UDC 67

RESEARCH ON VOICE INTERACTION DESIGN BASED ON THE ATTENTION MEASUREMENT METHOD OF STROKE

[1]Jiang Xinyu, PhD, Associate Professor, Department of Machinery
[2]Wang Xinyi*, 221 MSc in Mechanics, Year 1
[3]Li Zhijun, 228 MSc in Mechanics, Year 1
Donghua University
Shanghai

Abstract: Under the background of intelligent era, voice interaction technology has developed into the most convenient and the most humanized way of human-computer interaction. The popularity of mobile intelligent equipment has expanded the depth and breadth of voice interaction technology. In this paper, voice interaction mode is taken as an experimental variable to conduct a controlled experiment on the influence of voice assistant on people's attention. Based on the measurement method of STROKE in psychological assessment, quantitative and visual experimental results are output to analyze the causes of distraction caused by voice assistant in the interaction process. An interactive design strategy is proposed to deepen interaction patterns, accurately identify information and reduce attention consumption.

Keywords: Voice interaction; Attention mechanism; Measurement of attention; The product design

Introduction

In recent years, with the popularity of intelligent mobile terminals, intelligent voice interaction, as a new generation of information transmission mode based on voice input, changes the traditional touch input mode and improves the service efficiency of intelligent devices [1]. Voice interaction enables both "listening" and "speaking" capabilities in the human-machine interface, which expands the breadth and depth of the current interaction design.

With the maturity of the voice interaction chain, the expansion of the real scene data scale and the continuous improvement of cloud computing capability [2], the study of voice interaction is also transforming into the study of the whole life situation of users. In today's interconnected world, voice will liberate our hands and feet, make the user input more convenient, make the service more efficient, and become a milestone in the development of mobile Internet [3].

With the introduction of Attention mechanism, the overall performance of voice interaction technology has been further improved [4]. This study unfolds from the influence of voice assistant on human attention, the control test was set by the measurement method of STROKE, analyzes the causes of attention distraction during interaction, so as to propose interaction design strategies to accurately

identify information, deepen interaction patterns and reduce attention consumption.

1 The framework and technology of voice interaction

The essence of voice interaction is human-computer interaction. Through the interaction, communication and information exchange between human and computer, a series of output and input are generated, so as to achieve the purpose of the task, and the information carrier of voice interaction is voice. The key technologies of voice interaction include voice recognition, voice synthesis, and semantic understanding [5]. Voice recognition is the process of analyzing voice signals and converting them into text sequence. The process and framework of voice recognition are shown in Fig 1.

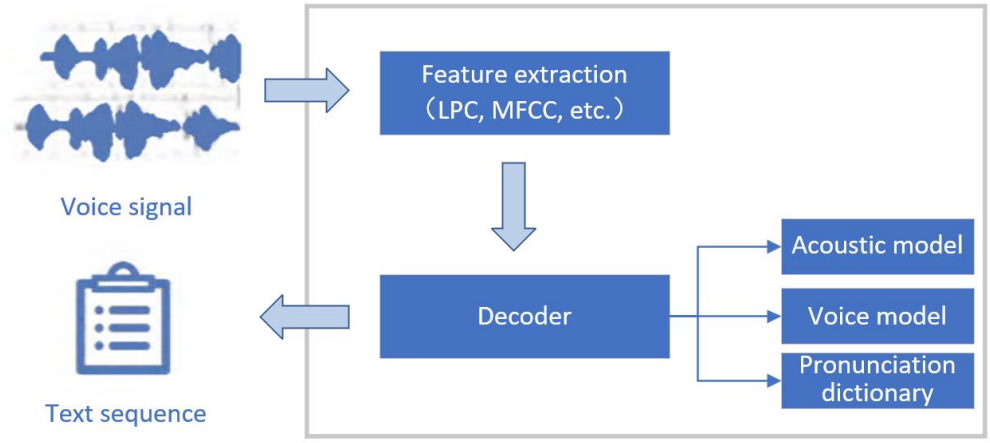

Fig 1. Voice recognition technology framework

The voice interaction process is as follows: The user uses the microphone to input the voice, and the voice is converted into text by the voice recognition subsystem, and the text is converted into the user's intention by the semantic understanding subsystem. The user's intention is completed by the executive subsystem, and then the execution result text is transmitted to the voice synthesis system, and finally the text and the synthesized voice are presented in front of the user respectively.

