



Application of Logistic Regression Analysis of Smartphone Speech Interaction Usage in China: A Questionnaire-Based Study of 622 Adults

Wen-jun Hou^{1,2} and Xiao-lin Chen¹(✉)

¹ School of Digital Media and Design Arts,
Beijing University of Post and Telecommunications, Beijing 100876, China
cx195@163.com

² Beijing Key Laboratory of Network Systems and Network Culture,
Beijing University of Post and Telecommunications, Beijing 100876, China

Abstract. Speech interaction as one of the contact-free input techniques has been applied in mobile devices (e.g. smartphone) for many years, which means that a considerable number of users were exposed to speech interaction. In China, speech interaction, while undoubtedly natural, has also entered users' life for more than 5 years but it is still perceived as "not that good". Curiosities are stimulated that what barriers are that prevent speech from becoming one of the mainstream interaction modality in China, yet there is no research on the user experience of mobile speech interaction. An online questionnaire was used to measure participants' speech interaction use of smartphone. Simple Logistic Regression Model and Ordinal Logistic Regression Model were used as the primary method of data analysis. This study concluded that speech interaction is an interactive modality for the future and need to give full play to the advantages of speech interaction by reducing interface exclusivity, offering services to users actively in combination with situational awareness and guiding users in a variety of modalities.

Keywords: Speech interaction · Questionnaire-based
Logistic regression analysis

1 Introduction

The past decade has witnessed dramatic changes in the input and output modalities between human and computer due to the ubiquity of pervasive and mobile computing and broadband Internet. We have seen the big opportunities happened with touch for mobile computing, which presented us that naturalness is a central theme for interaction. These kinds of developments lead people to pursue the more natural forms of communication and information processing that human possess such as speech, gestures even thoughts.

Speech interaction, communicating through speech and language, once hit a bottleneck due to the fact that speech is the highest-bandwidth two-way communication

channel human have [1]. Improving machines' ability to understand speech and natural language used to be difficult until the advances in speech processing and machine learning recently. People now are facing many more situations requiring hands-free or eyes-free interaction like driving. Speech interaction should be treated as more than an alternative to "traditional" input or output mechanisms.

Speech interaction is not far away from people's life. It has been applied to mobile devices (e.g. smartphone or smartwatch), which means that a considerable number of users can be exposed to speech interaction. Mobile devices, especially smartphone, have become an essential part of modern life, bringing huge benefits in terms of living and working flexibly. According to Zenith's Mobile Advertising Forecasts 2017, in 2018, 66% of individuals in 52 key countries will own a smartphone, and China as the country with the highest number of smartphone users will have 1.3 billion users. In China, speech interaction, while undoubtedly natural and reachable, has also been brought into users' life for more than 5 years but it is still perceived as "not that good". Curiosities are stimulated that what barriers are that prevent speech from becoming one of the mainstream interaction modality in China, yet there is no research on the user experience of mobile speech interaction.

There is a little user research on evaluating the recent applications of speech interaction. A pre-research was conducted to get an overview of the usage of speech interaction in China. This research observed that nearly half of the 196 respondents gave up using speech interaction after a tryout. The identification of factors for this result still needs to be further studied. Various methods can be applied to uncover the factors of this result, like user interview. Among them, logistic regression analysis is considered, in view of the fact that the response variables can be considered as binary (use or not use) or ordinal (5-point satisfaction scale).

Logistic Regression also known as Generalized Linear Models, is in the category of statistical models. Logistic Regression has the goal of finding the best fitting model to describe the relationship between explanatory variables and response variable, which can help to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these [2]. There are two kinds of Logistic Regression models for binary data, Simple Logistic Regression Model and Multiple Logistic Regression Model. The First type of these models involves modeling a relationship between one explanatory variable and the binary response variable. While the second type can be used to model k explanatory variables each with m levels. The application of Logistic Regression is also extended to the case where the dependent variable is in form of ordered categorical responses, which is also known as Ordinal Logistic Regression Model [3]. A few studies have been done using ordinal logistic regression model (OLR) to identify the predictors of child undernutrition [4] and in many epidemiological and medical studies, OLR model is frequently used when the response variable is ordinal in nature [5–10]. Ordinal logistic regression models also were applied in quality of life studies because procedures such as either dichotomization or misinformation on the distribution of the outcome variable may complicate the inferential process [8].

According to the dependent variable form in our study, Simple Logistic Regression Model and Ordinal Logistic Regression Model were used as the primary method of data analysis.

2 Method

2.1 Procedures

This study mainly adopted the methods of interviews and questionnaires. The entire research lasted for three months and was divided into two phases. In phase 1, pre-research, based on small sample interviews and questionnaires, was conducted to get an overview of the usage of the mobile speech interaction in China. In the phase 2, a well-design online questionnaire modified based on the results of the first phase were used for further study. In every phase, participants were notified what the purpose of the study was and that only the comprehensive statistical data of the survey were used and no personal data of the respondents was involved in the study.

2.2 Materials

In the first phase, the smartphone users were divided into three types: *never use speech interaction*, *gave up after tryout*, *use speech interaction*. We interviewed about 2 users for each type to understand their user experience or perspectives of the speech interaction, which helped to design the first version of the questionnaire. The questionnaire survey was conducted in a small range to establish a baseline about the usage of mobile speech interaction.

In the phase 2, the questionnaire, the second version, was used to measure participants' speech interaction use of smartphone, the functions for which they used speech to interact with the phone, the overall satisfaction and pain points of use. Three parts were included: the demographic information questionnaire, the mobile phone function questionnaire and the usage experience questionnaire of the speech interaction.

The demographic information questionnaire consisted of 5 questions about age, gender, education level, occupation and present location, which is designed to prove the validity of sampling.

