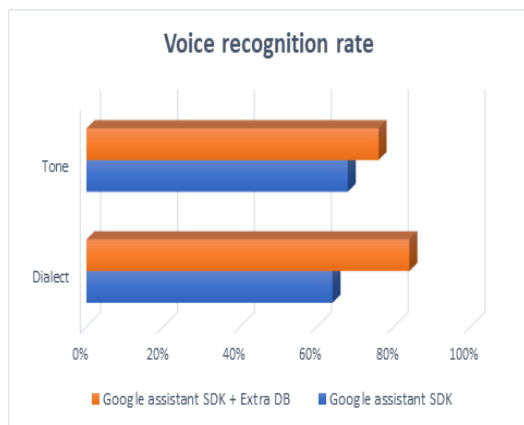


Fig. 8. Voice recognition rate comparison chart

Fig. 8 is a comparison chart for speech recognition rate. The existing Google assistant SDK showed the recognition rate of dial [$17/25 = 0.68$] and dialect [$16/25 = 0.64$]. On the other hand, the SDK that added the Extra DB showed the recognition rate of the tone [$19/25 = 0.76$] and the dialect [$21/25 = 0.84$].

V. CONCLUSION

In this paper, we increased the recognition rate of dialects/uncommon tones and implemented speech recognition that can be used by seniors and people who do not use standard words. Instead of going through the process of collecting and learning a large amount of data such as existing voice recognition, the existing dialects/tones with a lot of used and low voice recognition rates are selected, and an arbitrary small amount of data is selected. The voice recognition rate is increased by configuring the. As a result, it was

confirmed that the speech recognition rate increased by 8% in the tone and increased by 20% in the dialect.

We developed a smart home that can be easily accessed by equipping the voice recognition function to increase convenience and reduce loneliness at home in a modern society where the elderly and single-person households increase. When the user is in his or her space, he/she can control activities required for real life by voice recognition and automatic system to feel comfortable and convenient instead of loneliness and depression. It made me feel like I was talking to someone in a conversational format. It is expected that it will be available to users in various countries by enabling voice recognition in three languages: Korean, English, and Japanese.

REFERENCES

- [1] Eason, G., Noble, B., & Sneddon, I. N. (1955). On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 247(935), 529-551.
- [2] http://speech.cbnu.ac.kr/srhome/technology/continuous_speech_recog.html
- [3] Mardell, M. (2017). *Facing the robotic revolution*. Retrieved from <http://www.bbb.com/news/technology-3928030>
- [4] Maxwell, J. C. (1892). IP Multimedia Subsystem. *A Treatise on Electricity and Magnetism*, 2, 68-73.
- [5] Jacobs, I. S., & Bean, C. P. (1963). Fine Particles, Thin Films, and Exchange Anisotropy: (effects of Finite Dimensions and Interfaces on the Basic Properties of Ferromagnets). Research Information Section, The knolls.
- [6] Gilbert, T. L. (1956). Formulation, Foundations and Applications of the Phenomenological Theory of Ferromagnetism, Ph.D. dissertation, Illinois Inst. Tech., Chicago, IL, 1956, unpublished.
- [7] Arnold, D. P. (2007). Review of microscale magnetic power generation. *IEEE Transactions on Magnetics*, 43(11), 3940-3951.
- [8] Demokritov, S. O., & Demidov, V. E. (2008). Micro-Brillouin light scattering spectroscopy of magnetic nanostructures. *IEEE Transactions on Magnetics*, 44(1), 6-12.
- [9] Kaufman, C. J. Rocky Mountain Research Laboratories, Boulder, Colo., personal communication, 1992. *Personal communication*.
- [10] Yorozu, T., Hirano, M., Oka, K., & Tagawa, Y. (1987). Electron spectroscopy studies on magneto-optical media and plastic substrate interface. *IEEE translation journal on magnetics in Japan*, 2(8), 740-741.
- [11] Young, M. (2002). *The technical writer's handbook*. University Science Books.
- [12] Acuña-Villacorte, A. (2017). Standardization of Broadcasting Language Practices for International Comparability and Applicability. *International Journal of Emerging Multidisciplinary Research*, 1(1), 1-4.
DOI: 10.22662/IJEMR.2017.1.1.001.
- [13] Hong, S. (2017). Development of a Secure and Intelligent IoT System based on a Consortium Blockchain. *International Journal of Emerging Multidisciplinary Research*, 1(2), 9-14.
DOI: 10.22662/IJEMR.2017.1.2.009.

- [14] Eun-Sil Ha (2014). The implementation of Smart Care System for Dementia Patients," *Journal of the Korea Academy Industrial Cooperation Society*, 15(6), 3832-3840.
DOI : 10.5762/KAIS.2014.15.6.3832
- [15] SEOUL NPO Center, Archive Curator (2017). Introduction to Concept and Case study of 'ChatBot' based on Artificial Intelligence. Retrieved from <http://www.seoulnpocenter.kr/index.php>.
- [16] Jin-Wook Ha & Sang-Hyuk Kim (2017). Past, Present and Future of Artificial Intelligent Speaker. *Journal of the Korean Society Of Computer And Information*, 25(2),1-8.
- [17] Korean Consumer Agency (2017). Problems and Improvements of Artificial Intelligence (AI) Appliances : Focusing on Speech Recognition Speaker. Survey report, pp.1-33.