Informing the Design of Spoken Conversational Search

Perspective Paper

Johanne R. Trippas
RMIT University
Melbourne, Australia
johanne.trippas@rmit.edu.au

Damiano Spina
RMIT University
Melbourne, Australia
damiano.spina@rmit.edu.au

Lawrence Cavedon
RMIT University
Melbourne, Australia
lawrence.cavedon@rmit.edu.au

Hideo Joho
University of Tsukuba
Tsukuba, Japan
hideo@slis.tsukuba.ac.jp

Mark Sanderson
RMIT University
Melbourne, Australia
mark.sanderson@rmit.edu.au

ABSTRACT

We conducted a laboratory-based observational study where pairs of people performed search tasks communicating verbally. Examination of the discourse allowed commonly used interactions to be identified for Spoken Conversational Search (SCS). We compared the interactions to existing models of search behaviour. We find that SCS is more complex and interactive than traditional search. This work enhances our understanding of different search behaviours and proposes research opportunities for an audio-only search system. Future work will focus on creating models of search behaviour for SCS and evaluating these against actual SCS systems.

KEYWORDS

Conversational Search; Voice Interaction

ACM Reference format:
Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the Design of Spoken Conversational Search. In *Proceedings of 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, March 11–15, 2018 (CHIIR '18),* 10 pages.
DOI: 10.1145/3176349.3176387

1 INTRODUCTION

With the development of accurate speech recognition and text to speech synthesis, it has become possible to speak simple natural language queries and for an information retrieval (IR) system to verbally respond. However, simply speaking the textual output of a standard search engine result page (SERP) has been found to be insufficient [27]. The underlying components of an Spoken Conversational Search (SCS) system (where communication between user and system is mediated verbally through audio) will need to operate differently from a traditional IR system [36, 37].

Conversational search has been identified as an important new research direction at several meetings including the Second Strategic Workshop on IR [2]. At one recent meeting[1] it was indicated that there is a lack of understanding of search tasks, search result description, and evaluation of SCS. More importantly, the IR community lacks a broader understanding on how users will interact with these highly interactive search systems and which components maybe involved.

In this paper, we provide an insight of conversational search challenges and opportunities, specifically on what search interactions might look like. Note, we focus on audio-only SCS excluding multi-modal or visual interactions. Thus, we designed a study to observe characteristics of spoken interactions and use the observations of the study to examine differences between "conventional text" search and SCS. Hence, this research is not conducted in a (typical) statistical manner but explores the ranges of possibilities or actions in a SCS setting [8].

Radlinski and Craswell [26] define a conversational search system as "…a system for retrieving information that permits a mixed-initiative back and forth between a user and agent, where the agent's actions are chosen in response to a model of current user needs within the current conversation, using both short- and long-term knowledge of the user". With this definition we attempt to place our observations within the context of existing models of information seeking behaviour (e.g., Belkin's Anomalous State of Knowledge (ASK) [6] or Marchionini's Information Seeking Process (ISP) [20]).

Our main contributions are threefold, we identify: *(i)* the impact of the audio channel on interactions between the user and the system, and on search interactions; *(ii)* different levels of system involvement suggesting SCS systems will have to become *actively* involved in a users' search process; *(iii)* new research opportunities linked to the change of the information transfer channel.

In the following section, an overview is provided of our observational experimental setup and the participants. Section 3 describes observations related to the change in modality of interaction (i.e., audio channel) highlighting the importance of understanding the interactivity of this new search paradigm. Then we present observations in Section 4, which have a strong link to search. In Section 5 we provide an overview and discuss the suggested results from the

CHIIR '18, New Brunswick, NJ, USA
© 2018 ACM. 978-1-4503-4925-3/18/03…$15.00
DOI: 10.1145/3176349.3176387

[1] International Workshop on Conversational Approaches to Information Retrieval (CAIR) at SIGIR 2017 https://sites.google.com/view/cair-ws/

made observations. We also suggest ways of differentiating diverse search systems depending on their involvement with the users' search process. The final section provides a conclusion and outlines future work. This paper will provide related work throughout the observations allowing for a better demonstration and integration with our findings.

2 METHODOLOGY

An empirical laboratory study was conducted to investigate the aspects of an SCS system [37]. This study was designed to understand how users communicate in an audio-only search setting and focuses on the issues one could encounter when using such system. Thus, observing how people search in this setting provides initial insight into the interactions which take place. We also consider our observations with existing research and models creating a broader understanding of search in a spoken conversational setting. Thus, we combine previous research with our empirical observations in order to extend the general search expertise.

2.1 Observational Study

The study consisted of a series of sessions, each with two participants, one participant acting as the *seeker* (or user) and the other as the *intermediary*. The seeker received a backstory describing a *task* to find information on a certain topic and had no access to anything to satisfy that information need (such as a computer). The backstories were based on TREC topics (Q02, R03, and T04) and are described by Bailey et al. [4]. The intermediary had access to a search engine through a computer but did not have access to the seeker's backstory. Participants were not able to see each others' facial expressions. All tasks were randomized in order. The roles of the participants were randomly assigned.

The participants had to collaborate with each other in order to satisfy the information need. Before and after each scenario the participants filled out a short questionnaire and at the end of the experiment, a semi-structured interview was conducted. Participants could leave at any time and there were no adverse consequences apart from 90 minutes of the participants' time. All interactions were recorded and transcribed for analysis.[2] This process is described in more detail by Trippas et al. [38].