The mobile phone function questionnaire measured the different uses people have for mobile phone and the functions people use through speech interaction, taking into account the usage frequency of these functions. This part could collect data to compare the common mobile functions with functions used by speech interaction to know which functions users are more willing to use by speech interaction, and which functions users are more willing to use by traditional interaction methods.

The usage experience questionnaire designed to divide users into four categories and get to know their experience specifically. According to the extent of the user's understanding of the speech interaction, the participants were classified into four categories: *never heard of speech interaction*; *heard of it but never use it*; *gave up after tryout*; *still use speech interaction*. Questions include items such as "Why give up using speech interaction?" and we provided some alternative answers which we got from prior user interviews. Respondents are required to rate the answers they chose on a 5-point scale where 1 = slight impact and 5 = severe impact. The primary method of data analysis was logistic regression analysis.

2.3 Participants

A total of 666 respondents completed the questionnaires and 622 questionnaires were valid. Recruitment was via links placed on China-based social networking app called WeChat. The age range of the respondents was 18 to 50 years old and above with 51% 18–30 years old and 37% 31–50 years old. 57% respondents were women. 84% respondents have a bachelor degree or above.

2.4 Data Analysis Methods

The primary method of data analysis was correlational analysis. We aimed to explore the relationship between the pain points of usage and the overall satisfaction of the participants who have used speech interaction. For the participants who never used speech interaction, we tried to explore the relationship between the reasons why they didn't use speech interaction and whether they would try to use it in the future.

3 Results and Discussion

3.1 Function Analysis

In the mobile phone function questionnaire, all participants were required to choose the functions they might use in daily life by touch in the 25 provided functions and the participants who used speech interaction were asked to choose the used functions in the 24 provided functions. They were also asked to give the frequency of the functions they chose by 5-point scale (1 for “rarely”, 2 for “less than once a week”, 3 for “once a week”, 4 for “once for 2 or 3 days”, and 5 for “everyday”).

According to the users' distribution of frequency for each function they might use by touch, functions can be divided into 4 types (Fig. 1¹). The first kind is the most frequently and commonly used functions of smartphone like making a call, which were used by most mobile users. The second kind of functions are the functions that were closely related to the usage scenarios, such as driving navigation or using camera. The distributions of these functions were more dispersed comparing with the first class due to the close relationship with users' living status and habits. The third and fourth kinds of functions could be counted as niche applications compared with the first two class. The total number of users were relatively small, so the distribution of the actual users was mainly concerned. The third kind of functions was characterized by the high proportion of users who used every day, such as Stocks or Sports. The fourth category functions had relatively scattered frequency distribution, which means they might be relevant with some particular scenarios like finding a restaurant or using the flashlight.

The number of people using speech interaction was much smaller, and the frequency of use was also generally low, which was in accordance with the initial hypothesis we suggested that speech interaction was not one of the mainstream

¹ All the data in this paper can be viewed at the following web site:
<http://blog.chenxiaolintryme.com/HCI2018-Paper/>.

