The Application of Artificial Intelligence Voice Recognition on Helping Elders Use Mobile Phones More Easily

Qingyuan Yao
Shanghai Foreign Language School, Shanghai, China, 200083

Corresponding author’s e-mail address: yaoqy4504@qq.com

Abstract. The problem that elders struggle to use mobile phone devices and elders often refuse to gain in touch with contacting devices with newest technologies have attracted people’s attention for a long time. It has already become a social phenomenon that elders tend to refuse smart devices such as smartphones. The reason is that most smartphones own a complex system, which has too high a difficulty for elders to master. The elders have a strong need for a way to remove the obstacle between their demands and the input of system orders. AI voice recognition, on the other hand, can basically remove this issue from consideration, for it can replace the manual manipulation part with machine-learning-based understanding of commands. Through deep learning methods mentioned in this paper, AI voice recognition can analyze orders given by voice, pick out key words and form into an order, and eventually execute the order properly. Voice recognition programs can also detect the elders’ health conditions through the attributes of the voice and make fast reactions automatically. This article aims to find a way to build targeted AI voice recognition systems for a specific kind of people, for example, the senior citizens, which is the target of the system in this article. This article also proposes rational prospects for future AI voice recognition applications in mental treatment or psychological researches areas.

1. Introduction

Ever since the invention of smart phones, many smart phone users have been troubled by the sometimes puzzling and complex mobile phone systems. Some systems have lots of choices, which makes giving orders complicated. Some systems show a lot of words on the screen, and thus they have to use smaller words, making sentence reading harder than ever; others may have unclarified system descriptions, keeping phone users from making accurate orders. As we step into aging society, it’s crucial to improve the phone using experience of elders, and make sure that they can use it without any confusion. Mobile phone companies have long discovered the problem that mobile phone usage has been too complicated for some elder phone users. The elders have given up trying to use mobile phones frequently in their daily lives due to the difficulties or annoyances during their phone usage. Companies and system-developing groups have been dealing with phone-using improvements or simplifications for a long time. Measures including bigger words, AI voice assistants are all effective solutions for this problem. They have improved the phone using experience of the elders by closing the gap between their demands and the inputs...
of computer orders, making their demands better understood by the system. Voice assistants now are very ‘intelligent’ already. They are able to transfer the voice order to correct word orders, and can perform well for most adults or kids. It’s convincing to say, nowadays, that AI voice assistants have made phone usage easier than ever.

Yet existing AI voice assistants now are too general for the elders, as the voice commands given by the elders are often misunderstood by general voice assistants. A recent survey showed that for the elderly, the general voice assistant mistakenly believes most commands when calling up Apps and converting voice to words. This is because elders have different needs from adults and kids, and their voice have different tone, lower speech rates, different speaking habits, or even different intonations. This would cause misinterpretations of elders’ words, and therefore make voice assistants much less efficient. For instance, they analyze false noises, for they have general tones instead of lower tones of elders. They may stop order input during the speech because elders’ lower speech rates, and misinterpret words for that elders may pronounce differently than general people, etc. These problems have been troubling old mobile phone users, making AI voice assistants far less effective than they ought to be.

The main purpose of this research is to find a way to make a voice recognition system successfully interpret the orders given by the elders, and by letting a system adapt new training methods, make it possible to develop a voice recognition system specifically targeted for elders. The article provides possible solution. It is to use special training data, targeted training methods, and different test standards, each of which would be mentioned in the next periods while developing the training system.

2. Literature review

In year 1983, the inventor Edward Porter built the first applicable voice recognition system. All the following voice recognition systems followed the way this very system was built: record verbal inputs, transfer verbal strings into word strings, get keywords out of word strings, and processing the keywords to form a sentence.[1]

In year 2008, the researcher Hisayuki Nagashima presented the specific method of getting keywords out from the verbal inputs. He made a clear analysis on the methods and devices that shall be used on the voice recognition process. [2]

In year 2001, researcher Sean Doyle gave the description of how to let a voice recognition system improve by itself. This can be seen as the fundamental step towards the creation of voice recognition systems involving artificial intelligence. This has given advice towards all voice recognition systems that are speaker-based or needs continuous improvements. This is crucial towards the building of the system mentioned in this article.[3]

The research done in year 2011 has proved the result of a certain AI voice recognition system to be highly dependent on the language ability of the user. This has proposed a new topic: how to build systems for users with poor language abilities? And meanwhile, this article is trying to solve the answer of this system.[4]

The researcher Paul Angott has concluded the system and method that can be used for a voice recognition system. In his theorem, he pointed out that systems can be put in pre-recorded audio patterns. These patterns would be of high priority in voice recognition process, creating inclinations towards verifying these patterns out. [5]
3. Analysis
In order to build a voice recognition system designed especially for elders, study about elders’ life patterns is necessary. It is found that a large part of elders’ reaction with phones is related to social activities and contacts. Compared to general phone users, they also have a larger demand of health-related apps, and few requirement of entertaining apps, including video websites and game apps. They also have lower speech rates and tones, and more varied speaking habits. All of these attributes led to a completely different, and far larger database than normal voice recognition training systems, which lead to special training data, targeted training methods, and different test standards, as mentioned before.

3.1. Training data
The training data mainly consists of 6 sections: phone calls, word contacts, health-related commands, emergency calls, works and entertainment. According to the survey, the proportion of word contact section orders is far larger than other sections. Therefore more test data are needed to provide extra accuracy for this section. The four other sections, including phone call section, health-related demand section, work app section and entertaining app section, have declining frequency of calling, therefore they are provided with declining test data amount. The frequency of calling out emergency section, on the other hand, couldn’t be measured generally like other sections. Extra data are also provided for this section, for lack of accuracy may lead to extremely serious consequences. Meanwhile, the six sections would be given different weights, depending on both the big data and later learning of the user itself. More frequency of mentioning a special kind of section, or more frequency of reinputting the voice command in the application all make the weights fluctuate, depending on the using habit of the user himself.