In recent years, with the introduction of Attention mechanism, the integration of voice enhancement and acoustic modeling makes human-computer interaction closer to the communication mode of real human language, and further improves the overall performance of intelligent voice technology [6].

Studying the influence of voice interaction on human attention and quantifying its results from scientific data is of important reference significance for the design and development of voice interaction.

2 The study of attention

2.1 Factors and traits of attention

Attention is a common psychological feature accompanied by psychological processes such as sensory perception, memory, thinking and imagination, which is the direction and concentration of psychological activities to certain objects. Attention has two basic characteristics, one is directivity, which refers to the mental activities to selectively reflect some phenomena and leave the rest of the objects.

The second is concentration, which refers to the intensity or tension of mental activity staying on the selected object. There are four characteristics of attention, namely the breadth, stability, distribution and metastasis.

2.2 The cause of the distraction

Usually, the reasons for the distraction caused by the voice assistant are mainly divided into two aspects: first, easy to cause memory burden; second, the technology in the voice recognition is not complete, making the user has command recognition error, resulting in the user attention slack. The plot of the distraction analysis is shown in Fig 2.

Fig 2. Distraction analysis

2.3 Attention mechanism

The Attention mechanism originates from the study of human vision. In cognitive science, people tend to pay attention to only one part of the information, but ignore other visible information, which is due to the "bottleneck" of human processing of information, which is the Attention mechanism. Different parts of the human retina also process different information differently, with only the central retina being the most sensitive. In order to make full use of the limited visual information processing resources, people must select a specific area in the visual field and focus on it.

2.4 Voice technology approach based on the Attention mechanism

With the introduction of Attention mechanism into the field of voice interaction, the overall performance of intelligent voice technology is greatly improved by establishing the end-to-end voice recognition model based on Attention mechanism.

Compared with humans rapidly screening out valuable information from huge amounts of information in a limited time, attention models focus more on allowing task processing systems to find important information associated with the current output in the input data. Compared with traditional machine learning algorithms, attention patterns select the input subset in a structured way, thus reducing the dimensionality of the data and greatly improving the speed and accuracy of processing high-dimensional data. In 2018, Google released an end-to-end voice recognition model based on Attention mechanism. Based on seq2seq, combined with Attention mechanism, it realized the transition of voice sequence to text sequence with a single model. Its error rate is 5.6%, which is 1/18 of the conventional mode [7]. In 2019, scholars from Carnegie Mellon and Karlsruhe Institute of Technology proposed the in-depth Self-Attention mechanism technology, which tremendously improved the accuracy of the voice recognition system [8].

3 The STROKE Attention measurement experiment

3.1 The measurement of attention

Attention assessment is one of the important evaluation items of psychological assessment. The cancellation test is a psychological assessment method used for attention measurement [9]. The cancellation test is also known as STROKE. The cancellation test is a common test to compare the differences in the speed and accuracy of different individuals in complete the work. The experiment is easy to operate and the visualization can quantify the experimental results. In order to complete the cancellation task in the experiment, the subjects have to pay high attention and accurately and quickly identify the specified specific object in many similar objects.

The experimental methods and procedures are described as follows:

The experimental material is a table of randomly distributed numbers with 5*52 numbers, with 18 target numbers in each line. Set a number as the target. The subject is asked to check the number table line by line and to cancel the target number by stroking. The experiment ended when the subject thought the cancellation was complete. At the end of the experiment, the system records the time of the task, the number of strokes and the number of misstrokes. According to the formula, the cancellation accuracy and work efficiency of the subjects are calculated. Work efficiency, both time and accuracy, can be directly used to compare between different subjects.

The cancellation test provided in this experiment is a limited workload method, and the scholar G. M. Whipple proposed to use the work efficiency (E) as the index of the cancellation score.