2.2 Participants

The study involved twenty six participants recruited through a mailing list.[3] Fifteen participants were female and eleven were male. Participants' mean age was thirty (SD=11). Twenty two participants (85%) reported to be a native English speaker and four participants (15%) reported to have a high level of English. Eighteen participants reported that they held either a Bachelor's or Master's degree (69%) and eight participants reported their highest level of degree was awarded at high school (31%). The frequency of main fields of education were Science (19%), Engineering (19%), and Law (11%). The majority of participants were students (73%) or employed (19%). Computer use for over ten years was reported by 85% of our participants, while 15% reported use for 5-10 years. All participants reported that they used search engines daily with the majority of participants reporting that they used a search engine more than eight times per day (54%).

Participants rated their own search skills on a 5-point scale, where 1=novice and 5=expert. Participants' mean search skills was 3.9 (SD=0.5), with a minimum score of three and maximum of five. Participants reported their usage of intelligent personal assistants, such as Google Now, Siri, Amazon Echo or Cortana. The majority of the participants reported that they had used an assistant a couple of times in the past but did not use it anymore (27%). 19% of the participants reported that they used intelligent personal assistants one to three times per week.

3 NON-SEARCH RELATED OBSERVATIONS

Results are divided into two sections. First, Section 3 presents high level investigations, which are not constrained specifically to search but cover other aspects of SCS, such as communication and cognitive user models. The observations are linked to applicable more general models, which are also applicable to intelligent agents.

Second, Section 4 presents observations, which are framed in the ISP of Marchionini [20], allowing us to introduce our observations in a structured manner.

3.1 One Utterance Consists of Multiple Moves

Complexity appeared to be added in a search process by allowing users to convey their query verbally. In a traditional visual-textual interface, a mouse click or key press are single *moves*. Each action(s) a system needs to take is linked to an atomic move from the user. It could be said that we have a one-action search paradigm (action-response) in a visual-textual setting: if a user provides input (query) the system will respond (results). Search interactions in such a setting can be seen as a linear process.[4]

However, we observed that this paradigm does not hold in a verbal setting. Users were observed describing multiple moves in one utterance. Two such examples are shown in Figure 1. We also observed more than two moves in one utterance; however, this was rather unusual and needs further investigation. These two or more moves in a single utterance increase the complexity of seekers and intermediaries' interactions.

General information seeking behaviour models such as Wilson's information behaviour [41], Marchionini's ISP [20], or Saracevic's Stratified model [29] are too broad and not specific enough and therefore do not provide the necessary information about whether one utterance or interaction may have several moves. Belkin [7] however, describing his work with intermediaries, mentions that one utterance can contain several moves. Yet he does not elaborate on this aspect or on how this may translate in a non-human conversation and how a system should handle this conversation.

Several researchers have proposed ways to incorporate IR through dialogue [25, 31, 32]. Sitter and Stein [31] developed the COnversational Roles (COR) model based on dialogue acts as a general model for information-seeking dialogue combining it with a dialogue plan [3]. The plan is used to guide users through stages of IR

[2] Transcripts of the experiment are available at http://jtrippas.github.io/Spoken-Conversational-Search

[3] The mailing list is maintained by the Behavioural Business Lab at RMIT University, https://orsee.bf.rmit.edu.au/. RMIT University's Ethics Board approval number ASEHAPP 08-16.

[4] Note, this one-action search paradigm could be manipulated by seekers, for example opening several tabs from the SERP.

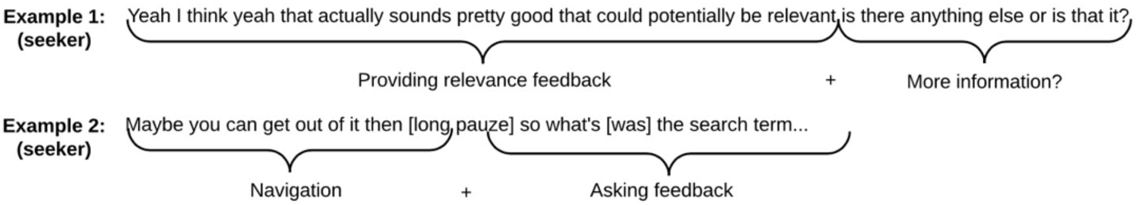
Example 1:
(seeker)
Yeah I think yeah that actually sounds pretty good that could potentially be relevant is there anything else or is that it?
Providing relevance feedback
+
More information?
Example 2:
(seeker)
Maybe you can get out of it then [long pauze] so what's [was] the search term...
Navigation
+
Asking feedback

Figure 1: Multiple moves examples.

with two actors. These actors are noted as A (information seeker) and B (information provider) as illustrated in Figure 2. The Figure also provides the main overview of the COR model where the bold lines are the optimal path taken to solve an information need.

The moves (for example from one to two, Figure 2) consist of atomic dialogue acts [33]. By way of illustration, we take Example 2 from Figure 1 and attempts to apply it in the COR model. We can assign the first action of Example 2 to the notion of *request (A,B)* (Number one from the COR model). However, Example 2's second request action cannot be fit into COR's sequential model.

