Results Presentation Methods for a Spoken Conversational Search System

Johanne R. Trippas
johanne.trippas@rmit.edu.au

Damiano Spina
damiano.spina@rmit.edu.au

Mark Sanderson
mark.sanderson@rmit.edu.au

Lawrence Cavedon
lawrence.cavedon@rmit.edu.au

School of Computer Science and Information Technology
RMIT University, Melbourne, Australia

ABSTRACT

We propose research to investigate a new paradigm for Interactive Information Retrieval (IIR) where all input and output is mediated via speech. Our aim is to develop a new framework for effective and efficient IIR over a speech-only channel: a *Spoken Conversational Search System (SCSS)*. This SCSS will provide an interactive conversational approach to determine user information needs, presenting results and enabling search reformulations. We have thus far investigated the format of results summaries for both audio and text, features such as summary length and summaries documents (noisy document or clean document) generated from (noisy) speech-recognition output from spoken document. In this paper we discuss future directions regarding a novel spoken interface targeted at search result presentation, query intent detection, and interaction patterns for audio search.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]; H.3.3 [**Information Search and Retrieval**]

Keywords

Spoken Retrieval; Search Result Summaries; Crowdsourcing; Spoken Conversational Search

1. INTRODUCTION

The popularity of mobile technology, such as smartphones and wearable devices, has made information access on mobile devices a topic of growing relevance [3]. The usage ergonomics differ from desktop computers due to a smaller screen and the lack of a physical keyboard. The characteristics of mobile devices make them challenging for displaying search result summaries. As a result, a a Spoken Conversational Search System (SCSS) could be suitable for information access on mobile devices due to the the system's interactive nature.

NWSearch'15, October 23, 2015, Melbourne, Australia.

DOI: http://dx.doi.org/10.1145/2810355.2810356.

Speech-based search has received more attention in recent years with the development of systems such as Siri, Google Now, and Cortana, where input and output are mediated via speech. Although these systems allow users to pose queries by speech, they are limited in their response capabilities. For example, such systems are often able to reply via speech to a "factoid" style search query (e.g. *"What is the capital of Australia"*), but if a non-factoid style search query is posed, the system reverts to displaying the results as a ranked list on screen. In some situations, a speech-only interface is of interest to the user, for example when no screen or keyboard is available [22], or when users are on the move [11, 13, 18], operating machinery [6, 7], or using wearable devices [3]. Results presented on screen may also cause some accessibility issues for people with a specific learning disability such as dyslexia, people with limited literacy skills, or people with a visual impairment [15].

A key challenge of a SCSS is result presentation. Many ways of presenting search results over audio exist. A SCSS should be able to present the search results in a way that enables users to listen to results, compare them after listening and decide which search result they want to know more. This requires many insights into how users process information-dense speech. Thus designing a SCSS involves different complexities compared to designing graphical user interfaces. For example, speech as output requires information to be presented concisely to minimise the information load on users [18]. Dealing with both speech output and its content is also challenging, as noted in [15]; a solution suggested therein involved lowering task complexity by adding structure to the search task to avoid cognitive overload.

To allow users to search using a SCSS we propose guiding them not only through the search results but also through query refinements. Therefore we will have to study the interaction patterns for audio search closely and understand how users search. This will help us to understand at which stages users change the intent of their query and the dialogue move they use to do this.

2. RELATED WORK

Much research has been conducted into supporting searching by voice input. However, only a few studies focused on the presentation of search results over a speech-only channel [15]. In this section, we briefly outline some of the challenges associated with (cognitive) processing of search results over speech, and briefly point to work in related fields.

2.1 Cognitive Challenges of Spoken Search

Search over speech presents the user with several cognitive challenges. Since speech is a temporal medium, it is taxing for users [21].

