Users' Perceived Difficulties and Corresponding Reformulation Strategies in Voice Search

Wei Jeng
School of Information Sciences,
University of Pittsburgh
wej9@pitt.edu

Jiepu Jiang
School of Information Sciences,
University of Pittsburgh
jiepu.jiang@gmail.com

Daqing He
School of Information Sciences,
University of Pittsburgh
dah44@pitt.edu

ABSTRACT
We report on users' perceptions on query input errors and query reformulation strategies in voice search. The perceptions were collected through a controlled experiment. Our results reveal that: 1) users' faced obstacles during a voice search that can be related to system recognition errors and topic complexity; 2) users naturally develop different strategies while dealing with varying types of words that are problematic for systems to recognize.

Keywords
Voice search; voice input errors; query reformulation.

ACM Classification Keywords
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *query formulation*

General Terms
Human Factors; Design; Experimentation

INTRODUCTION
Voice search recently became increasingly popular for both mobile and desktop devices. Compared to conventional search systems when using the keyboard for query input, a user's interaction with voice search systems can be more complex. However, there are few contemporary studies on users in voice search.

Our recent study focused on typical query input errors and users' query reformulation behaviors [3]. Based on search logs, we found that voice input errors were prevalent with the state-of-the-art voice search systems which resulted in substantial declines in search performance. Users adopted both lexical (query term addition, substitution, removal, and re-ordering) and phonetic query reformulations (overstating a part of or the entire query). Some of them are closely related to those previously misrecognized words (e.g. query term substitution and overstate a part of the query). In this paper, we intend to augment the previous study by looking into data collected from the surveys and interviews conducted during the same experiment presented in [3]. Specifically, we explore the following research questions:

- What are the difficulties in voice search as reported by the participants?
- What are users' query reformulation strategies to resolve the voice input errors?

The results of this paper and those in our article [3] provide better insights about voice input errors and users' interactions in current voice search systems. This study will improve the future design of an ideal voice search interface.

Proceedings of HCIR 2013, Vancouver, BC, Canada, Oct 3-4, 2013

RELATED WORKS
Voice search is a relatively new research topic: this includes studies that specifically focus on user interactions such as query reformulation behaviors in voice search. For example: although Crestani et al. [2] conducted a user experiment comparing voice queries with written queries, their experiments did not involve user interactions; Schalkwyk et al. [5] reported on statistics about individual queries from Google Voice's search logs. Our previous study [3] did examine user behaviors in voice search, but the employed data analysis method was primarily log analysis.

Another group of related studies deals with users' responses in spoken dialog systems. For example, Swerts et al. [7] categorized users' responses to the recognition errors in dialog systems including repeating, paraphrasing, adding relevant content, omission and hyperarticulation, which are similar to the lexical and phonetic reformulation: these were similar patterns to those we observed in voice search. Similar findings were reported in [1, 4, 6]. However, spoken dialog systems differed significantly from voice search systems. The former is usually designed to handle structural query input (e.g. the location and time), and solve a specific task (e.g. flight information inquiry), while the latter deals with far more diverse information needs and flexible query inputs. Overall, there are a limited number of studies on user interactions and query reformulation strategies in voice search. A deeper understanding of these issues can foster the more effective design of voice search systems.

METHODOLOGY
Twenty English native speakers were recruited to work on TREC topics using voice search. We used the Google voice search app on an iPad to record all participants' behaviors, including spoken queries, system transcribed queries and clicking history. Each participant worked on 25 topics selected from a pool of 50 topics in total. For each topic, the participants could freely interact with the voice search system within a 2-minute session (e.g. click and check results, reformulate voice queries), but typing on the iPad was restricted throughout the experiment. We reported experiment settings in detail in [3].

