Advances in Intelligent Systems Research, volume 146
2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)

Visual Feedback Design and Research of Handheld Mobile Voice Assistant Interface

Wenjun Hou and Yang Yang*
Beijing Key Laboratory of Network and Network Culture, Beijing University of Posts & Telecommunications, Beijing, China
*Corresponding author

***Abstract*—Based on the limitation of handheld mobile voice assistant in usability, this paper analyzes the important role and characteristics of visual feedback in voice interaction interface, and discusses the research and the application status of voice assistant usability. On this basis, taking the interaction of Siri voice assistant as a case to study, the task characteristics and design defects of voice interaction are found through interview 12 users. Based on the general design principles and voice interaction design principles, the voice interaction feedback interface in handheld mobile devices is optimized to improve the validity and reliability of the intelligent voice assistant.**

Keywords—Siri voice assistant; voice interaction; visual feedback; usability

I. INTRODUCTION

With the strengthening of the dominant position of mobile Internet, voice interaction technology has been widely used in hand-held mobile devices. Language is the most primitive and most instinctive way of communication between people. As early as 2007, Bill Gates predicted that voice technology will replace traditional keyboard and mouse as the mainstream interactive mode in the future. The terminal device will break the limitation of the screen and the operation, and voice interaction will be the preferred choice of human-computer interaction technology [1].Voice interaction is a technology that studies the conversation between human and computer that is the interactive technology that human input natural language and computer output synthetic speech answer. Voice interaction is characterized by high efficiency, naturalness, flexibility, sensitivity and transience, and has the advantages of low cognitive load and high interaction efficiency. However, there are some inevitable shortcomings, such as the high demand for the surrounding environment, that is, the noise should not be too large, the information is fuzzy, and the memory load is heavy. These problems have a significant impact on the convenience of voice interaction, but also increase the complexity of human-computer interaction.

In order to make up for the deficiency of the research on the interaction of the intelligent voice assistant on the touch-screen mobile phone, this paper, from the aspect of interaction design, based on the theory of usability, tries to sum up the task mode suitable for the intelligent voice assistant of the touch-screen mobile phone through interview method. Find out its design defects in usability, and analyze the specific problems in depth, and propose a specific improvement scheme based on usability general design principles.

II. RELATED WORK

Currently, Nuance, the world's largest voice technology company, has developed Naturally Speaking, the world's most advanced computer voice recognition software [2]. The company's flagship product is the T9 smart text input method, powerful enough to support more than 70 languages, more than 3 billion mobile terminals are built into this input method. Microsoft achieved a major breakthrough in speech recognition through deep neural network technology, reducing the error rate to 18.5% and the accuracy rate 33% higher than the traditional technology [2]. The Google Cloud Speech API, covering over 80 languages and compatible with any real-time streaming or batch mode application, will provide the application with a complete set of APIs to bring "Look, Listen & Translate" Aspects of the function [3]. This technological breakthrough from Google will have a big impact on the entire industry. Siri Apple introduced the mobile intelligent voice assistant on behalf of almost.

At present, most of the researches on voice interaction are focused on the technical level, and the establishment of interaction theory system is still in the initial stage of exploration. Some scholars from the use of the crowd, make the survey on the children and the elderly these special groups of people using voice assistant behavior habits and obstacles encountered [4, 5]. Role-playing and focus group methods have proved to be effective in the design of voice assistants [6, 7]. Julia Kiseleva et al have done a continuous and in-depth study on the user satisfaction of intelligent assistant, and found that the factors affecting task satisfaction are different in different scenarios. They also try to use interactive signals to construct a method to predict the satisfaction of intelligent assistants [8, 9, 10]. Other studies have found that there is a relationship between the personality of users and their preference for different intelligent assistants [11]. Most of the time, scholars only use sound as a secondary interactive channel to assist other forms of interaction, and apply speech as an interactive mean to specific scenarios or tasks. For example, using auditory feedback to assist pen or gesture interaction, or to shape the user's touch behavior [12, 13, 14].

III. METHOD

This study uses a non-structured interview, the main content is the overall feeling of users using Siri to complete daily interactive tasks and the user's willingness to use voice assistants. Extract valid information from the interview results

to find flaws in usability feedback from the voice assistant interface.

This study invited 12 users to participate in the interview, all of them were graduate students, including 6 males and 6 females. The age of the subjects ranged from 20 to 25 years old. All of the subjects had experience using Siri voice assistant, but they were not skilled users in Mandarin standard. Considering that the vast majority of users have not yet cultivated the habit of voice interaction in touch-screen mobile phones, and the development of intelligent voice assistants is not complete, there is a large room for development. Therefore, the lack of experience in using speech assistants has little effect on the results of this study.