In a limited manner:

$E=100\times A/T;$

$A=(c-w)/(c+o);$

Where T represents the time spent, A represents the accuracy, c represents the number of strokes, o represents the number of missed strokes, and w represents the number of misstrokes.

3.2 The design of experiment

Based on the measurement method of STROKE, this experiment will be divided into one experimental group and two control groups, in which the way of voice interaction is taken as the variable factor of the experiment. Among the variables, Siri, the most common daily object, was selected as the voice interaction object. The total number of participants was 150, 50 in the experimental group and 100 in control groups. The measurement content of the experimental group is the ordinary cancellation test. The number of participants in the control group I was 50, and the test content is to ask Siri simple questions according to the test arrangement, such as "what is 1 + 1 equal" and so on; The control group II was also 50, the test content requires asking Siri for more complex questions according to the experimental schedule while completing the cancellation test, and complete some dialogues, such as "recommend the nearby restaurant with the highest score to me".

The experimental site was selected as a closed and quiet room with closed windows without interference. The experimental contents and objectives of the

experimental group and control groups are shown in Tab 1.

The subjects selected adults with more than two years of experience in Mandarin as daily language, and used Mandarin as the language used in the experiment. The purpose of the experimental group is to measure the average completion time, work efficiency and task accuracy of normal adults. The goal of control group I is to measure the effect of voice interaction in trigger mode. The control group II aims to measure the effect of voice interaction in task mode.

Tab 1. The effect of voice interaction on attention

Group name	**Experiment content**	**Experimental objective**
Experimental group	Complete the cancellation test	Calculate the relevant data and its average in general
Control group I	Complete the cancellation test; Communicate with the voice assistant in trigger mode	Calculate the relevant data and the average of the trigger mode
Control group II	Complete the cancellation test; Communicate with the voice assistant in task mode	Calculate the relevant data and the average of the task mode

4 Analysis of research results

In the study of attention stability, the attention concentration index E measured in the cancellation test is used to measure the level of attention stability. E is proportional to the accuracy of the subjects in the cancellation test and inversely proportional to the time taken by the subjects to complete the test. One-way ANOVA analysis was analyzed for the attention stability of participants in experimental group, control group I and control group II, the results are shown in Tab 3. The exponential curve of attention concentration is shown in Fig 3.

Tab 2. Comparison of attention stability among three groups

	n	M	SD	SEM
Experimental group	20	1.3901	0.2393	0.057
Control group I	20	1.1050	0.1230	0.015
Control group II	20	1.0006	0.1116	0.012
Total	60	1.1652	0.2343	0.055

Tab 3. One-way ANOVA analysis of attention stability

	SS	df	MS	F	Sig
Inter-group	0.109	2	0.109	3.851	0.000
Intra-group	1.612	57	0.028		
Total	3.238	59			