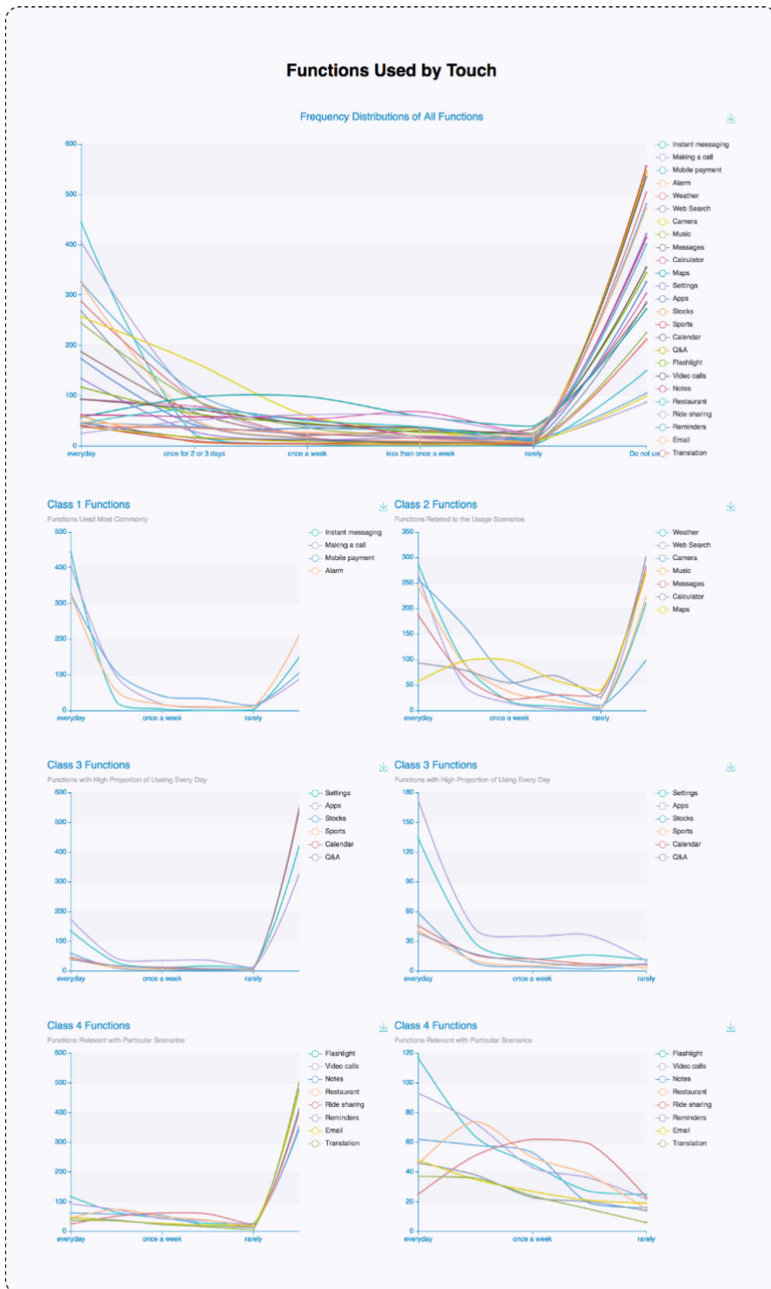


Fig. 1. Frequency distribution of functions used by touch

interaction modalities in China for mobile. Functions that less than 30 users chose weren't included in the scope of consideration because the small sample wasn't representative. Same as the classification of the functions used by touch, the functions used by speech interaction were divided into two types (Fig. 2). Type 1 was the functions users tended to use with high frequency, while type 2 was the functions with distributed frequency distributions.

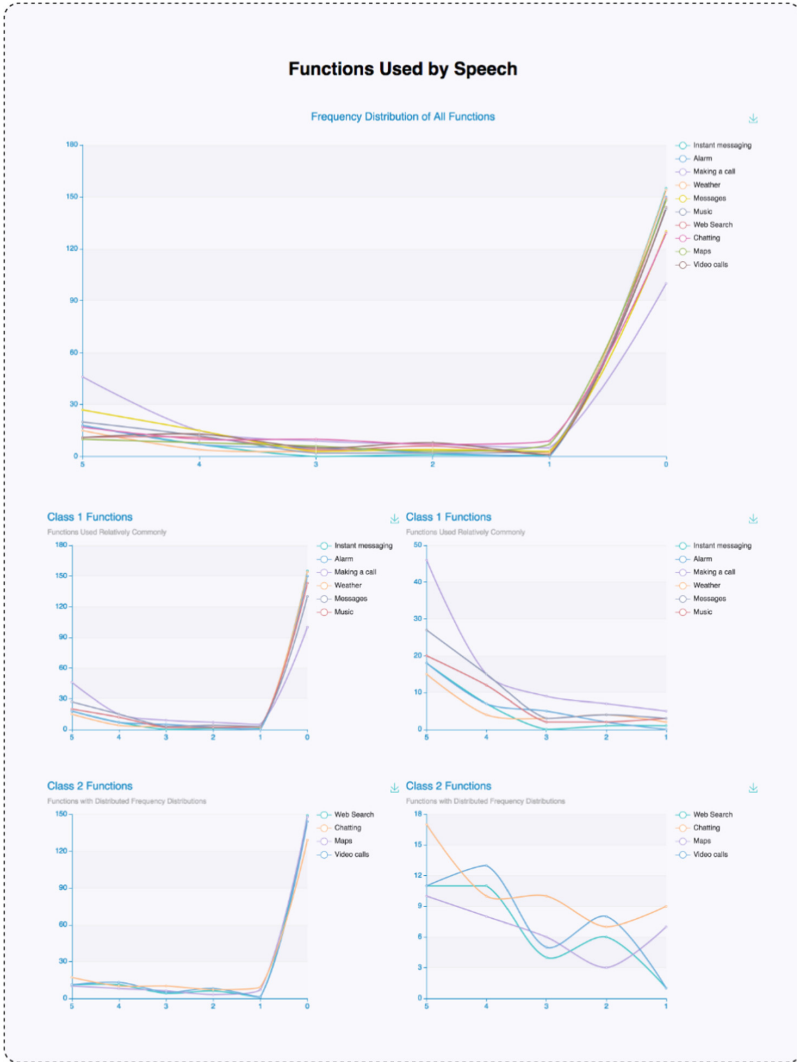


Fig. 2. Frequency distribution of functions used by speech

Compared with the results of the two interaction modalities, some interesting findings could be got:

- The functions with high frequency usage in speech interaction were generally the functions with high frequency in touch interaction.
- Users preferred to use speech interaction in functions with complex input and simple output.
- Some functions might have a lot of space to develop in speech interaction, such as instant messaging, photographing, changing phone settings and so on.

In a speech interaction task design, the usage scenarios need to be well designed as far as possible to enlarge the advantage of the speech interaction so that speech interaction could input a large amount of information at once which have to input by multiple steps by the traditional interaction mode.

3.2 User Experience Analysis

The usage experience questionnaire designed to divide users into four categories and get to know their experience specifically. According to the extent of the user's understanding of the speech interaction, the participants were classified into four categories: never heard of speech interaction; heard of it but never use it; gave up after tryout; still use speech interaction. Questions include items such as "Why give up using speech interaction?" and we provided some alternative answers which we got from prior user interviews. Respondents are required to rate the answers they chose on a 5-point scale where 1 = slight impact and 5 = severe impact. As shown in Fig. 3, of the 622 respondents, 93 were classed as Type 1 (Never heard of the speech interaction); 224 were classed as Type 2 (Heard of it but never use it); 113 were classed as Type 3 (Gave up after tryout); 182 were classed as Type 4 (Use speech interaction). These four types of users actually represented four types of users who had different degree