Table. 2 Users' Strategies for Difficultly Recognized Words

Types of Words	Example (errors/ used times)	Users' Strategies for Given Words
Acronym Single-worded queries without context	ER (29/29); AVP (11/11) sun (24/41), theft (14/14), art (24/53)	I ▪ Use full name (e.g. AVP, Association of Volleyball Professionals) ▪ Add extra key word (e.g. ER George Clooney) ▪ Change the part of speech (e.g. tax and taxing, S03; use to using, S04)
Two syllables can slide together easily Diphthong Unvoiced/ voiced consonants may fail	Rap in "rap and crime" (13/36) Fraud (12/14), horse (10/36) **V**iolence (19/27) "**t**alks" in "Irish Peace talk" (9/15), ethnics (10/21)	II ▪ Repeat the same query with the same tone ▪ Repeat the same query but speak differently in terms of: o Pauses between words o Slowing down o Add an Emphasis
Non-English words Named Entity "I don't think I pronounced it properly"	El Niño (31/46) Ralph (22/36), Owen (25/26) Culpeper (18/27), polygyny (8/8)	III ▪ Try different pronunciations (/ninjoʊ/ and /nino/, S05, S07, S11) ▪ Spelling (e.g. Niño and n-i-n-o, S09) ▪ Avoid perceived difficult words in terms of: o Pick a synonym (e.g. theft and espionage, S09) o Describe associated things, but nothing directly related (e.g. polygyny to one man two wives, S19)

This paper focuses on studying users' perceptions of the difficulties and query reformulation strategies in voice search. The data used in the analysis included:

1. Participants' background information collected at the beginning of the experiment;

2. The participants rated each topic as to the difficulty of query formulation after finishing all of the topics;

3. Participants' answers from a concluding semi-structured interview which featured six overarching interview questions. The survey covered: 1) the efforts and difficulties of using voice input versus using a keyboard in search; 2) the most difficult topic(s); 3) types of words that are less likely to be recognized; 4) the solutions or strategies to address recognition errors noted by users; 5) users' affective feelings when recognition errors happened; and finally 6) occasions when it is better to use/ not use voice-based search. Due to the user-centered emphasis of this paper, we focus on reporting results of questions 1-4.

RESULTS

Participants

Among the 20 participants, 65% (N=13) were undergraduate students and the remaining were graduate students. The average age of the participants was 23.7 (SD=4.72), and 14 of them were female. When asked about the frequency of using search engines, 85% (N=17) reported that they used search engines on desktop or laptop computers on a daily basis, whereas only 40% (N=8) used search engines on mobile devices every day. Half of our participants reported that they had never used a voice search systems neither on computers nor on mobile devices.

Users' Perceived Difficulties

In this section, we focus on difficulties experienced and reported by the participants.

Voice Input Errors

In our study, we define voice input error as the case when the search query received and recognized by the voice search system is different from the query intended by the user. In [3], we discovered two types of errors, as observed in our experiments. Eighty-nine percent are speech recognition errors, i.e. the automatic speech recognition system cannot provide a correct transcription. Eleven percent is errors caused by improper system interruptions, i.e. the user is interrupted by the voice search system before completing articulation of the query. This happens when the system "believes" that the user has finished speaking while in reality they have not (e.g. the user pauses for a relatively long period of time but would like to continue speaking).

In the interview, the majority of the participants (12 of 20) explicitly expressed that voice search is more challenging and it costs more in terms of effort than conventional search (using the keyboard for query input) due to voice input errors. For example, S16 expressed: "I'd rather type. It takes forever for them (the search engine) to pick up what you're saying." S14 noted: "In numerous times I had to repeat. Actually, this topic right here, I didn't search for Philippines. It just sort of popped up." This is consistent with our previous article [3], in which voice input errors were not only responsible for a significant decline in search performance for individual queries, but also led to an increased amount of efforts and users' negative feelings.

The participants did not specifically report that one is more serious or troublesome than the other (even though in [3] we did find a lower level of performance for queries that were improperly interrupted), although there are clear difference between the two types of errors,

Topic Familiarity and Complexity

As with conventional search systems (using a keyboard for query input), users also found that topic familiarity and complexity are factors affecting search difficulty in voice search. Four participants stated that topic familiarity was a major obstacle they perceived during the experiment: "I didn't know enough about those topics to re-word the speech properly. (S01)". S07 also reported on topic complexity for the topic related to "marine vegetation": "… I mean, finding marine vegetation was easy but how it …

Table 1. Users' perceptions on the easiness of topics and the influence of voice input errors on users' search performance.

Perceived Difficulty	% missing words	Jaccard Similarity	Drop of nDCG@10
6 (the least difficult)	0.3304	0.4900	0.1023
5	0.2805	0.5140	0.1045
4	0.3274	0.3725	0.1411
3	0.3336	0.4147	0.1187
2	0.3825	0.3261	0.1464
1 (the most difficult)	0.4658	0.1365	0.1831

but I couldn't find anything on how it was used in relation to food and drug and it kept …[sic]"

Query Formulation

After finishing each topic, we also asked the user to rate (using a 6-point Likert scale) the topic regarding on whether or not it is difficult to formulate queries. We found that users' ratings do correlate with the seriousness of the errors and their actual search performance on the topics.