Speech is transient, leaving no trace of the message to which the user can later refer [12]. As a result, it is difficult to convey large amounts of information via speech without overloading the user's short-term memory [12, 18]. A lack of graphical support, in the form of text, means much less information can be transmitted to the user at one time [22]. The fact that text can be re-read means that most people absorb written information better than spoken information [22]. In addition, information delivered via speech is linear, making it difficult to present complex structure [18]. Simultaneously, as a series of sounds, speech can blend with other environmental sounds making it challenging to recognise [18].

To address some of these challenges, [22] suggest keeping spoken output brief by eliminating extra words and using menus instead of lists to avoid stressing users' short-term memory. Though such menus can act as shortcuts, they still need to be memorised [9]. By limiting the number of options, it becomes easier for users to remember all options thus reducing the working memory load [20]. Another way to avoid lists of information is by incorporating a conversational style [20]. Conversational style involves the system using questions to filter the options of the query results, presenting fewer options per turn and providing confirmations [20]. This helps users navigate the relevant information rather than requiring them to remember all options [9]. However, the longer the conversation, the greater the number of refinements the user needs to remember [20].

Systems such as screen readers and Spoken Dialogue Systems (SDSs) are areas where research has been conducted into how to allow users to achieve their goals without cognitively overloading them. Users of screen readers are often presented with a flood of information resulting in an information overload which is cognitively taxing [15]. To mitigate information overload, SDSs address this issue by guiding the user through a dialogue to allow them to select the right option. Thus SDSs overcome the problem of information overload by refining options whenever there are too many options available [19]. By studying screen reader users and the information seeking process, [15] observed that screen reader users interact differently with search engines than non-screen readers. Screen reader users are often limited due to cognitive challenges and the lack of support to over overcome those challenges.

3. CURRENT AND FUTURE WORK

This section first describes experiments which have already been conducted and then discusses future research directions to develop a SCSS.

3.1 Length of Web Search Result Summaries

In an initial experiment [17], we aimed to better understand how to present search results over a speech-only communication channel without overwhelming the user with information, nor leaving the user unsure as to whether what they heard covered the information space. The length of search summaries plays an important role in the presentation of search results. To investigate the impact of the search result summary length in a spoken retrieval scenario, we used a crowdsourcing platform[1] to present queries and search result summaries with different lengths to users. The queries in this experiment were query topics from the Text REtrieval Conference (TREC) 2013 Web Track to reflect common web search tasks [4]. The queries used were either *single-faceted queries* (queries with a clear intent) or *faceted queries* (queries with a broader intent and represented in subtopics). We investigated whether these two categories had any impact on the preference of the search result summary length.

The search result summaries presented to the users were either full-length or *truncated* versions of Google search result summaries. Users were presented with both versions of the summaries and were asked which version they most preferred. The preference was asked for both text and audio presentation of the summaries. This allowed us to use the text presentation as a baseline measure of the system.

Users reported that they preferred the full-length search result summaries in text format over their truncated counterpart. For the audio format, no clear preference was reported between full-length or truncated search result summaries. However with single-faceted queries in audio format users showed a clear preference for truncated summaries.

Joachims et al. [10] found that the first and second search result summaries receive the most attention from users. Our findings support these results for single-faceted query judgement distributions. However, this distribution was not found in the faceted search result summaries. Instead, the first and last search result summary received the most attention.

We believe that there is a need of further research on how to present results over audio for faceted queries.

3.2 Generating Podcast Summaries from Noisy Automated Transcriptions

In a second experiment, we investigated whether summaries generated from automated speech recognition (ASR) transcripts would allow users to effectively judge document relevance. We also investigated whether those judgements were as accurate as non-ASR transcript summaries and if there was a preference in the different summaries (ASR summaries vs. non-ASR summaries). We used podcasts and corresponding manual transcripts from the Australian Broadcasting Commission (ABC)[2] for this experiment. Thus, the text collection consisted of the manual transcripts transcribed by the ABC and automated transcripts transcribed with the AT&T WATSON Speech API[3]. A *known-item* document scenario [1] was used to design manual queries. Topic modeling was performed to minimise the likelihood of selecting documents associated with underrepresented topics. We used the same crowdsourcing platform as mentioned above to collect user judgements and user preferences of the different summaries. The results from the relevance task suggested that summaries generated from ASR transcripts—with correction of ASR errors in the summaries themselves—were as effective for users performing relevance judgements as were summaries generated from the manual transcripts. Note that corrected summaries simulates the playback of such summaries in an audio channel. The ASR corrected summaries were also no less preferred than summaries generated from the original podcast transcriptions. Therefore we suggest that transcripts with ASR errors can be effectively used for selecting segments of podcasts for use as audio summaries for a speech-only podcast search system.