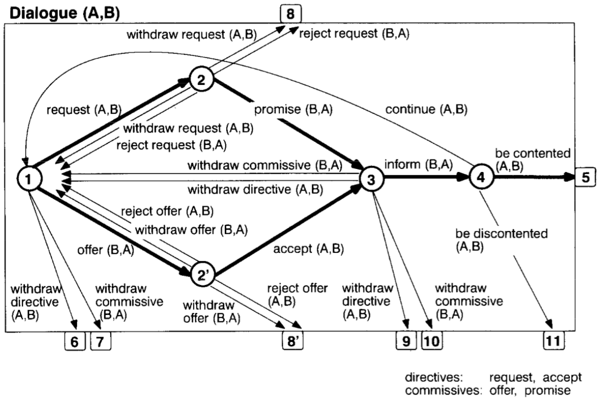

Figure 2: COnversational Roles (COR) model [31].

An alternative approach are scripts from Belkin et al. [10], which are effective interactions between the system and the user on an information seeking strategies (IIS) level [9]. The authors argue that depending on the kind of information need, different interactions may be appropriate. Thus, providing an (ideal) abstraction of the problem allows an understanding of the problem, from which responses (scripts) can be created.

Both the COR model and the scripts enable a form of prediction of which kind of interaction will be necessary following on from a previous move. One could argue that this is a form of advanced slot filling of Spoken Dialogue Systems [22]. Hence, if we could predict and simplify the input given from the user, we may be able to provide appropriate responses generated by the system. Dialogue scripts are a good idea and have worked previously, however, we will need to develop new scripts for this new spoken interaction paradigm.

Other features of the COR model include the flexibility in *mixed-initiative*, meaning that at any given time one of the actors can decide what happens next or ask questions. Mixed-initiative dialogues allow for a more natural interaction but are more complex for the system to handle [21]. The model also allows for meta-communication by permitting the conversation to go through one of the loops at any point in time.

Allowing users to talk freely to a system will come with challenges from a system's perspective since the system will not be able to control the users' input utterances. The aspect of "freedom of speech" means that the system cannot guide or constrain users' options as easily as in a visual setting. For example, it will be more challenging to provide query refinement options or to check whether a user used the right search terms when browsing images. Simultaneously, it may be challenging from a user's perspective if the system provides multiple moves in one utterance.

However, notwithstanding these difficulties, allowing users this "freedom of speech" may encourage an information need expression that more closely represents their real knowledge gap rather than formulating a query for a box. It is important to keep in mind that this freedom of information need expression may be challenging at first for many users who are accustomed to expressing their need in a search box.

The *naturalness of the interaction* with a SCS system can be an aspect of the evaluation measure which is a measure in Spoken Dialogue Systems [19, 39]. We suggest that one of the aspects of this measure could be users uttering multiple moves in one turn. In a human-human interaction this is a behaviour which is observed and expected, and which the other actor can handle. Therefore, allowing users to utter multiple moves in one turn which the system can handle is likely to lead to positive interactions with the system.

3.2 User and System Model or Memory

> *"The overall approach is based on the idea of cognitive models or images that the components of the system have of one another and of themselves"* Belkin [7, p. 111]

We observed users building cognitive models of their partner during the course of the experiment:

User building model of intermediary: Some examples include seekers creating ideas of which actions intermediaries can perform. In one instance, the intermediary offers a function to the seeker by asking if they would like to open a link in a new tab. The seeker now knows that this is an option of the 'system' and later in that session the seeker requests several links to be opened in different tabs. Later in that session the seeker examines the extent of the function by asking ‘‘Could I open the recyclers recycle `uhm in a new tab... if it allows that''` and thus challenges the built intermediary model.

Intermediary building model of user: Other instances were recognized where intermediaries started creating a view of what users may want to hear as output. From the intermediaries we noticed two distinct differences in their utterances. Firstly, intermediaries had formed a cognitive model of *how the information should be presented* to the seeker. For example, through the interaction between the participants, one of the intermediaries was able to form a model of how the seeker preferred to pose queries (this particular seeker posed her queries in a distinctive way with Boolean aspects). As such, the interactions allowed the intermediary to establish a model of how the seeker would form or structure her information and was able to mimic this to satisfy her need.

Secondly, the intermediary had formed a cognitive model about *which information should be presented* to the seeker. In this instance, the intermediary reported names of objects. When the seeker posed another information need, the intermediary checked whether the seeker wanted object names again, even though it was not specified in the seeker's information need. Coincidentally this pair had another search task related to similar objects where the intermediary checked once more whether the seeker wanted the names.

As such, we make a distinction on the system side of how the information should be presented (**form**) and which information should be presented (**content**).

Creating memory over multiple turns/sessions: In this example, the seeker asked for "numbers" (i.e., numerical information) for a particular backstory. In the next task, the intermediary directly asked whether the seeker would like to navigate to the statistics section. This demonstrates an example of creating memory over multiple turns. In an other example, a participant pair had learned from a previous backstory that they could use Google Scholar which the seeker preferred. In the next search task, the intermediary explicitly mentioned that scholarly articles were available for their information need. This demonstrates that memory may be created over multiple sessions as well as multiple turns [26].

Cognitive models are concerned with the cognitive process underlying the search. Much research has been conducted in cognitive IR models [7, 11, 13]. Cognitive models represent search situations at a particular time where the user creates an image of what the system would respond to a particular action. However, the image of this system can change over time. More appropriate to our research is work by Belkin [7] which focuses on the cognitive model of a librarian as an intermediary to a database, particularly at how a librarian forms a cognitive model from the user through dialogue.

