

HUMAN FACTORS RESEARCH OF VOICE SEARCH ON MOBILE INTERNET

KEWEI YUAN

*School of Automation Beijing Univ. of Posts & Telecoms (BUPT)
Beijing China, 100876
yuankw@gmail.com*

WENJUN HOU

*School of Automation (BUPT) Beijing China, 100876
houwenjun2113@sohu.com*

AND

YUJIA ZHAO

*School of Telecoms Engineering, BUPT Beijing, China, 100876
zhaoyujia930@gmail.com*

Abstract. Voice search is a kind of promising search approach in the era of mobile internet. Currently, most of voice search services are still manual work, which are expected to be replaced by human-machine voice search based on voice recognition technology. Despite numerous disadvantages of human-human voice search, its interaction approach as natural language communication is obviously superior to human-machine voice interaction. In order to get better human-machine voice search interaction scheme, we conduct an investigation into current human-human voice search service with the purpose of understanding human habits and patterns in using voice search, elements that impact user experience, and getting enlightenment for designing automatic voice search interaction. Based on the findings, we continue to explore underlying rules and principles by building models which include two dimension model on search intention, state transition model and user experience hierarchical model. Finally, we propose a blueprint of automatic voice search interaction scheme according to investigation findings and models, hoping it could elicit better design.

1. Introduction

Voice search service, which takes advantages of talking as interaction approach, is widely used in daily life due to the convenience of interaction by

natural language. Along with the growing of mobile internet services, many companies such as Google and Baidu are putting forward their voice search service (Zhang, 2006). People can search for information simply by a call using mobile devices and telephone instead of desktop computer, which has significant meaning for people without computer or access to Internet in China.

The voice search services available nowadays are mainly within two categories which are manual service (human-human) and automatic service (human-computer). Currently, Google411 voice search developed by Google is automatic service using voice recognition technology, and it mainly provides search based on map and yellow pages, and also connection and reservation services. The voice search service of Baidu is informing the information found by baidu search engine to users by an operator. Additionally, Microsoft (Yuan, 2007) and many Chinese communication operation enterprises also putted forward their own voice search services, such as “114 know-it-all”[3] and “Voice Google”.

In this paper, first we try to find helpful information such as human habits and patterns in voice search through an exploratory investigation of current human-human voice search services. Then some models are built based on the findings of the study from different perspective, which are aimed at exploring the principle and typical voice search process and provide guidelines for designing voice search system and service. The final part is a blueprint of our voice search interaction design and related considerations.

2 An Investigation into Current Voice Search

Our investigation contains three parts. First, we design a series of tasks for test takers and record the whole conversation. From the conversation we hope to dig out human’s language habits and patterns. Then test takers as well as enough number of other people fill out related questionnaires who used voice search before. And we interviewed test takers and other users finally. We wish to find out user’s feelings and thoughts, and aptitude about information display and interaction. In this part, we describe our investigation in detail and represent out findings.

2.1 TASK DESIGN

In order to simulate the real circumstances, we weaved the tasks into a common story rather than a simple task list. Thus, the test takers may get involved easier and remember the tasks rather than simply read them out (hardly happen in real life). The story involves most aspects of information requested frequently in daily voice search such as restaurant information and

entertainment information. And different difficulty levels of tasks can help to discover the response of users in various circumstances.

The story is like this: one of your old friends is coming to Beijing to attend a training course in a place A you don't know. He plans to visit you after the training. You are a student of Beijing University of Posts and Telecommunications (BUPT) living in campus. And you are required to firstly find out how to travel from A to BUPT for him, secondly book a room in a hotel near BUPT from which he can get to the airport conveniently. Thirdly, he plans to visit Tiananmen, so that you should find a Peking style restaurant near Tiananmen Square for the lunch that day. Fourthly, you would like to have fun at night, so you have to find entertainment information.

The first task is at easiest level. *96166 Li Suli hotline* is very convenient to find it out. The second task is at hardest level because it is hard to determine convenience degree of reaching the airport from a particular hotel. We expected to observe the response to failure in getting results in this task. The third and the fourth tasks are at medium level, which involves numerous results and various aspects of information such as names, prices and locations. The flexibility of the third and fourth tasks is helpful to observe communication in flexible situation.