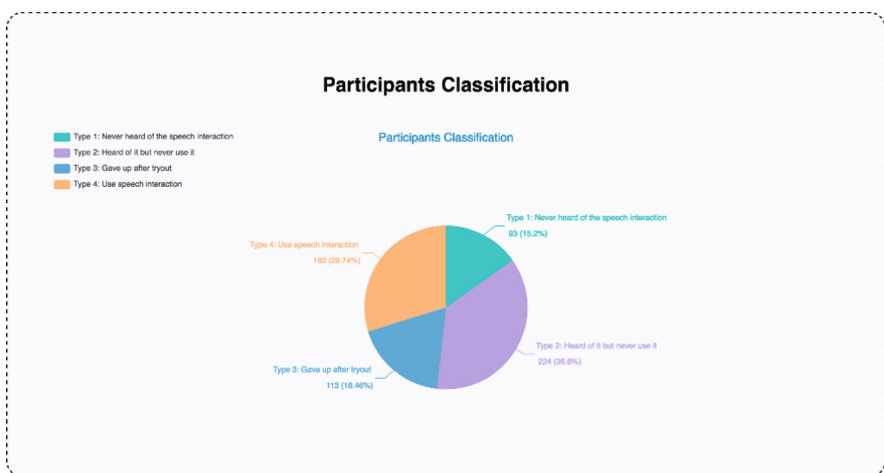


Fig. 3. Participants classification

understanding of the speech interaction. By analyzing the perspectives of these four groups users on speech interaction, which factors impeded the users using speech interaction can be preliminarily judged. For the users who never used speech interaction, the reasons for preventing the usage of speech interaction were tried to be found out. For the users who had used the speech interaction, the reasons that affect the subjective satisfaction of the speech interaction were more concerned.

Type 1: Never heard of the speech interaction. Of the 93 respondents of Type 1, 52 would try to use the speech interaction in the future. A logistic model was fitted to the data to test the research hypothesis regarding the relationship between the reason they didn't use speech interaction and whether they would try to use it in the future. The logistic regression analysis was carried out by the Logistic procedure in SPSS. The result showed that

$$\text{Predicted logit of (Future use)} = 0.234 + (-0.526) * \text{type1r1} + 0.815 * \text{type1r2} \quad (1)$$

According to the model, the log of the odds of the future speech interaction usage of a user who never heard of speech interaction was negatively related to reason 1, "I never thought of using speech interaction" ($p < .05$), and positively related to reason 2, "I don't know I can use speech to operate my phone" ($p < .05$; Table 2). In other words, people who chose reason 2 with high score are more likely to use speech interaction in the future and people who chose reason 1 won't (Table 1).

Table 1. Variables in the equation for Type 1

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	type1r2	1.046	.342	9.366	1	.002	2.845
	Constant	-.406	.276	2.156	1	.142	.667
Step 2 ^b	type1r1	-.526	.248	4.504	1	.034	.591
	type1r2	.815	.303	7.260	1	.007	2.260
	Constant	.234	.379	.379	1	.538	1.263

a. Variable(s) entered on step 1: type1r2

b. Variable(s) entered on step 2: type1r1

Table 2. Omnibus tests of model coefficients for Type 1

		Chi-square	df	Sig.
Step 1	Step	15.734	1	.000
	Block	15.734	1	.000
	Model	15.734	1	.000
Step 2	Step	5.566	1	.018
	Block	21.299	2	.000
	Model	21.299	2	.000

Table 3 showed the usual significance test for the logistic model based on the log-likelihood chi-square test. It therefore led to the conclusion that the linear relationship between the explanatory variables and the Logit P is significant ($P = .000 < .05$) and the model was reasonable.

Table 3. Hosmer and Lemeshow test for Type 1

Steps	Chi-square	df	Sig.
Step 1	.114	1	.736
Step 2	2.214	4	.696

The statistical significance of individual regression coefficients is tested using the Wald chi-square statistic (Table 2). According to Table 2, both reason1 (type1r1) and reason 2 (type1r2) were significant predictors of the future usage of speech interaction for the people who never heard of it ($p < .05$). The test of the constant merely suggests whether an intercept should be included in the model. For the present data set, the test result ($p > .05$) suggested that an alternative model without the intercept might be applied to the data.

Goodness-of-fit statistics assess the fit of a logistic model against actual outcomes. The Hosmer–Lemeshow (H–L) test is an inferential goodness-of-fit test that yielded a $\chi^2(4)$ of 2.214 and was insignificant ($p > .05$; Table 4), suggesting that the model was fit to the data well. In other words, the null hypothesis of a good model fit to data was tenable.

Table 4. Parameter estimates for Type 3

		Estimate	Std. error	Wald	df	Sig.	95% Confidence interval	
							Lower bound	Upper bound
Threshold	[type3satisfy = 1]	-1.030	.267	14.921	1	.000	-1.553	-.507
	[type3satisfy = 2]	.051	.259	.038	1	.845	-.457	.558
	[type3satisfy = 3]	3.169	.472	45.174	1	.000	2.245	4.093
	[type3satisfy = 4]	5.172	1.043	24.598	1	.000	3.128	7.216
Location	type3r1	.102	.065	2.494	1	.114	-.025	.229
	type3r2	-.235	.096	5.969	1	.015	-.424	-.047
	type3r3	.251	.075	11.320	1	.001	.105	.397
	type3r6	.142	.078	3.328	1	.068	-.011	.295
	type3r7	-.070	.096	.530	1	.467	-.258	.118
	type3r8	.075	.078	.914	1	.339	-.078	.227
	type3r9	-.125	.091	1.865	1	.172	-.304	.054
	type3r10	-.044	.075	.349	1	.555	-.191	.102
	type3r12	-.164	.085	3.752	1	.053	-.331	.002
	type3r14	-.031	.091	.120	1	.729	-.209	.146

Link function: Negative log-log

Speech interaction was still a very strange way of interaction for some users, which hadn't caused much impact on their lives. It can reflect that speech interaction is highly substitutable and the usage of speech interaction lacks rigid demand. Only by greatly improving the experience of the original interaction, can users try to use speech interaction, which is the first step of speech interaction into people's life.

Type 2: Heard of speech interaction but never use it. Of the 224 respondents of Type 2, 143 would try to use the speech interaction in the future, and 176 chose that they never use speech interaction. A logistic model was also fitted to the data to test the research hypothesis regarding the relationship between the reason they didn't use speech interaction and whether they would try to use it in the future. However, none of the reasons we provided determined to be significant relevant to whether people would use speech interaction. Only the result of descriptive statistics was focused on.