3.3 Future Work

This section discusses various techniques which we propose to use in the design of a SCSS.

3.3.1 Query Intent and Result Presentation

The results from the experiment in Section 3.1 suggested that different kinds of queries (single-faceted vs. faceted) benefit from a summary optimised for the type of query. We therefore suggest researching query intent recognition, allowing us to categorise

[1] http://www.crowdflower.com

[2] http://www.abc.net.au

[3] http://developer.att.com/apis/speech/docs

queries by type of intent, which will help to present users with the appropriate summary for a specific query type. Queries with a clear query intent (single-faceted queries) could be represented by shorter summaries than queries with a broader query intent (faceted queries). Our focus will be on addressing the faceted queries to help a user refine the query in a conversation to narrow down search. We also plan to investigate how a system can narrow down users' search without *directing* them unduly.

Instead of presenting query results in a ranked list, we plan to investigate alternative ways of organising and presenting results. One possible method is by clustering the query results where documents are clustered based on document similarities or topic. Clusters can be used to maximise coverage of an information space, allowing users to understand the important concepts as well as potential relationships between the search results [14]. Such structure will provide a scaffold for conversational approaches to navigating a set of results. The system may also present options which are of no interest to users, making it easier for users to explore and dive deeper into the cluster that is most relevant to them [5]. Another method would be to present facets to the user to allow them to narrow down their search, similarly to the clustering approach [8].

3.3.2 *Designing New Patterns of Interacting with Information*

Since we aim to understand a *dialogue-oriented interface* [16], the system needs to be able to identify which dialogue acts the user makes use of. Once a dialogue act of a spoken query is classified into a certain group, the system will be able to support the user with presenting the right information in the right presentation style for that query. Thus we will use the extensive knowledge base from speech act identification to determine the function of an utterance of the query logs to which we have access. These logs consist of many users interacting with a conversational search system to find and listen to podcasts, books and news articles. The captured interactions are organised in sessions and include queries, navigation through search results and play/stop commands, among others.

Developing a spoken user interface differs in many aspects from developing a graphical user interface. In developing a spoken user interface, [12] suggests first listening and learning how people talk. We therefore propose using a Wizard of Oz (WOZ) methodology instead of using sketches and drawn storyboards [11]. The WOZ technique allows us to capture the dialogue flow and user responses to information presentation style. With the logs we will be able to compare the human-computer spoken interaction styles to optimise the dialogue flow.

3.3.3 *Developing New Search Models for Audio*

As we are researching the development of a SCSS, we need to be able to understand the cognitive demands which are put on the users when using the system. It has been mentioned that demands on a user reduce performance [2]. In our case we could see a decline in performance such as speed, response time or errors. Thus it is important to understand and minimise the impact of these factors. Many cognitive models have been proposed, though they are often for a conversation between two people. Our aim is to identify which model is appropriate to a SCSS and if necessary modify the model. One of the tools we will use is the widely-used NASA Task Load Index (NASA-TLX) which will allow us to understand the workload of a user when using a SCSS. The NASA-TLX could be used in some of the crowdsourcing tasks.

4. CONCLUSION

This paper has outlined a research plan for developing a Spoken Conversational Search System. As little research has been conducted on this topic, we need to gather different methodologies from different fields to develop techniques to handle issues with search result presentation, query intent detection and interaction patterns for search over speech.