We formed the notion that both seeker and intermediary construct *images* or *models* of what the other person can do or which components they have. However, these models are influenced by the seekers' own lens and belief of the world. Similarly Ingwersen and Järvelin [18] mention that a document's author is influenced by their context while the recipient of that document will view it through their lens and belief in their context. Thus, the intended message and the received message may differ. Even though the message itself has been sent across without a noise source [30], the interpretation of the document may vary.

Understanding the cognitive model users form of a system is important for a variety of reasons. For example, understanding what users expect will happen next will allow us to create a system which conforms to the users' model and therefore does not surprise them if something unexpected occur.

We explored the idea of cognitive models of the system; however other ways exist of defining cognitive models [7]. For example, Brooks and Belkin [11] used a reference interview between the intermediary to create a mental model of the user's information need. The idea of using reference interviews to elicit information is not new; however, it could be an interesting "old" approach to a "new" problem. Thus using the conversational interactions would allow us to build a model of the user's information need while utilizing the system's search strategies.

3.3 Decision Offloading and Taking Control

We observed that intermediaries applied many different techniques to deal with the challenge of transferring information through an audio channel. Examples include reading out search results sequentially, summarizing a SERP, or requesting feedback as to whether more information had to be transferred, e.g.

Intermediary: `''Uhm do you want some information about the cinnamon from that company''`
Seeker: `''No that is I think that's enough''`
Intermediary: `''Ah, what else do we need''`

We also noticed that intermediaries became more involved in assisting to express the seeker's information need and taking a leading approach. In the following example, the intermediary *rewrites* the utterance of the seeker into a specific query.

Seeker: `''... cinnamon is from Europe, so I was trying to look uhm is it from Europe or from other places''`
Intermediary: `''I look up cinnamon suppliers... in Europe''`

We observed intermediaries becoming actively involved in trying to satisfy the seeker's information need by making decisions. These observations suggest that intermediaries have a significant role in deciding which information is transferred. The intermediaries are making decisions as to what information is appropriate to share at a given moment. This may also suggest that intermediaries have to make a *cost-benefit* calculation associated to each strategy in order to decide which one would be more likely to benefit the seeker.

These observations corroborate that, given the high cost of delivering information via a linear channel such as speech, it is not optimal to present everything. The system needs to decide which information it should present at each interaction by continuously estimating the satisfaction of the user.

We also observed seekers explicitly requesting the intermediary to make decisions for them, e.g.: `''uhm do you think that should be enough to know where it actually came from or do you think we should carry on''`. It could be suggested that this particular decision offloading example is an artefact of the seeker being aware that there is an intermediary (i.e., human), however, this would be something to explore in a Wizard of Oz setting.

3.4 Effective Information Transfer

Sometimes actors misheard each other (information transfer was not successful) and had to *repair* their conversation. To repair, actors requested a repetition of a previous utterance: `‘‘sorry say that again’’` or `‘‘can you repeat that please’’`. Actors were also observed hesitantly repeating back what the other had said. In other situations, actors interpreted a message incorrectly and were later corrected by the other. Many different instances were noted where the information transfer was disturbed.

The idea of information transfer is a well studied problem. Shannon [30] proposed a model, which shows that a signal can be sent over a noisy channel and the receiver has to construct the signal again with a probability of error (Figure 3). Many researchers have used this model to add probabilities to each of the stages in order to measure the information transfer [12].

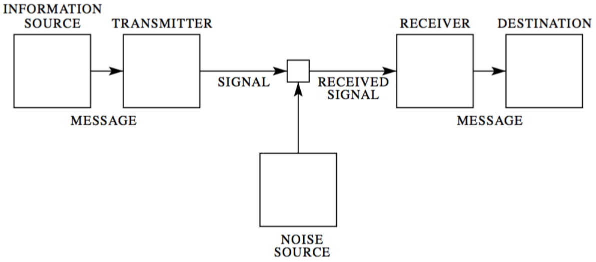

Figure 3: Shannon’s general communication system [30].

Moving away from an explicit form of communication (i.e., typing) allows us to express our thoughts more freely. However, this non-explicit communication is prone to more errors in the transfer of the message, adding an extra layer of complexity to search. Meaning that effectively transferring information becomes an even more important feature of search. We expect that effective information transfer will impact greatly how an SCS system will be evaluated. Therefore, including measures of uncertainty of effective information transfer in evaluation metrics will be beneficial.

3.5 Linking Non-search Related Observations

In this section we provided observations which suggest that the audio channel impacts on the interactions between the user and the system. Interacting verbally increases the flexibility of what users can provide as input, which was illustrated with the observation that one utterance can consist of multiple moves. This flexibility also increases the complexity of the belief regarding what a system or user can do (cognitive user model) as there are no conventional interaction paths. Simultaneously, the responsibility for making decisions could be shared between actors or shifted from one to an other. However, all this is only possible when the information transfer is successful and effective as shown by Shannon’s model [30].

This section also covered some suggested evaluation aspects for SCS, such as measuring the naturalness of a conversation, forming “expected paths” or scripts for cognitive models, or adding a measure of information transfer uncertainty.