2.2 FINDINGS

Five test takers participated in this test using 96166 Li Suli hotline and 12580. The complexity of tasks elicited responses and interactions in various unexpected situation. Here we summarize some most important aspects:

1. Failure in getting results that meet the demands proposed by user. In this situation, users usually choose to change demands to less restrictive ones or change keywords. For example, test takers who failed to find out hotel near BUPT and convenient for getting to airport, usually change to find hotels near BUPT only.
2. Too many results are found. In this situation, users will either add more conditions to sift results, or choose a number of results arbitrarily. For example, if there are too many restaurants in list, users may ask more about prices and locations and get the most suitable ones.
3. Information needed is incomplete. This type means that users need the location, telephone number and prices of a particular restaurant, but there is no price information stored in the system. In this situation, users will act according to whether the missing information is crucial. If yes, they will response like in situation 1. If no, they may accept the incomplete results reluctantly.

4. Too complicated for voice search service. If the task is too complicated for voice search to finish, users usually choose to refer to other approaches such as Internet, and asking friends.

The investigation in this part only represents important findings, which enlighten the design of human-machine voice search system. However the enlightenment is not clear yet. Thus we also explore deeper continually in the following parts.

3 Models in Voice Search

The previous parts discussed our investigation and associated findings concerning manual voice search, which has important meaning to the design of automatic voice search. So in this part, we continue to discuss principles and models about automatic voice search service flow and user experience based on the investigation.

3.1. TWO DIMENSION MODEL ON SEARCH INTENTION

In order to find universal search patterns in voice search, we examine every conversation and discovered that the conversations have similar structure in general.

First take a look at part of a conversation in testing:

Test taker: Is there any restaurant near Tiananmen Square with Peking style?

Operator: Near Tiananmen, and having Peking style?

Test taker: yes.

Operator: Hello, the restaurants with Peking style near Tiananmen Square are Quanjude Roast Duck, Jiumen snacksWhich one would you like?

Test taker: En, Quanjude is ok. Could you tell me the address and telephone number of Quanjude?

Operator: The address is, and the telephone number is

This kind of dialogue happens nearly in every conversation of every test takers. They all first indicate the conditions that results must meet which we call constraints. Then they ask for detail information (address, telephone number, prices and etc) of targeted results they picked from all the results presented previously. They had already been familiar with tasks before search, which means that whether they realized or not, they have already settled intention in two dimensions-----constraints and details needed. They know which attributes should be constraints and which ones are the details. They all happen to deliver constraints to get abundant results, then target on some ones, and ask for details of the targets. In sum, two dimension model on search intention has meaning in two aspects. Firstly, users formed two dimension search intention; secondly, the search process is divided into two phases accordingly, which is consistent with the discovery of Robert G.

Capra III and Manuel A. Perez-Quinones in their research on mobile refinding of web information using a voice interface (Capra et.al., 2005).

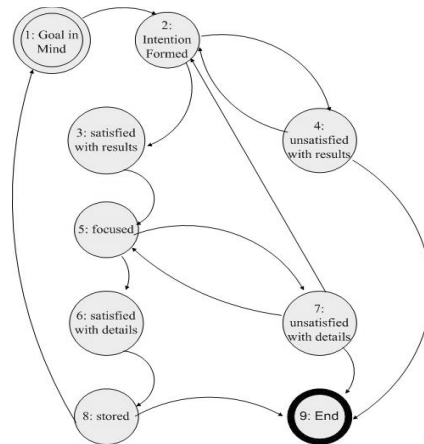


Figure 1. User States Transition.

3.2 STATE TRANSITION MODEL

User States transition

The first two stages of Norman's model of interaction (Norman, 1986) are establishing the goal and forming the intention. It is also what happens in voice search. Thus, state 1 is that users formed search goals and requirements. In this state, users realized that they need to turn to voice search for some information but having no clear intention and action in mind. State 2 is that users formed clear search intention both in constraints dimension and details dimension.

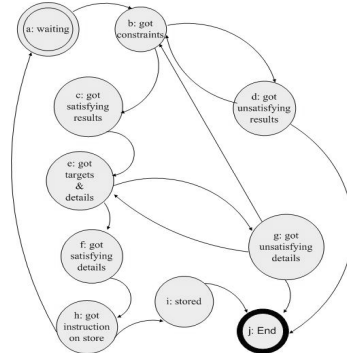


Figure 2. System States Transition.

In state 2, users are ready to use voice search to find results within constraints. After system feed back, they may get into two states: state3-----got satisfied results, state 4-----got unsatisfied results. Here satisfaction means differently according to different user and tasks. But in general, it contains two aspects which are meet constraints and enough number of results.