5. ACKNOWLEDGMENTS

This research was partially supported by Australian Research Council Project LP130100563 and Real Thing Entertainment Pty Ltd.

6. REFERENCES

[1] L. Azzopardi and M. de Rijke. Automatic construction of known-item finding test beds. In *Proc. of SIGIR'06*, pages 603–604, 2006.

[2] C. Baber, B. Mellor, R. Graham, J. M. Noyes, and C. Tunley. Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication*, 20(1):37–53, 1996.

[3] E. Chang, F. Seide, H. M. Meng, C. Zhuoran, S. Yu, and L. Yuk-Chi. A system for spoken query information retrieval on mobile devices. *Speech and Audio Processing, IEEE Trans. on*, 10(8):531–541, 2002.

[4] K. Collins-Thompson, P. Bennett, F. Diaz, C. L. Clarke, and E. M. Voorhees. Trec 2013 web track overview. In *Proc. of 22nd Text REtrieval Conference (TREC)*, 2014.

[5] V. Demberg and J. D. Moore. Information presentation in spoken dialogue systems. In *Proc. of EACL'06*, 2006.

[6] V. Demberg and A. Sayeed. Linguistic cognitive load: implications for automotive uis. In *Proc. of AutomotiveUI 2011*, 2011.

[7] V. Demberg, A. Winterboer, and J. D. Moore. A strategy for information presentation in spoken dialog systems. *Computational Linguistics*, 37(3):489–539, 2011.

[8] G. Drzadzewski and F. Tompa. Enhancing exploration with a faceted browser through summarization. In *Proc. of DocEng 2015*, 2015.

[9] M. Hearst. *Search user interfaces*. Cambridge University Press, 2009.

[10] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. of SIGIR '05*, 2005.

[11] S. R. Klemmer, A. K. Sinha, J. Chen, J. A. Landay, N. Aboobaker, and A. Wang. Suede: a wizard of oz prototyping tool for speech user interfaces. In *Proc. of UIST'00*, 2000.

[12] J. Lai and N. Yankelovich. *Speech Interface Design*, pages 764–770. Elsevier, 2006.

[13] L. J. Najjar, J. J. Ockerman, and J. C. Thompson. User interface design guidelines for speech recognition applications. In *Proc. of IEEE Virtual Reality Annual International Symposium (VRAIS '98)*, 1998.

[14] H.-T. Pu. User evaluation of textual results clustering for web search. *Online Information Review*, 34(6):855–874, 2010.

[15] N. G. Sahib, D. Al Thani, A. Tombros, and T. Stockman. Accessible information seeking. In *Proc. of Digital Futures '12*, 2012.

[16] A. Stein and E. Maier. Structuring collaborative information-seeking dialogues. *Knowledge-Based Systems*, 8(2):82–93, 1995.

[17] J. R. Trippas, D. Spina, M. Sanderson, and L. Cavedon. Towards understanding the impact of length in web search result summaries over a speech-only communication channel. In *Proc. of SIGIR*, 2015.

[18] M. Turunen, J. Hakulinen, N. Rajput, and A. A. Nanavati. Evaluation of mobile and pervasive speech applications. In *Speech in Mobile and Pervasive Environments*, pages 219–262. 2012.

[19] S. Varges, F. Weng, and H. Pon-Barry. Interactive question answering and constraint relaxation in spoken dialogue systems. *Natural Language Engineering*, 15(01):9–30, 2009.

[20] M. Wolters, K. Georgila, J. D. Moore, R. H. Logie, S. E. MacPherson, and M. Watson. Reducing working memory load in spoken dialogue systems. *Interacting with Computers*, 21(4): 276–287, 2009.

[21] N. Yankelovich and J. Lai. *Designing speech user interfaces*. ACM, 1998.

[22] N. Yankelovich, G.-A. Levow, and M. Marx. Designing speechacts: Issues in speech user interfaces. In *Proc. of SIGCHI'95*, 1995.