4 SEARCH RELATED OBSERVATIONS

We present observations at three stages of the information-seeking process: *Query Formulation*, *Search Result Exploration*, and *Query Reformulation* as defined by Sahib et al. [27]. These stages are equivalent to Marchionini [20]’s *Express*, *Examine*, and *Reformulate*. The model provides broad stages for the collected observations while still providing a structure.

For each of the three stages, we describe the observations and present an analysis of how the observations are explained linking them to existing research and models.

4.1 Query Formulation

We provided the seekers with a backstory for each query, allowing them to verbalize their own information request.[5] In this section we cover the *initial* information requests, i.e., the first iteration of information requests after the user has read the backstory.

Naturalness of Information Request. Participants varied in the way they verbalized their information request: from uttering a query-like expression to describing a detailed and carefully crafted information request. The examples in Table 1 illustrate the range.

Typed queries are usually short [40]. One recent query log analysis by Guy [16] suggested that the average text query length is 3.2 words. Guy [16] also suggested that voice queries are on average longer (4.2 words) stating that the queries are richer in language because they are closer to natural language.

How people formalize a cognitive information gap into a query was modelled by ASK [6]. Once a user has identified a gap, they can start formulating their information need. Taylor [34] proposed four stages of expressing an information need: *Visceral, Conscious, Formalized, and Compromised.* Firstly the need for information (visceral) is formed and its mental description emerges (conscious). The two last stages involve expressing the need (formalized) and then formulating it in a way which can be presented to a search engine (compromised).

Many different expressions of information requests were observed that did not conform to the typical textual query. These information requests included natural language requests, instructions, or additional information to the original information request (Table 1). It could be argued that some of these complexities are observed because users are not restricted to a typical search box and do not have to translate their thoughts into queries as was suggested by Taylor [34]’s stages of information need. We suggest that an SCS information request often will not go through the four stages of information need [34]. Instead SCS information requests will be uttered before they have conformed to textual queries.

Other observations include users wanting to spell keywords in their queries or use advanced search mechanisms such as Boolean syntax. Note that in audio-only settings, allowing spelling may be an important feature, given that typing or copying/pasting keywords are not (or hardly) available.

[5] We use the notion of information request because these expressions were often not precise queries but more an explanation of what the users were looking for.

Table 1: Example initial information request utterances.

Example utterance	Characteristic
‘‘Turkish river control’’	Query-like
‘‘Which jobs from the United States have been outsourced to India’’	Natural language type query
‘‘So the count part in uhm a biscuits that you are get from Europe uhm it contains cinnamon and I want to know where the cinnamon is coming from are there is this uhm is this coming from Europe uhm so how to uhm search for uhm cinnamon Europe biscuits’’	Query babbling [24]
‘‘Maybe start of with uhm type in the origins of cinnamon’’	Instructions plus query-like
‘‘Can you please search car tyre recycling [long pauze] and in the results I am looking for examples of what uhm recycled car tyres are used for’’	Instructions plus query-like plus additional information on what to look for in the results (step-wise information request revealment)
‘‘Have Turkish river control projects affected Iraqi water resources [long pauze] so we are looking for if dams or irrigation schemes have affected uhm any of the Iraqi people’’	Natural language type query plus additional information on what to look for in the results (step-wise information request revealment)
‘‘Uses for old car then the query or, passenger vehicle tyres TYRES (user spells tyres) or in caps tires TIRES (user spells tires) ... and I wanna uhm do a date range so the data is from a recent twelve months, so uses for old car caps or passenger vehicle or tyres TYRES (user spells tyres) caps or tires TIRES (user spells tires) and data in the last twelve months that's the query’’	Detailed and carefully crafted information request (teleporting [35]) plus utilizing extra features such as date range from the system

4.2 Search Results Exploration

In the previous stage (Query Formulation) we investigated the first action of the user. In this section we investigate the interactions between the user and the intermediary after this initial utterance.

We investigate the concept of the boundaries between the SERP and the documents. We then cover how both user and intermediary are actively involved in the relevance judgments, followed by what happens when previously encountered results are seen. Finally, we investigate how graphical information can be useful in an audio-only setting.

SERP and Document Boundaries. In traditional IR, the SERP and documents linked from the SERP are thought of as quite different entities. In an SCS system, the differences faded for several seekers during their search.

There were instances where seekers asked intermediaries to access a particular document assuming the intermediary was reading from the SERP. However, the intermediary was already reading from the document in the previous turn without the seeker realizing this –which could be referred to as ‘non-hyperlink click’ [40]. In other instances, the seeker asked for clarification about information on the SERP thinking the intermediary was reading a document. The lack of visual feedback was a major aspect. As identified earlier, the cost of effective information transfer increased and it may be beneficial for transparency for the seeker to indicate when something is hyperlinked or not.

We also observed intermediaries providing an overall summary of the SERP or document. Some of these summaries covered aspects of multiple documents without the intermediary indicating this to the seeker. This may suggest that incorporating *multi-document summarization* [5] may be beneficial in transmitting information in an audio-only search setting.

The idea of a SERP (the tool) and the document (the goal) is not distinctively presented in an audio setting.[6] The notion of fading boundaries between the "tool" to get to the relevant document may introduce different cost benefits for the user depending on whether they want to listen more to that particular document. Removing this boundary and provide better integration between the system and the document may have profound impacts on how people perceive "searching" since they may not have to deal with either documents or search engines.