Similar to [3], we characterize the influence of the voice input errors by: the average proportion of words spoken by users that were missed in the system's transcription (% missing words), the Jaccard similarity between results of the voice queries and the transcribed queries, and the decline of nDCG@10 in the transcribed queries compared to the voice queries' actual content. As shown in Table 1, voice input errors are less severe when the topic is perceived to be easier by the user; in addition, the search performance is less affected by the errors (although results of two adjacent rating values are sometimes inconsistent). This indicates that users can correctly perceive difficulties in query formulation: such difficulties have a negative effect on users' search performance.

The reasons for difficulties in formulating queries can be diverse. In addition to topic familiarity and complexity (as discussed in the last section), it happens when the topic has theme words that are essential and cannot be replaced yet those words are particularly difficult for the system to recognize. For example, S03 reported that the topic "Culpeper national cemetery" because: "I could not pronounce. I couldn't get the name. I could not even find anything on it. [sic]" We will report the typical difficult words in details in the next section.

Difficult Words and Reformulation Strategies

We asked participants whether or not they noticed any types of words or phrases that were especially difficult to recognize as well as potential reformulation strategies. In each of the following sub-sections, we describe one type of difficult words and the corresponding strategies.

Acronyms and Single-worded Queries: Create More Clues

As shown in Table 2, we categorized acronyms and single-worded queries in the same group because users pointed out their common characteristics: the lack of context. Several participants (N=5) mentioned that acronyms, abbreviations, or very short words can lead to serious recognition issues. For example, S02 reported: "Oh, when you're using abbreviations or saying just a single word or there are short words, like art, was really hard for it to pick up." In the search log, we also found that the queries "ER" (the TV show "Emergency Room") and "AVP" (acronym for "American Volleyball Professionals") had a 100% error rate (see Table 2).

To address these types of difficult words, the participants reported that they tried to use the full name for the acronyms; or to add additional clues: "If I know the word, like ER for example. I kind of like use a key word that makes it obvious what I'm referring to. ER George Clooney. [sic] (S17)".

Frequently Misrecognized Words with Observable Phonetic Features: Repeat

Some of the participants were able to describe certain phonetic features of the frequently misrecognized words. S17 reported some words with syllables that can "slide together" were hard for the system to recognize, such as "horse hooves" or "rap and crime". As we examined the search log, we found that the word rap in "rap and crime" was misrecognized 13 times out of 36 uses. S17 and S18 both reported that a diphthong word (i.e. two adjacent vowels) would cause confusing results (e.g. "hooves" was misrecognized as "who" or "whose"). Participants S04 and S07 also noted errors when using voiced and unvoiced consonants, respectively: "consonant, P, T, K, those are… it doesn't hear them as well and so for example saying Irish Peace Talks. (S04)"; "Violent. I guess where it… words that don't have kind of like sharp consonants in them … to them, it has trouble finding those words, I would guess. (S07)".

In response to this group of errors, participants (N=3) reported that they would just repeat or overstate the error words (e.g. speak slower, clearer, louder) (Type II in Table 3). For example, S07 was asked about how she dealt with the errors when using the word "violence": "I would speak clearly and enunciate. I would definitely speak in a manner that I wouldn't speak to control. [sic]"

Words with Uncertain/Unknown Pronunciations

Participants also noted some words as difficult when they were uncertain about the pronunciation. For example, non-English words such as "El Niño" can result in a high error rate (31 of 46 times being misrecognized). Some users tried to pronounce it as the "ninjoʊ" sound: "my voice's trying to mimic the sounds of the Spanish language, didn't come across as well, as the English words [sic] (S17)". Users also reported that they were not familiar with the proper pronunciation of some relatively rare words, such as "Culpeper" (18/27) and "polygyny" (8/8).