Figure 4 showed that the main reason for the most users who don't use speech interaction was that they were used to existing operation modes. It was found that many mobile users have been very skilled in touch-screen interaction by user interviews. Hence, using speech interaction may be a fresh attempt for them. If the usage of speech interaction requires a relatively high cost of learning, users will not use speech interaction. For those who are not very skilled in touch-screen interaction, such as older mobile users or younger mobile users, voice interaction may be more promising. For older mobile users, the decline in various physical functions such as vision leads to the possibility that they may be more likely to rely on speech interaction. As for the younger mobile users, they may have begun to use speech interaction on mobile when

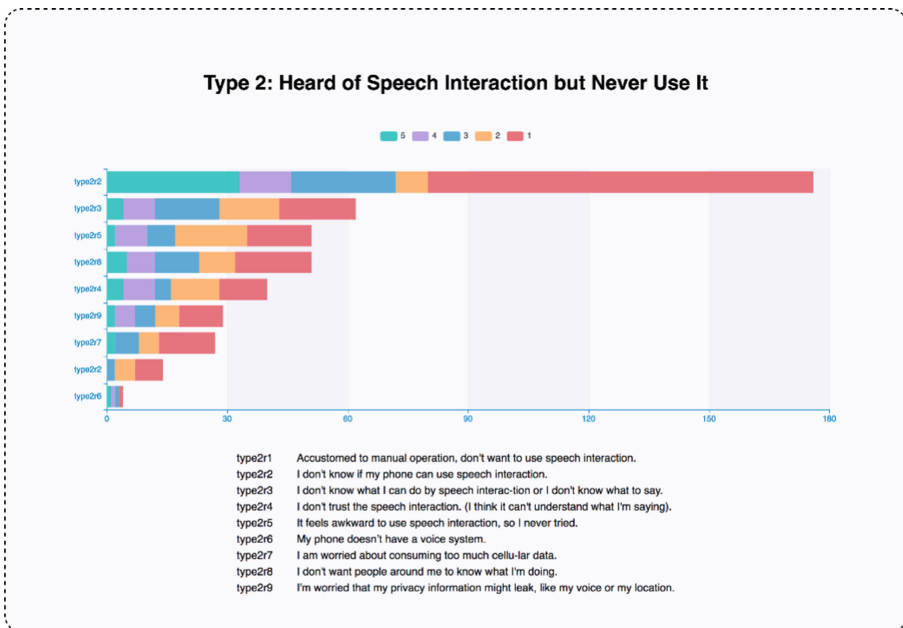


Fig. 4. Importance distribution of provided reasons for type 2 users

they have the first mobile phone. For them, speech interaction is just one of the interaction methods mobile has. In this case, the learning cost and operation cost of the speech interaction are not higher than the touch-screen interaction, and the performance of the speech interaction is even better in some functions. Speech interaction is an interaction form for the future, and the next generation of mobile phone users may greatly contribute to the acceptance of speech interaction.

Type 3: Gave up using speech interaction after tryout. Of the 113 respondents of Type 3, an ordinary logistic model was fitted to the data to test the research hypothesis regarding the relationship between the pain points of usage and the overall satisfaction. The logistic regression analysis was carried out by the Logistic procedure in SPSS. 17 alternative answers were provided to the participants. Descriptive statistics were used to analyze the users' selections of those answers (Fig. 5). From the results of frequency statistics alone, there were two reasons why users are most likely to give up the use of speech interaction on a mobile phone. The first reason was that when users