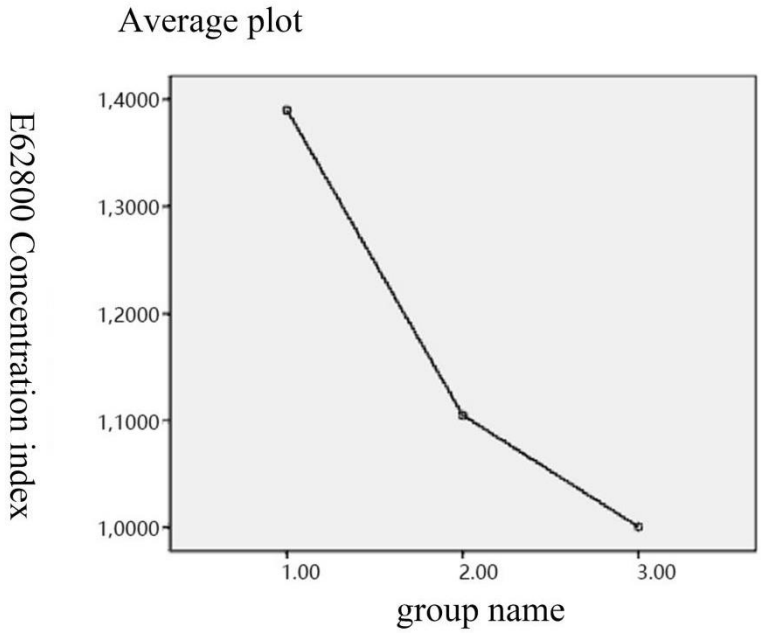

Fig 3. The exponential curve of attention concentration

One-way ANOVA of attention stability showed that the attention stability of users varies significantly between different voice interaction patterns ($p < 0.05$). LSD method was used to make multiple comparisons after the test, and there were significant differences between experimental group and control groups ($p < 0.01$). The difference between control group I and control group II was not significant ($p > 0.05$), as shown in Tab 4.

Tab 4. Multiple comparisons of attention stability in the three groups

（I）	（J）	（I-J）	Difference significant
Experimental group	Control group I	0.2852	0.000
	Control group II	0.3895	0.000
Control group I	Experimental group	-0.2852	0.000
	Control group II	0.1044	0.055
Control group II	Experimental group	-0.3895	0.000
	Control group I	-0.1044	0.055

It can be seen from the experimental results that the two modes of voice interaction have significant impacts on the stability of human attention, and the task

mode is more obvious than the trigger mode. The causes of distraction as mentioned above are caused by the large amount of information of the interaction process and the command recognition errors.

5 Conclusion

In this study, the voice interaction mode was taken as the experimental variable, and a control experiment was based on the influence of voice assistant on human attention. Based on the measurement method of STROKE in psychological assessment, the quantitative and visual experimental results were output. The following conclusions can be summarized on the design of voice interaction interface by analyzing the results of the cancellation test:

(i) The design and error correction of voice interaction interface.

The common reason for voice assistant recognition errors is the user's instinct to provide more information to drive the conversation process, so the designer needs to predict when the VUI will provide more information and when to prompt the user to provide more information. When the user provides more information and the recognition mechanism cannot identify, the user usually only needs to prompt through a short word to guide the answer needed by the recognition mechanism, without emphasizing the error, so as to avoid over distract the user's attention.

(ii) Voice interaction interface design needs to match the product service.

Voice products need to be familiar with the usage scenarios and the characteristics of target users, and more importantly, to reasonably match the voice interaction with the services of the products. From the perspective of service design reasonable planning voice interaction interface design, take the user as the center, effectively planning and organization services involved in the people, equipment, process and environment related factors, deepen the interaction mode, reduce user attention consumption, improve the user experience and service quality, avoid the waste of excess resources.

References

1. Gao Jingyang. Next battle: Man-machine Dialogue —— Dialogue Siri founder Norman[J]. Tsinghua Management Review, 2017, (7-8): 8-13.

2. Chen Xiaoliang. Why voice interaction iterates so quickly [J]. Science and Technology Herald, 2017, (3): 92.

3. Voice: a new revolution in human-computer interaction [EB/OL]. (2012-03-13)[2014-02-20]. http://www.leiphone.com/siri-ifly.html

4. DAVISK H. Automatic Recognition of Spoken Digits [J]. Journal of the Acoustical Society of America, 1952.24(6): 669.

5. Yuan Bin, Xiao Bo, Hou Yuhua, etc. Status quo and development trend of mobile intelligent terminal voice interaction technology [J]. ICT, 2014, (4): 39-43.

6. Li Sun, Cao Feng. End-to-end framework model analysis and Trend Research of Intelligent Voice Technology [J]. Computer Science, 2022,49 (S1).

7. WATANABE S, HORI T, KARITA S, et al. ESPnet: End-to-End Speech Processing Toolkit[C] //Interspeech 2018. 2018.

8. PHAM N Q, NGUYEN T S, NIEHUES J,et al. Very Deep SelfAttention

Networks for End-to-End Speech Recognition [J/OL]. 2019. https://arxiv. org/abs/1904. 13377.

9. MANCUSO M, ROSADONI S, CAPITANI D, et al. Italian standardization of the Apples Cancellation Test[J]. Neurol Sci, 2015, 36(7): 1233-1240.