Figure 1. The proportion of each section in daily mobile phone usage
Figure 1 was reached from a survey conducted in nursery homes and neighborhoods, covering the sample amount of 1000 senior citizens in Shanghai. It has shown the ratio of frequencies that the key words of each section is estimated to be used. This has, meanwhile, set up several basic coefficient while building up the model, including the amount of database needed to train every keyword section, the weight of each section, and the required accuracy of each section in order to provide the best user experience, etc.

3.2. Training methods
There would be two big parts of the process of transferring voice commands to computer orders: The first part is to separate target sound track from noises. For this part, the system would first divide the target data into several sound tracks, and then pick out the required sound track using K-Nearest Neighbor methods, for which attributes include volume, sound clarity, tones and speech rates. Compared with other modules like vector learning and etc., K-Nearest Neighbor method has no less accuracy. The coefficients are far more easy to change, and the system would have more flexibility adopting the K-Nearest Neighbor method. For example, if the user has more frequency in calling out working section orders, the system using K-Nearest Neighbor method can make changes in the lowest attempts, making weight for the working section larger than before, for that only the weight coefficients need to be adjusted.

The second part is to pick out key words from the sound track, and therefore form a computer order. For this part, the system would firstly divide the sound track into words, and then pick out key words, using Bayesian learning to count the probability of each separated word being a key word. It is worth noting that this learning process may result in different results according to the weight of different sections in the data. Bayesian learning has already been tested to be the most effective method, since it gives clear reflections of probabilities, and may be able to give multiple choices of orders, depending on the final probabilities of each word being a key word. For instance, if there are two potential key words detected in the same voice order, the system would give three different orders, allowing the user to choose from activating the first order, the second order or both of the orders.

The difference between this training method and training methods for general voice systems is that there exist the process of valuing speech rates, for that elders tend to speak slower when giving voice commands to the AI voice recognition system. Therefore, the standard of ‘generality’ might vary in the two systems, reflecting in that this training method may allow pauses or lower speak rates, while normal training systems don’t.

3.3. Test standards
The difference between test standards of this system and general voice recognition system is that the test standards are more elder-based, and requires higher degree of fitting, for the system might save lives in emergency conditions. The expected rate of accuracy is over 95 percent.

3.4. Test results
Due to the limit of computing skills, the training and testing of the system did not end up in a actual voice recognition system. Yet some of the fundamental researches gained one conclusion: In the transferring process from verbal inputs to word strings, a main recognition frequency is
needed in order to provide accurate capture of the target verbal input. Audio tracks with frequency
near the main recognition frequency will be prioritized, as a method to determine which audio
track should be analyzed.

The main recognition frequency is best located between 350 to 500 Hz. This is because most
elders’ frequency of speech is between 350 and 400 Hz. The highest peak of accuracy is located
near 370 to 380 Hz. The second highest peak occurs at near 430 to 450 Hz. This is probably
because most male participants speak at a frequency near the first peak, and most female
participants speak at a frequency near the second peak.

4. Conclusion
In the future, the voice transferring method for elders can be made into an app, providing elders
with better and simplified phone-using experience, and enabling them to accept smart phones
gradually. Methods for voice recognition apps or programs for a specific kind of people have
tremendous future potentials. Similar methods can also be used in improvements of mobile phone
using for kids, or patients suffering from mental diseases like autism. It would be not only a more
convenient way for them to use mobile phones, but also would probably help us discover more
about worlds inside their minds.

The emergency section and the health-related commands section may be integrated into one
section, which mainly serves to detect the uses’ health condition. The new ‘health section’
can collect data of the user’s frequency of pausing, the amplitude of the sound trembling, or the
way that the voice sound differently with normal days to discover potential diseases or health
threats. Working together with extra thermal modules and extra pulse valuation facilities, if
possible, the voice recognition module may even be able to give part of the normal health checks,
save the elders trouble to go to the hospital and have a health check. The previous emergency
section, at the meantime, would still take its duty to provide automatic emergency callings and
show first aid instructions on the screen of the mobile phone.

5. References
[1] Edward W. Porter, Boston, Mass. Voice recognition system. United States Patent. Publication
number: US4829576A. May 9, 1989. https://patents.google.com/patent/US4829576A/en#title
[2] Hisayuki Nagashima, Wako. Voice recognition device, voice recognition method, and voice
recognition program. United States Patent. Publication number: US8005673B2. Aug. 23,
2011. https://patents.google.com/patent/US8005673B2/en?oq=US8005673B2
[3] Sean Doyle. Automatically improving a voice recognition system. United States Patent.
Publication number: US7103542B2. Sep. 5, 2006.
https://patents.google.com/patent/US7103542B2/en?q=US7103542B2
[4] Tyler K. Perrachione, Stephanie N. Del Tufo, John D. E. Gabrieli. Science.[J] 29 Jul 2011: Vol.
333, Issue 6042, pp. 595 DOI: 10.1126/science.1207327.
[5] Paul Angott. System and method for voice recognition. Publication number: US8510103B2
Aug. 13, 2013. https://patents.google.com/patent/US8510103B2/en?q=US8510103B2.
Acknowledgement

Here I would like to thank my instructor Professor Rakesh Kumar for his instructions and his inspirations for me during my process of writing the essay. Meanwhile, I’d also like to appreciate the teaching assistant, Jiatong Shi, for his support and advice towards this essay. They have given me lots of suggestions, most of which have been adopted in this essay. Without their effort, this essay couldn’t have been written out.