We found participants used different strategies when they encountered unfamiliar or non-English words. According to the experiment log, S09 spelled "n-i-n-o" by letters when she performed her sixth attempt on the topic. The example

below shows a participant's (S19) search log when she tried to input the query "polygyny" (sounding "dʒənɪ" at "gyny"):

#	*Voice Query*	*Transcribed Query*	*Reformulation Strategy*
1	polygyny	poligamy	- -
2	polygyny	paul inca ny	Emphasis
3	polygyny	polly guinea	Emphasis
4	polygyny	call gary	Try different pronunciations
5	polygyny	polygamy	Emphasis
6	one man two wives	1 man to live	Describe associated things

First, the user offered "gənɪ", but the system did not return the result that she expected. After repeating the same sound (gənɪ) twice while overstating, in the fourth attempt, she pronounced it differently as "gaɪnɪ". However, the "gaɪnɪ" sound seemed to have a critical error as well. Finally, she abandoned the word and used "one man two wives" instead. Therefore, we anticipate that, if a user continued to fail after too many attempts at the same word, it is very possible for the user to use several of the Type III strategies in Table 2. S18 stated that sometimes the Repeat strategy would not work very well because "I feel that if you were to say it again there's not going to be a big difference [sic]". At this point, Type III strategies seemed to be "a shelter of last resort" for any type of difficult words as it will at least generate "some differences".

DISCUSSION

By analyzing participants' responses in the interviews, we found that the difficulties in using voice search systems (as recognized by the users) come from both those related to voice query input and unrelated (e.g. topic familiarity and complexity). The users perceived difficulties in query formulation are associated with the seriousness of the error and the query performance. As shown in previous sections, users reported distress with the voice input errors. They also noted a tendency to use alternative input methods when encountering errors ("I'd rather type (S16)"). This suggests the necessity for offering multi-modal query inputs in current voice search systems. As we restricted our experiments to voice inputs, it would be interesting to further explore user interactions in systems with multi-modal query inputs.

We noted three types of difficult words reported by the users, which helps us to explain (Table 3 of [3]) and the author's categorization of the words with high error rates. Most of the categories in [3] were also reported by this study's participants (i.e. acronyms, named entities, and non-English words). Participants also provided possible reasons for some of the uncategorized words with high error rates. This may provide those studying automatic speech recognition with first-hand examples of errors.

While in [3] we found that specific query reformulation strategies best address certain speech recognition errors, this work confirms that query reformulation strategies indeed are used to solve specific speech recognition errors Although "partial emphasis" (overstating a part of the voice query) and query term substitution are the two reformulation patterns used most often to correct error words [3], our participants did not specifically report this phenomenon. We suspect that this was due to the fact that participants were able to recognize and summarize some acoustic features of those words most often incorrectly recognized. In this case, the participants' first response is to repeat or improve their pronunciation rather than switch to alternate words. However, as found in [3], such strategy had limited effectiveness and was less useful than query term substitution. As it is feasible to automatically detect the adopted query reformulation strategies, voice search systems may benefit by providing better guidance on the selection of reformulation strategies, e.g. reminding the user that it is probably more effective to try other words when speech recognition errors happen. In addition, voice search systems should support other input query strategies: two of our participants adopted spelling as their input strategy.

CONCLUSION

In this paper, we confirmed and expanded upon many of our findings in [3]. We found that users perceived voice input errors, topic familiarity, and topic complexity as the major obstacles to effective voice search. The users also reported typical types of words that were difficult for systems to recognize, as well as the corresponding reformulation strategies to solve those issues. These findings will help us to better understand the current issues with user interactions in voice search.

REFERENCES

1. Bohus, D. and Rudnicky, A.I. Sorry, I didn't catch that!-an investigation of non-understanding errors and recovery strategies. *6th SIGdial Workshop on Discourse and Dialogue*, (2005).
2. Crestani, F. and Du, H. Written versus spoken queries: A qualitative and quantitative comparative analysis. *Journal of the American Society for Information Science and Technology 57*, 7 (2006), 881–890.
3. Jiang, J., Jeng, W., and He, D. How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search. *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR'13)*, (2013).
4. Raux, A., Langner, B., Bohus, D., Black, A.W., and Eskenazi, M. Let's go public! taking a spoken dialog system to the real world. *in Proc. of Interspeech 2005*, (2005).
5. Schalkwyk, J., Beeferman, D., Beaufays, F., et al. "Your Word is my Command": Google Search by Voice: A Case Study. In A. Neustein, ed., *Advances in Speech Recognition SE - 4*. Springer US, 2010, 61–90.
6. Shin, J., Narayanan, S., Gerber, L., Kazemzadeh, A., and Byrd, D. Analysis of user behavior under error conditions in spoken dialogs. *Proceedings of ICSLP*, (2002).
7. Swerts, M., Litman, D., and Hirschberg, J. Corrections in spoken dialogue systems. *Proceedings of the International Conference on Spoken Language Processing*, (2000), 615–618.