IV. EXISTING PROBLEMS

Take Siri as an example, summarizes the following several common problems that exist in voice assistants.

A. There is a "Pseudo phonetic" Interaction

Usually the user says that opening an application is a desire to continue to use voice control after the application is opened. Siri, on the other hand, automatically exits after opening, then returns to the gesture interface completely.

When using Siri, there are also a lot of "Pseudo speech" interactions. When Siri gives multiple options for the user to choose from, it is expected that the user will manually click on the selection rather than the form of the dialog, where a manual trigger button is required to continue the conversation. Most users do not notice this detail, and they will continue to input information until a sentence is finished but found not recognized, can only say again, and affect the efficiency of the interaction.

B. Unable to "Fast-forward" Dialogue

Siri cannot interrupt when it outputs the result or it will terminate the task process. The process of getting information quickly with the eyes is much faster than listening to the Siri voice reading screen, especially when the Siri gives a long paragraph of text feedback, users have a "Fast forward" requirement.

C. The Way of Feedback Is Inappropriate

- Interface feedback: Users expect different feedback on different tasks, such as book a car, text messages, such as tasks that need to look at the screen check output results, and tasks such as looking at the weather, users do not want to disperse the visual channels to see the results. In order to find the right user requirements, the output of the interface results is redesigned according to the voice scene.
- Content feedback: First, Siri should pay attention to simplifying the content when it comes to feedback. According to the existing way, the voice reading screen will waste a lot of time, and occupy the user with too much attention, then it will increase the memory burden of the user, and it is not easy for the user to focus on the key information quickly. Second, too much text feedback effect is not good, users will not read carefully, but think that the operation is wrong, in most cases will directly ignore the feedback information to try again. Third, the conversation is not coherent, often not giving the user the opportunity to speak, such as checking the weather task in which Siri directly outputs a list of weather conditions for each time period, but the user's visual channel may be occupied. Users may say, "Please read for me." But Siri can't do this right now. Siri indicated that he could not provide bus trips or schedules in the search for routes, but could look for bus routes from the current location. When the user lets it execute without a word, the boot does not work well.
- Technical feedback: First, when the network is not smooth, Siri does not explicitly say that the network is wrong, but presents a state of long thought, showing "I am listening, please continue". Cause the user to wait and try to say "Hey Siri" to trigger it; Second, in the completion of such more complex tasks as set reminders, there will be a user input the same command, but there are two different results; Third, the fault-tolerance rate for interactive terms needs to be improved, for example, when the user has entered the "Taxi" keyword, Siri cannot judge the user's intention, it should be more intelligent.

D. Operational Guidelines are Less Practical

Siri provides users with a detailed guide to native application voice operations, but users often have no patience to take a closer look, and prefer to learn through their own trial and error. In the course of specific interactions, Siri should provide corresponding operational guidance feedback in the first place if users encounters difficulties, rather than having the user manually look it up in all the instructions. This increases the user's interaction burden.

E. Continuous Dialogue Context Confusion

When the user enters a piece of information, the system outputs the wrong result due to some reasons, such as incorrect identification, ambiguous understanding or beyond the ability of the voice assistant. At this time, if the user wants to change through the voice, the result will be more confused. The user will want to start over after the change fails, but the voice assistant can't tell if it's a new conversation, and his memory will stay in the last conversation, so there will be a situation where the answer is wrong and the user is farther away from the user's intention. In most cases, the user has to exit the voice assistant and reenter, so as to ensure the correct dialogue context.

V. DESIGN PRINCIPLES

This paper summarizes and supplements the improved design principles from the general design principles and the voice interaction design principles, in order to solve the shortcomings in the current voice interaction.

A. General Design Principles

- Performance load minimization.

  Effective load includes cognitive load and exercise load. Reducing cognitive load means removing redundant information and forming memory module with information needed to be memorized. The problem of small memory capacity can be overcome by compiling longer information into several small units. Increasing the amount of information stored in conscious memory. The reduction of exercise load depends on reducing the operation steps, reducing the overall movement and energy consumption, and automating the repetitive work.

- Recognize, not recall: Identifying memory is easier to cultivate than recalling memory.

  Recognition of memory by receiving information, and not necessarily related to the source, content, or relevance of the memory, which has been experienced in the past (such as a scene, a sound, a taste, a touch) [15]. The interactive interface feedback should follow the "recognition easier than recall" rule, and minimize the need to recall information from memory. Use existing menus, decision aids, or similar devices to create clear options.

- Picture Superiority Effect.

  People has a better memory of pictures than words or breakthroughs. Therefore, when the voice interface to feedback more information, we should try to use the form of charts rather than large sections of text stacked. When people inadvertently receive information, and receive a limited time, this principle will be better. Users usually do not look at the screen carefully but glance it roughly when using voice interaction. At this time, this principle is used to increase people's discernment and help the user to recall important information.