System States Transition

Similar to user states transition, we can analyze the states and transitions as following: state a-----initialized and waiting for input. When user input constraints, system enters into state b-----got constraints. Based on the results, system will also get into two states: state c-----get satisfied results and state d-----get unsatisfied results, which are corresponding to state 3 and 4 of users. If in state d, system may get into state b due to user's adjustment on constraints, or end state because of termination by user. If in state c, users will continue to focus on targets and details and system will enter state e-----get instructions on targets and details. After searching for details, system may again enter two states: state f-----get satisfied details and state g-----get unsatisfied details corresponding to state 6 and 7 of users. Again, if in state g, system may go back to state b due to user's adjustment on constraints, or go to state e because of refocusing, or enter end state. If in state f, users will give instruction on storing, and system got it entering state h-----get instructions on storing. Upon finishing storing, system enters state i. Then, system enters state a if user begins another inquiry, or go to end state. The transition is showed in figure 3.

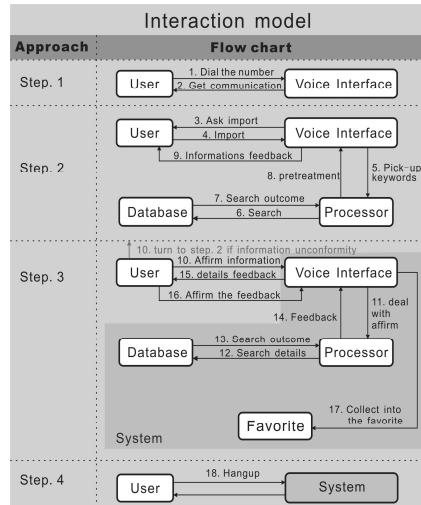


Figure 3. Voice search interaction scheme.

3.3 USER EXPERIENCE HIERARCHICAL MODEL

The Element of User Experience discussed by Jesse James Garrett and Robert Robinoff's theory on how to quantify the user experience explore the important factors that influence user experience in using web. There are also particular elements of user experience within voice search scope. In the investigation, especially from questionnaires and interviews, we get the following key elements of user experience and classify them into three layers: 1. System layer elements 2. Cognition layer elements 3. Emotion layer elements.

4. Voice Search Interaction Scheme

Working as an interface to mobile internet, the design of voice search should consider the features of cell phone and other mobile devices and related network resources. The feature of cell phone is complex. The basic functions are making call and sending message.

In step 1, user initiate connection and get connected with service. In step 2, system prompts user to input constraints and pick up keywords by analyzing voice input. System searches for results according to keywords. Before feeding back results to user, the system must pretreat results in order

to organize information presentation. For example, if there is no result, the system should analyze reason and promote alternative means to handle it such as loosing some constraints. And if there are too many results, system should count the number of items, try to classify them and make recommendations. The representation of results should involve both voice and visual display in screen. In step 3, after feed back results, user may focus on targets. Voice input is not absolutely appropriate in this focusing process because selecting by touch the screen or keyboard is more time-saving and convenient. The search process and pretreatment of results are like those in step 2 expect the collection part. Because the information in voice search is usually in daily use, thus like web collection, voice search results collection may also help people to refind them. Favorite, web address book and other functions can be used here. Also, the storage of results can be short message, voice inbox, and etc.

At the end of search, storage can not be neglected. First, storage deals with keeping information retrieved this time such as sending message to user. Secondly, storage should take refinding into consideration. To obtain information again from that they had searched for before is very likely to happen according to our investigation, which is consist with the result of Robert G. Capra III and Manuel A. Perez-Quinones (2005) in their research on mobile refinding of web information using a voice interface. Thus, how to organize and keep search results for future use, and how to realize refinding function still needs further discussion.

Reference

- Capra, R.G. III, Pérez-Quinones, M. A.: 2005, Mobile Refinding of Web Information Using a Voice Interface: An Exploratory Study, in *Proceedings of the 2005 Latin American conference on Human-computer interaction*, ACM International Conference Proceeding Series, 12, pp.88-99.
- Ma, J.: 2007, "114 know-it-all" leading the voice search, C-enterprise Management.
- Norman, D. A.: 1986, *User Centered System Design: New Perspectives on Human-Computer Interaction*.
- Yuan, M.: 2007, Microsoft buy Tellme for confronting with Google's mobile search, IT Times.
- Zhang, B.: 2006, *Time of Voice Search Coming*, China New Telecommunications.