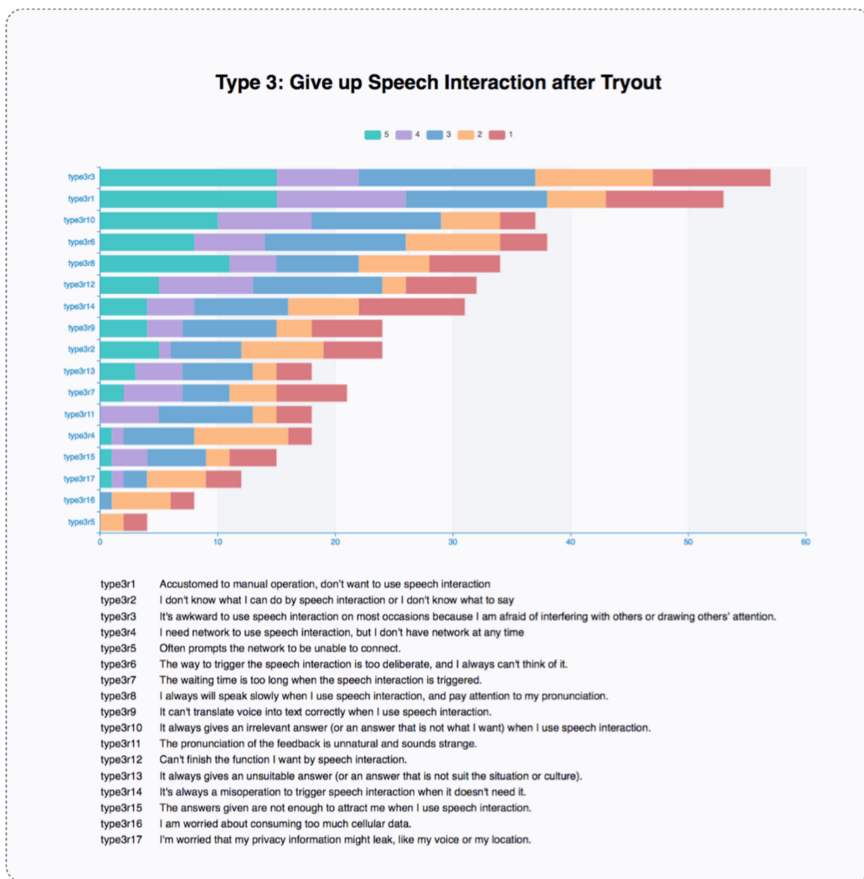


Fig. 5. Importance distribution of provided reasons for type 3 users

used speech interaction in non-private places, they might be worried about affecting others or arousing others' attention. The psychological pressure brought by this situation might lead to a higher expectation of speech interaction. The second reason was that users who had been accustomed to manual operation felt no need to use speech interaction. It is mainly for mobile phone poisoning users who were also the most likely to give up after trial speech interaction. They might have an open mind to try out new functions in the process of mobile development. However, the requirements for new interaction modalities would be relatively high. If there was no obvious improvement in user experience in some aspects, it is difficult for them to give up the skilled interaction form to migrate.

The analysis results showed that a part of the answers was not common, which meant that not all of the reasons for the majority users is significant. Consequently, only the reasons that had more than 20% users choose were interpreted as the explanatory variables in logistic regression which was mainly to explore whether there was a clear correlation between the various reasons and the user's usage of satisfaction. Users look for a explanation of their behavior, associating either internal attributes or external attributes. Internal attribution refers to the fact that users do not use speech interaction for some reason, but they think these reasons are personal ability defects or habits rather than the defect of speech interaction. External attribution refers to that users will ascribe their pain points in the use process to the shortcomings of speech interaction itself, which will directly affect users' satisfaction with speech interaction. The estimate in Table 4 was the reflection of internal and external attribution, which means that positive values represent external attributions and negative values represent internal attributions.

Table 4 showed that reason 2, "I don't know what I can do with the voice system or I don't know what to say" and reason 12, "I cannot perform tasks that I want to do" that had negative coefficients, were more likely to cause a decline in user overall satisfaction based on the 0.1 significance level. These two reasons showed that users had relatively high demands for the guidance of speech interaction or the guidance of speech interactive on mobile wasn't good enough and couldn't help users understand the function range of speech interaction, which resulted in the user's expectation of speech interaction was not consistent with the actual usage. It was easy to cause the decline of the overall satisfaction of speech interaction. As for Reason 3, "It's awkward to use speech interaction on most occasions because I am afraid of interfering with others or drawing others' attention" and reason 6, "The way to trigger the speech interaction is too deliberate, and I always can't think of it" were also the reason why users don't use speech interaction, but these two reasons were considered as internal attributions. These two reasons put forward the difficulties that should be overcome in the further design of speech interaction, which enabled speech interaction to enter the life of users more naturally by combining context awareness which could provide active service for users.

The following tables are the evaluations of the Logistic Regression Model.

Table 5 showed that the model is shown to a significant difference between the final model with all predictors included and the model with only the intercept fitted ($p < .005$). It therefore leads to the conclusion that the fitted model gives better

Table 5. Model-fitting information for Type 3

Model	-2 Log-Likelihood	Chi-Square	df	Sig.
Intercept only	233.212			
Final	205.403	27.81	10	.002

Link function: Negative log-log

predictions than if interpretations were based on marginal probabilities of the categories of the response variables which is satisfaction.

Table 6 aimed at testing whether the observed data were inconsistent with the fitted model. Table 6 showed that large significant values compare to the significance level of 0.05 – a result that could have led to the conclusion that the fit of the model is good, had it not been for the limitations that the goodness of fit of the model cannot be ascertained due to the large number empty cells.

Table 6. Goodness-of-fit for Type 3

	Chi-Square	df	Sig.
Intercept only	361.16	410	.960
Final	201.008	410	1.000

Link function: Negative log-log

Large R^2 values (i.e. closer to 1) indicate that more of the variation is explained by the model. However, the Pseudo R^2 values shown in Table 7, although not too small considering the inclusion of the five interval scale variables in the fitted model, gave some reason for revision of the model in order to generate better predictions.

Table 7. Pseudo R^2 values for Type 3

Cox and Snell	.218
Nagalkerke	.248
McFadden	0.117

Link function:

Negative log-log

The large significant value of the test of parallel lines indicated that the null hypothesis is to be accepted, which is all slope coefficients, also known as location parameters, are the same across the categories of the satisfaction (Table 8).

In general, this model had a certain reference value, but there was still room for revision of the model in order to generate better predictions.

Type 4: Use speech interaction. Of the 182 respondents of Type 4, an ordinary logistic model was fitted to the data to test the research hypothesis regarding the relationship between the pain points of usage and the overall satisfaction. The logistic

Table 8. The test of parallel lines for Type 3^a

Model	-2 Log-Likelihood	Chi-Square	df	Sig.
Null hypothesis	205.403			
General	199.109 ^b	6.293 ^c	30	1.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories

- a. Link function: Negative log-log
 - b. The log-likelihood value cannot be further increased after maximum number of step-halving
 - c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model.
- Validity of the test is uncertain.