- The principle of readability.

  Clarity of visual feedback is usually determined by font size, font size, contrast, text modules, and spacing. The readability design of language interface should meet the needs of the users to read quickly.

B. Principles of Voice Interaction Design

- Use voice response instructions to avoid "Pseudo phonetic" interaction.

  The voice assistant assumes that the user has not looked at the device screen all the time and responded to the user's instructions as fully as possible. For example, when looking at the weather, the user's expectation is for the system to broadcast today's temperature directly, with human cues such as "Colder weather and more clothes." Instead of giving a temperature list for each time period, users can use visual channels to get information. If the target contact has multiple phone numbers during send a short message, the voice assistant can read out the first three digits directly, or "Home number or work number."

- Voice Assistant and device Application parallelism.

  The main reason for the creation of voice interaction is to liberate human hands, although the current level of technology does not completely separate it from other organ aids, but designers still need to make improvements in this direction. As far as possible, complete the user's operation instructions within the voice assistant to avoid "Pseudo phonetic" interaction with GUI-VUI-GUI, if the voice assistant is unable to do so, ask the user if he or she is allowed to open other applications.

  In the case of Siri, when the user enabled it, the interface of the device was occupied and no other application could be operated. But after opening other applications through Siri, you can't continue to use voice to manipulate the application. This is the concrete embodiment of "Pseudo phonetic" interaction, but also a pair of irreconcilable contradictions. Therefore, voice assistants should avoid exclusive interfaces, and users can still use voice control after opening other applications on the device.

- Dialogue can be interrupted at any time to achieve "Fast forward".

  At present, the phenomenon of voice assistant reading screen is serious. Users must wait for the voice assistant to stop speaking before they can do the next operation, which not only wastes time but also consumes the user's patience. The voice assistant should allow the user to interrupt it at any time and resume a new conversation. Considering that the user is not convenient to look at the information on the screen, the voice assistant can report the output results to the user concisely and stop reading the screen at length.

- Real-time feedback operation mode.

  One of the concerns of users when using voice interactions is that they are not sure what terms can be recognized. When there are multiple operational errors, most users will choose to give up the voice operation directly, rather than viewing the operation guide to learn. Voice assistants should prompt users immediately when they make mistakes or become hesitate, such as "You can ask me: What's the weather like today?" The timing and content of the feedback given by the voice assistant is very important to improve the user's satisfaction.

- Provide personalized customization.

  Many voice assistants need fixed statements as activators, such as "Hey! Siri". In fact, let users give their own voice assistant a personalized name, each time the user awakened like a pet call, and the voice assistant's answer can be changed from "Hello" to "Master" and so on. The addition and change of this

kind of person will increase the interest of the speech experience, cultivate the user's usage habit, and enhance the user's stickiness.

VI. Design Applications

According to the interaction design principles and design points of the voice assistant interface proposed above, the interface of the Siri voice assistant is chosen as the analysis target. The main advantages and disadvantages of the voice interface are summarized, and suggestions for improvement are put forward.

A. Interaction Mode

- Interaction scope expansion.

  In the design process of voice interaction, many people naturally think that speech liberates the hands, thus they neglect the thinking of visual interface. Voice assistants should avoid interface exclusivity, that is, after opening native and third party applications, they can also use voice control operations. For example, a user uses Siri to turn on a music playback software and continue to search for the target song and play it. In other applications, the voice assistant should have a minimal suspension prompt that allows the user to return to the voice interface at any time without interrupting the current operation.

- Step by step guidance.

  For more complex tasks, the user wants to have a step-by-step boot that can split the steps and execute them separately. Otherwise, users input a large amount of information at one time, not only to consider how to organize the language, but also to worry about whether the voice assistant can identify, to bear the burden of excessive. If there is an error, the user does not know where the error is, and the step-by-step operation helps the user to modify it specifically.

- Proactively check and correct.

  One of the problems in voice interaction is that it is not convenient to modify. At present, the main modification way in Siri is to click on the screen to modify manually, which is a problem of "Pseudo phonetic" interaction. At this point, Siri should provide the user with a modified prompt, such as adding "Am I right" after the output?" Like a dialogue, there is a return to avoid taking up the user's physical resources.

- Can be interrupted at any time.

  When people communicate with each other, they can interrupt the conversation at any time, and the voice assistant also needs to do this. For example, when a text message is received, the voice assistant reads the text, and the user can pause or switch to the next one at any time. When playing music, the user can cut the song or pause the play.

B. Feedback Methods

- Scene processing.