Explicit Relevance Feedback. Relevance feedback allows searchers to provide implicit or explicit feedback about relevant information and these judgments may enhance subsequent searches [28]. Implicit relevance feedback is where users' interactions with the SERP are recorded and integrated in the search. Explicit relevance feedback is where users provide clear feedback on the relevance of items.

Researchers have made the assumption that when a user does not engage with a search result, then that particular result may be irrelevant to the user, or the relevant part is displayed in the SERP. However, in our observations we noted users were actively involved in both rejecting and accepting results and therefore provided explicit relevance feedback.

In a spoken search environment, we observed that explicit feedback was provided by users without prompting them. For example a seeker provided positive feedback by saying: ‘‘Yeah I think yeah that actually sounds pretty good, that could potentially be relevant, is there anything else or is

[6] Keeping in mind that search engines now provide cards on the SERP which have become often the goal.

that it?’’. We also observed utterances which may be interpreted as negative relevance feedback: ‘‘OK alright that’s probably not relevant then so yeah we wanna just find something actually where does the spice cannanon [sic.] cinnamon come from’’.

Users were not forced in any way to provide relevance feedback in our experiment; however, they provide it nonetheless. Incorporating such feedback may lead to better performance of the spoken search system and may reinforce users to provide more relevance feedback. We observed that the users who provided relevance feedback and received responses from the retriever provided relevance feedback more often.

Novel vs. Previously Seen Information. Changes in link colour are used to indicate whether a particular link on a SERP has been clicked before or not. The change provides feedback to users on whether they have visited the underlying document, reducing their memory load. We observed several groups indicating that the same search results were displayed: e.g., an intermediaries would state ‘‘I keep on getting the same [search result]’’ or ‘‘we’re back to that [search result] again’’.

Observations suggest that information about whether search results have been already visited or not is also important in an audio-only setting. However, in an audio-only setting this may be more difficult, given that providing visual feedback is not possible.

Interpretation of Graphical Information. Graphical information such as images, charts, or videos are for the majority of search engine users accessible. However, in an audio-only setting accessing graphical information is more challenging. This problem was also observed in previous studies among people with a visual impairment. Abdolrahmani and Kuber [1] indicated that images without description would be inconvenient for people with screen-readers and would lead to increased cognitive load. In our study intermediaries interpreted images and graphs in order to convey the presented information. Most of the interpretations were made of images and graphs in a document. However, we also observed another interpretation of images whereby the intermediary navigated to the image tab on the SERP in order to quickly gather insight of an object which she then described to the seeker. Thus, graphical information will need descriptive information in order to allow for the full potential of audio-only systems.

4.3 Query Reformulation

Automated Repetitive Search. To save time and effort, people try to find ways to automate repetitive tasks into batches (e.g., defining macros) which saves them time and effort instead of performing each task individually. We observed instances of this notion of “automation” during the conversational search setup. One pair wanted to find more information about the health benefits of eating seaweed. The seeker had different seaweed in mind that she wanted to look up and therefore created a short query loop for these different kinds of seaweed as illustrated in Algorithm 1.

Another pair created a repetitive search task with multiple conditions. The seeker wanted to investigate rivers in Turkey and Iraq before searching for dams among those rivers. For each river that had a dam, the seeker wanted to know the construction date and water volume. The example is given in Algorithm 2.

Algorithm 1: Automated Repetitive Search (Seaweed)

```
Result: Which are the health benefits of different seaweeds
foreach Seaweed do find health benefits;
else
    Seaweed not relevant to search
end
```

Algorithm 2: Automated Repetitive Search (Rivers Turkey)

```
Result: Did Turkish river control projects affect Iraqi water
        resources
foreach River in Turkey and Iraq do
    if They have a dam in Turkey then
        if Building date of dam and volume is stated then
            Compare river’s volume in Iraq before and after
              building of the dam
        end
    end
end
```

It could be suggested that seekers had made a plan before starting the search of what their search path was going to be or had formed a model of the intermediary’s capabilities. These two examples could be seen as one way of “taking control” over the search interactions as explained in Section 3.3. The seeker has set out a clear path of how they want to search without handing over any decision making responsibilities to the intermediary.

Information Requests Within a Document. We already observed different behaviour in posing initial information requests (Section 4.1), where seekers provided their information need in two steps. First they presented a query-like utterance and then enriching the utterance requesting supplementary information. In addition to providing further information in that initial turn, we also observed seekers providing an information request once they navigated to a SERP/document. Here, seekers requested information about the document that was being inspected by referencing to the given backstory or pieces of information within the document.

In some cases, seekers requested information within the navigated SERP/document with reference to the given backstory, thus, revealing their information need in step-wise fashion.

Seeker: (Initial information request) ‘‘Health benefits of marine vegetation’’

Intermediary: ‘‘... It just says a lot of comparing and uhm there are some articles that start to talk about like uhm plankton plants and stuff’’

Seeker: ‘‘Uhm do some articles mention the use of marine vegetation as a drug as in like in medicine’’

In other cases, intermediaries presented some information from the given document and seekers wanted to know more about a certain entity given in that document.

Seeker: ``Does the data uhm illustrate per capita consumption... by country?''