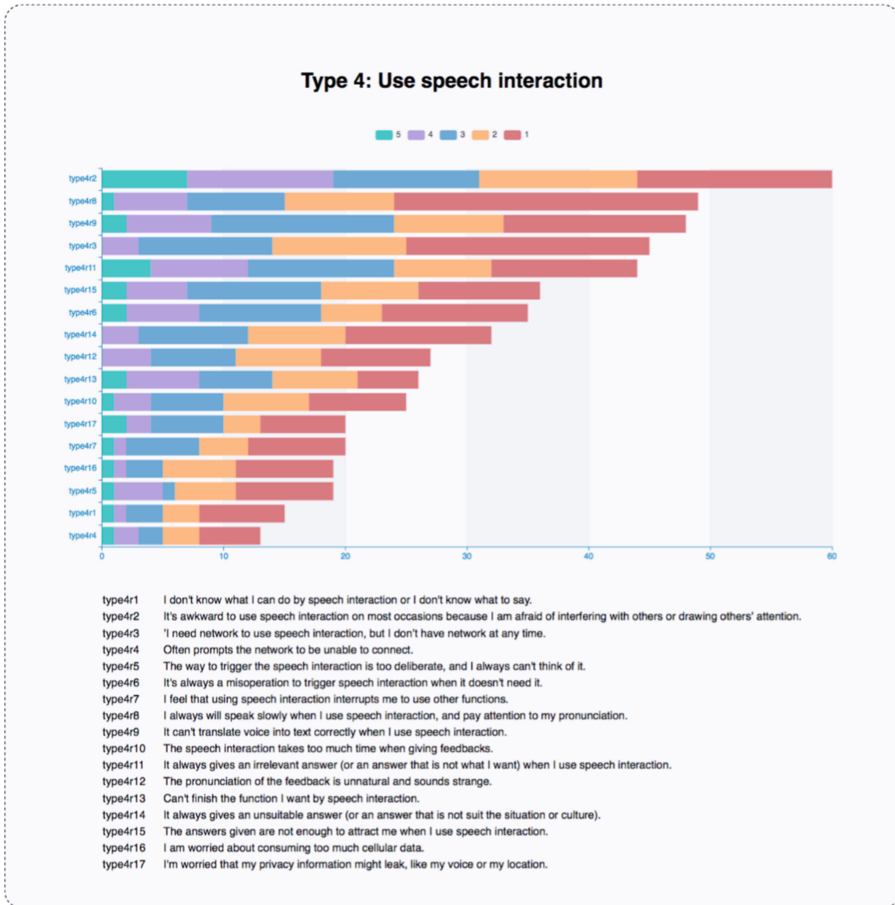


Fig. 6. Importance distribution of provided reasons for type 4 users

regression analysis was carried out by the Logistic procedure in SPSS. 17 alternative answers were provided to the participants. Descriptive statistics were used to analyze the users' selections of those answers (Fig. 6). From the results of frequency statistics alone, the pain point in the process was the same as the third type of users who chose to give up speech interaction, which was that when users used speech interaction in non-private places, they might be worried about affecting others or arousing others' attention. This showed that this was a common problem for all the users who had used speech interaction. There were four other reasons that also had a higher proportion. Reason 9, "It can't translate voice into text correctly when I use speech interaction" and reason 11, "It always gives an irrelevant answer (or an answer that is not what I want) when I use speech interaction" were the problems of validity. There is still a lot of space to improve for voice interaction in speech recognition and semantic recognition. Reason 8, "I always will speak slowly when I use speech interaction, and pay attention to my pronunciation", was a common phenomenon in the use process. Although many users selected this reason, it has little influence on users' actual usage. However, it can be seen from the side view that users have psychological burden when they use speech interaction and that is why they change their usual language habits to adapt to the shortcomings of speech interaction on speech recognition and semantic recognition. Reason 3, "I need network to use speech interaction, but I don't have network at any time", was a problem related to the social status and would disappear naturally with the popularization and the trade of low-price of mobile data network in China.

Similar to the analysis of the third type of users, the analysis results showed that a part of the answers was not common, which meant that not all of the reasons for the majority users is significant. So, only the reasons that had more than 20 users choose were interpreted as the explanatory variables in logistic regression which was mainly to explore whether there was a clear correlation between the various pain points and the user's usage of satisfaction. The estimate in Table 4 was still considered as the reflection of internal and external attribution, which means that positive values represent external attributions and negative values represent internal attributions.

Table 9 showed that the reason 10, "The speech interaction gives feedback too slowly", was more likely to lead to a decline in overall satisfaction of people who still use speech interaction based on the 0.1 significance level. In the current mobile speech interaction, most of the functions are performed by full screen, which means that users need to interrupt the ongoing tasks and switch to the speech interaction interface. Speech interaction is usually conducted by the form of dialogue. The feedbacks of the speech instructions are also speech. However, speech content is a slower form of feedback than visual content. This may lead users to feel that using speech interaction will take longer time to wait for feedback. From this point, two inspirations can be obtained:

- According to different functions, a non-full screen interaction form is used selectively to avoid affecting the users' ongoing tasks as much as possible. In this case, the speech interaction is used as an auxiliary interactive form to support the user's operation of parallel task flow. For example, when users read articles and want to play music at the same time, they can stay in the reading interface and perform speech instructions to play music.

Table 9. Parameter estimates for Type 4

		Estimate	Std. error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[type4satisfy = 1]	-6.349	1.983	10.250	1	.001	-10.235	-2.462
	[type4satisfy = 2]	-2.268	.474	22.943	1	.000	-3.196	-1.340
	[type4satisfy = 3]	.587	.218	7.243	1	.007	.160	1.015
	[type4satisfy = 4]	2.883	.659	19.123	1	.000	1.591	4.175
Location	type4r2	.099	.115	.742	1	.389	-.126	.324
	type4r3	.391	.151	6.695	1	.010	.095	.688
	type4r6	-.096	.157	.371	1	.543	-.404	.213
	type4r8	-.216	.165	1.728	1	.189	-.539	.106
	type4r9	-.027	.135	.039	1	.843	-.292	.238
	type4r10	-.342	.207	2.716	1	.099	-.749	.065
	type4r11	-.125	.157	.628	1	.428	-.433	.184
	type4r12	.066	.168	.154	1	.695	-.263	.395
	type4r13	.079	.150	.280	1	.597	-.214	.372
	type4r14	-.047	.183	.066	1	.797	-.407	.312
	type4r15	-.077	.162	.223	1	.637	-.395	.241

Link function: Cauchit

- When the full screen is needed for speech interaction, visual feedback is combined as much as possible to allow users to get feedback faster.