  The voice assistant should be humanized and actively guess the user's intentions. For example, when viewing the weather, the user expects to receive information directly from the auditory channel without looking at screen feedback .And when you check the route, the user has to check the screen information to be sure. In addition, the voice assistant provides friendly reminders at the right time, which can neither disturb the user nor guide the learning. When the user rummages in the music application, the voice assistant prompts the user to say the song name or the singer’s name to look for the song; When the user looks for pictures in the album, the voice assistant prompts the user to try to "Look for photos taken at home yesterday".

- Guided feedback.

  The voice assistant acts as a "Teacher" and teaches the user how to use it. When Siri says "I'm sorry, I can't figure it out" when it comes to unrecognized content, users need Siri to tell them what to do to be recognized, rather than this useless feedback. Similar to the case of selecting a target from multiple output results, both click and voice operations are possible. If at this time the user's hands may be inconvenient to click on, and he does not know that voice operations can also complete the task, this time the voice assistant needs to give a prompt feedback. The key words are extracted and the voice information input by the user are inferred reasonably. Siri, for example, does not recognize what "Reserve a car" means, but it can guess that the user's behavior is to make an appointment or search for car-related information, it can provide the user with a guess and ask if the user is correct. The most important thing is to teach users how to organize the speech to be recognized.

- Content structured feedback.

  We can highlight the key information by highlighting the color bar, such as the time, place and other information that the user will definitely view with the blue strip after the car appointment. The voice assistant takes the place of the user to bear the cost of information acquisition and improves the interaction efficiency.

VII. Conclusion

Interface visual feedback plays an important role in voice interaction. Through targeted usability interview, this paper summarizes the task mode of language interaction, discusses the problems of handheld mobile voice assistant, and complements the principle of voice interaction with relevant theories. This paper proposes a multi-channel interaction mode with visual feedback as an auxiliary channel, hopes to provide new ideas and directions for the design and research of handheld mobile language assistants.

REFERENCES

[1] Gu Yaping. Wisdom voice assistant system based on intelligent voice interaction technology [D]. Nanjing University of Posts and Telecommunications, 2015.

[2] LU Tian-zeng.Research on Interaction Technology of Intelligent Robot Based on Android [D] .China Ocean University, 2015.

[3] Internet of Things Think Tank [EB / OL]. [2016-05-05] .http: //news.rfidworld.com.cn/2016_05/a02750c38c553788.html

[4] Lovato S, Piper A M. "Siri, is this you?": Understanding young children's interactions with voice input systems[C]// International Conference on Interaction Design and Children. ACM, 2015:335-338.

[5] Wulf L, Garschall M, Himmelsbach J, et al. Hands free - care free: elderly people taking advantage of speech-only interaction[C]// Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. ACM, 2014:203-206.

[6] Lee S S, Lee J, Lee K P. Designing Intelligent Assistant through User Participations[C]// Conference. 2017:173-177.

[7] Meurisch C, Ionescu M D, Schmidt B, et al. Reference model of next-generation digital personal assistant: integrating proactive behavior[C]// ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2017 ACM International Symposium on Wearable Computers. ACM, 2017:149-152.

[8] Kiseleva J, Williams K, Jiang J, et al. Understanding User Satisfaction with Intelligent Assistants[C]// Conference on Human Information Interaction and Retrieval. 2016:121-130. Kiseleva J, Williams K, Jiang J, et al. Understanding User Satisfaction with Intelligent Assistants[C]// Conference on Human Information Interaction and Retrieval. 2016:121-130.

[9] Kiseleva J, Williams K, Awadallah A H, et al. Predicting User Satisfaction with Intelligent Assistants[J]. 2016:45-54.

[10] Jiang J, Hassan Awadallah A, Jones R, et al. Automatic Online Evaluation of Intelligent Assistants[J]. 2015:506-516.

[11] Patrick Ehrenbrink, Seif Osman, Sebastian Möller. Google Now is for the Extraverted, Cortana for the Introverted: Investigating the Influence of Personality on IPA Preference[C]// Proceedings of the 29th Australian Conference on Computer-Human Interaction. 2017:257-265.

[12] Andersen, T.H. and Zhai, S. “Writing with music”: exploring the use of auditory feedback in gesture interfaces. ACM TAP 7, 3(2010), 1-24.

[13] Sarah, M.S., Ruiz, J., Using Audio Cues to Support Motion Gesture Interaction on Mobile Devices, In Proc. of CHI, ACM Press (2014), 1621-1626.

[14] Tajadura-Jimenez, A., Liu, B., Bianchi-Berthouze, N., and Bevilacqua, F. Using sound in multi-touch interfaces to change materiality and touch behavior, In Proc. of NordiCHI, ACM Press (2014), 199-202.

[15] William Ridewell, Kristina Horton, Jill Butler. Universal Design Rules [M]. Central Compilation and Translation Press, 2013