Intermediary: ``Uhm... the first column... OK this is the list of countries by alcohol consumption measured in equivalent litres of pure ethanol consumed per capita per year''

Seeker: ``Fantastic.. please read out the top ten''

Intermediary: ``Uhm Belarus, Moldova, Lithuania, Russia, Romania, Ukraine, Andorra, Hungary, Czech Republic''

Seeker: ``Where is Australia in the list?''

The notion of "within-document" retrieval is not new and often used by people with the "find" *(Control+F)* function [17]. However, this find function is embedded in the browser and is not part of the search engine. The integration of different search related aspects such as the find function may be important in SCS.

5 SUMMARY

In Sections 3 and 4 we investigated the impact of the audio channel on the interactions between user and system during a search. We also discussed new research opportunities which are a result of the different mode of information transfer (i.e., the audio channel). The observations suggest that SCS has the following *increased complexity and interactivity* between system and user.

Increased complexity. Verbal communication is a major aspect of interacting with this new search paradigm. Since results are no longer displayed but sent through an audio channel which is non-persistent, the complexity of the interaction increases immediately. However, not only is the channel (audio) challenging, but what goes into that narrow channel also becomes increasingly complex. For example, a user is not confined to a search box and can freely express what the system should perform. Our study suggests that the complexity of a system increases by allowing users to express their needs more naturally, for example, by specifying multiple moves in one utterance, uttering non-specified needs, or providing feedback throughout the interactions.

Results also suggested that systems need to make more decisions in this new search paradigm. This increases the complexity of the system and simultaneously the complexity of what users and systems expect the other actor may perform. This then leads to more complicated user- and system models.

Increased interactivity (collaboration). Even though there is an increase in complexity with this new search paradigm, the paradigm also provides new opportunities. Since all results are presented in audio, the idea of static boundaries between the SERP and the documents appears to fade. At the same time, since the user and the system are actively involved in a conversation, this discourse could be used to extract the information need from the user. On the other hand, the user can request, in a more natural manner, information from within a document directly from the system. Thus, integration between search engine and document is important.

Conversation (i.e., interaction) and collaboration are crucial to communicate messages, such as interpreting photos, indicating that documents have been seen before, or explaining the information need. The willingness to collaborate and structure a conversation will be crucial in providing a satisfactory search interaction.

We now discuss the future vision and impact of involvement these search systems may have in the users' search process.

5.1 Existing Search Behaviour Models and SCS

We had aimed to form a clear view of whether any existing information seeking models fit SCS. However, to our knowledge many well-known models such as Belkin's ASK [6] or Marchionini's ISP [20] do not include the system's "responsibility" of interacting with the user. An exception is Saracevic's stratified model [29].

Other models, such as Sitter and Stein's COR model [31] or Belkin's scripts [10], encompass the interaction between two actors. However, these models lack the flexibility of the speech aspect such as multiple moves in one turn. We believe that novel models could be necessary for this new interaction paradigm, allowing for the development of new hypotheses which can be tested to inform the audio search paradigm [12].

5.2 Important Aspects of SCS

We suggest that SCS systems will become more actively involved with the users' search process. This involvement is needed to overcome the imposed complexity of the audio channel. System and user will rely more on aspects such as verifying effective information transfer through feedback. Thus, SCS will require interactivity from both actors while they collaborate on the shared task. Simultaneously, this dialogue allows for supporting and structuring the search from the system while allowing the user freedom to express their needs and wants more naturally.

SCS allows for progressing from an "action-response" search paradigm to a paradigm which has shared responsibilities between actors to succeed in the task. In other words, users have to share their information need and ideally provide direct feedback to the system. Simultaneously the system will have to become more actively involved in deciding which results to present in a narrow audio channel.

SCS allows for single- and multi-participatory search with the system which is similar to collaborative search as previously researched by Evans and Chi [15] or Morris [23]. However, it is widely accepted that people communicate and behave differently between human-human and human-machine communication [14].

5.3 System Differences

The results of this study suggest that SCS is more complex than a "traditional" IR system. Overall, we could argue that we are moving towards a search process where the system is more involved with the users' search process as a whole. The following differentiations can be made:

Passive System. The traditional search system where users have full control over the interactions and decisions. These search systems have many different added components and make decisions for ranking results, query suggestions, or spelling suggestions. Nevertheless, these search systems still leave the majority of the decisions with the users. Simultaneously, not all users (e.g., users with a visual impairment) can make use of these additional features [27]. In addition, the initiative taken in a *Passive System* comes mainly from users. For example, a user submits a query and can resubmit queries, however, the system has limited capabilities to interact with the user in order to elicit the information need. The idea of the one-action search paradigm (action-response) is very much ingrained.

Active System. Search systems become more active due to their involvement with the user, thus shifting away from the passive system as described above. We observed that the interaction between the system and the user becomes important in systems which are based on auditory information transfer. Thus, the system and user have to be engaged with each other in order to effectively transfer information. Simultaneously, the user is not confined to predefined actions (query, mouse-clicks, or pressing enter) and can express their desires more freely. Which means that the system can generate multiple responses to a given action and create a common ground for collaboration. In other words, the "passive" system is becoming more active in the interaction with the user.

Pro-active System. Up till now, the user has initiated a search. The next search paradigm are systems which are actively involved without users having to start the search. Instead, a "Search Engine That Listens (SETL)" could be a system that continuously monitors and listens what the the user does. This way the system can identify information needs/tasks and pro-actively provide content which could support users while satisfying their information need or completing a particular task.