The following tables are the evaluations of the Logistic Regression Model. The analysis method is the same as the analysis of Type 3. To avoid repetition, it's not described in this paper. In general, this model had a certain reference value, but there was still a lot of room for revision of the model in order to generate better predictions (Tables 10, 11, 12 and 13).

Table 10. Model-fitting information for Type 4

Model	-2 Log-Likelihood	Chi-Square	df	Sig.
Intercept only	391.211			
Final	374.206	17.005	11	.108

Link function: Cauchit

Table 11. Goodness-of-fit for Type 4

	Chi-Square	df	Sig.
Intercept only	516.071	525	.601
Final	335.087	525	1.000

Link function: Cauchit

Table 12. Pseudo R² values for Type 4

Cox and Snell	.090
Nagelkerke	.097
McFadden	.036

Link function: Cauchit

Table 13. The test of parallel lines for Type 4^a

Model	-2 Log-Likelihood	Chi-Square	df	Sig.
Null hypothesis	374.206			
General	326.344 ^b	47.862 ^c	33	1.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories

a. Link function: Cauchit

b. The log-likelihood value cannot be further increased after maximum number of step-halving

c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model.

Validity of the test is uncertain

4 Conclusions

This study presented the current situation of smartphone speech interaction usage in China to a certain extent. It was clear that the barriers that different type users encounter were different when it came to speech interaction, which could help to improve the user experience of smartphone speech interaction pertinently. Conclusions can be drawn as followed through this study.

Speech interaction is an interactive modality for the future, which is mainly reflected by two points:

- The fundamental of speech interaction lies in the development of speech recognition technology. As long as the accuracy of speech recognition can be further improved and users can be better understood, the speech interaction can be widely accepted gradually.
- Speech interaction has developed relatively mature when the new generation users of mobile use the first mobile phone. This generation has a higher acceptance of speech interaction because there is no interaction mode migration problem. They will become the mainstream users of future speech interaction.

Suggestion for the design of speech interaction is to give full play to the advantages of speech interaction by creating rigid demand to increase the opportunities to use speech interaction and cultivating users' habit of using speech interaction. The following points are included:

- Reduce interface exclusivity. Speech interaction can be used as an auxiliary interaction mode to help users to carry out parallel task flow. Therefore, according to different usage scenarios, reducing interface exclusivity means allowing users to use speech interaction for parallel operation, which is not substituted by other interaction modalities.
- Offer services to users actively in combination with situational awareness. For example, when a user sliding screen on the screen page repeatedly, if there is no other subsequent operation, it is more likely to look for a certain application. It is a good opportunity to remind users that they can start the application quickly by using speech interaction, which can not only improve the exposure of the speech interaction, but also not disturb user excessively.
- Guide users in a variety of modalities, such as visual guidance. The guidance is to allow users to know what can be done by speech interaction and how to use speech interaction, especially when users use speech interaction for the first time. They also need guidance when they make mistakes. Not only to explain what the error is, but also to give the appropriate guidance to help users to complete the anticipated operation according to the error content.

Speech interaction has advantages and disadvantages compared with traditional modalities of interaction. Under the current technical conditions, it is necessary to create more opportunities for users to use voice interaction and to cultivate users' habit of using speech interaction by making the best use of the advantages of speech interaction. Of course, these suggestions are also aimed at the current situation of speech recognition technology, mobile computing ability and mobile network speed, which has some limitations and needs further improvement and development.

References

1. Munteanu, C., Cowan, B., Nakamura, K., Irani, P., Oviatt, S., Aylett, M., et al.: Designing speech, acoustic and multimodal interactions. In: *The 2017 CHI Conference Extended Abstracts*, pp. 601–608. ACM Press, New York (2017)
2. Hosmer, D.W., Lemeshow, S.: *Applied Logistic Regression*, 2nd edn. Wiley, New York (2000)
3. Lokosang, L.: *Statistical analysis of determinants of household food insecurity in post conflict Southern Sudan*, University of KwaZulu-Natal (Master's thesis) (2009)
4. Tharakan, C., Suchindran, C.: Determinants of child malnutrition: an intervention model for Botswana. *Nutr. Res.* **19**(6), 843–860 (1999)
5. Armstrong, B., Sloan, M.: Ordinal regression models for epidemiologic data. *Am. J. Epidemiol.* **129**, 191–204 (1989)
6. Harrell, F., Margolis, P., Gove, S., Mason, K., Mulholland, E., Lehmann, D., et al.: Tutorial in biostatistics: development of a clinical prediction model for an ordinal outcome: the World Health Organization ARI Multicenter Study of clinical signs and etiological agents of pneumonia, sepsis, and meningitis in young infants. *Stat. Med.* **17**, 909–944 (1998)
7. Adeleke, K., Adepoju, A.: Ordinal logistic regression model: an application to pregnancy outcomes. *J. Math. Stat.* **6**(3), 279–285 (2010)

8. Abreu, M., Siqueira, A., Cardoso, C., Caiaffa, W.: Ordinal logistic regression models: application in quality of life studies. *Cad Saúde Pública*, Rio de Janeiro **24**(4), 581–591 (2008)
9. Haojie, L., Huiman, X., Aryeh, D., Reynaldo, M.: Effects of early childhood supplementation on the educational achievement of women. *Pediatrics* **112**(5), 1156–1162 (2003)
10. Gemeroff, M.: Using the proportional odds model for health-related outcomes: why, when, and how with various SAS® procedures. In: *The Proceedings of the Thirtieth Annual SAS Users Group International Conference*, Paper # 205-30. SAS Institute Inc., Philadelphia (2005)