As presented in Figure 4, the essence of search (the Passive System) is not going to change. The idea of posing an information need (by explicitly posing, query extraction through dialogue or extraction through listening) and presenting information will stay. However, it is how users will interact with these systems that is going to change.

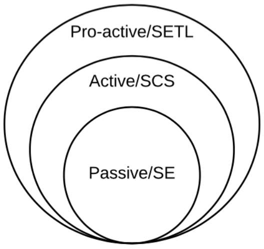

Figure 4: From a passive IR system to a pro-active Search Engine That Listens (SETL).

Examples of Passive and Active Systems. In Figure 5 the information need is expressed in a "search box query" by the seeker. The system ranks all the documents and presents the highest ranked document to the user. As suggested by Taylor [34], the user goes through stages of forming this information need, whereby the last stage is to create a query reflecting their cognitive need.

The observations in this study suggest that people are not expressing their information need exclusively through a "search box query". Instead, users express their need through a more natural language statement. Thus, seekers can benefit from the audio channel to present their information need.

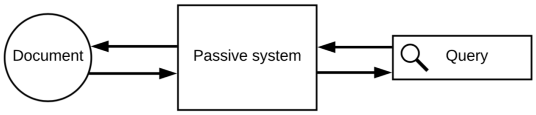

Figure 5: Passive IR with activated components of information need expression.

Simultaneously, the audio channel could also be an advantage on a system level. As suggested by the observations, the boundaries between the SERP and the documents become vaguer. For example, the system could utilize this aspect by not just presenting the highest ranked document, but by generating a summary of similar information in many different documents (multi-document summarization [5] as discussed in Section 4.2) (see Figure 6). Thus, the system would integrate technologies, both existing and non-existing, to create a more advanced interactive search system. Therefore we could suggest that in order to fulfil the difficulties of the audio channel the system may have to become more active and more strongly involved in the users' search process.

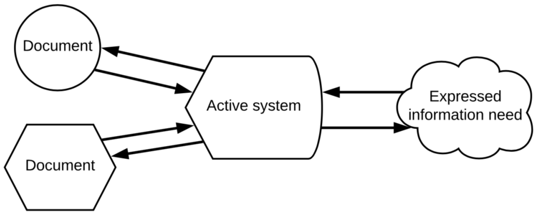

Figure 6: Active IR, combining multiple documents as one representation for the user.

We illustrate that a search system can become actively involved in the user's search process. During this process users may transfer the control they possess in this process. However, both in the passive and active system examples, the system will not be able to act autonomously since the information need may not be completely transferred to the system. This means that we may only be able to have an autonomous search system once the information need extraction can be automated.

6 CONCLUSIONS

This paper explored SCS, an emerging interactive search paradigm wherein all interactions are performed through audio. We conducted an observational study which showed that new information seeking models are needed for SCS. We concluded that this new paradigm is much more complex and interactive than the search scenarios/paradigms covered by existing models. We also suggested several new research opportunities, illustrating that this new search paradigm provides the opportunity to bring together many technologies which have been created into a single integrated model. One limitation is that our results are based on human-human interaction that simulates an ideal situation with search tasks which were designed for a textual setting. We plan to extend our observational experiment into a Wizard of Oz setting in order to understand whether our findings still hold and capture information needs which people would like to solve with this system. Nevertheless, to our knowledge, this is a first major study to provide insight into what SCS may look like.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments to improve this paper and the participants who took part in the study. This research was partially supported by Australian Research Council Project LP130100563, Real Thing Entertainment Pty Ltd and Microsoft Research Asia under MSR CORE-12 Project.

REFERENCES

[1] A. Abdolrahmani and R. Kuber. Should i trust it when i cannot see it?: Credibility assessment for blind web users. In *Proc. of ASSETS'16*, pages 191–199. ACM, 2016.

[2] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. In *ACM SIGIR Forum*, volume 46, pages 2–32. ACM, 2012.

[3] J. Allen and M. Core. Draft of DAMSL: Dialog act markup in several layers, 1997.

[4] P. Bailey, A. Moffat, F. Scholer, and P. Thomas. User variability and IR system evaluation. In *Proc. of SIGIR'15*, pages 625–634. ACM, 2015.

[5] R. Barzilay, K. R. McKeown, and M Elhadad. Information fusion in the context of multi-document summarization. In *Proc. of ACL'99*, pages 550–557. ACL, 1999.

[6] N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1):133–143, 1980.

[7] N. J. Belkin. Cognitive models and information transfer. *Social Science Information Studies*, 4(2-3):111–129, 1984.

[8] N. J. Belkin. A methodology for taking account of user tasks, goals and behavior for design of computerized library catalogs. *ACM SIGCHI Bulletin*, 23(1):61–65, 1991.

[9] N. J. Belkin, P. G. Marchetti, and C. Cool. BRAQUE: Design of an interface to support user interaction in information retrieval. *Information processing & management*, 29(3):325–344, 1993.

[10] N. J. Belkin, C. Cool, A. Stein, and U. Thiel. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Syst. Appl.*, 9(3):379–395, 1995.

[11] H. M. Brooks and N. J. Belkin. Using discourse analysis for the design of information retrieval interaction mechanisms. In *ACM SIGIR Forum*, volume 17, pages 31–47. ACM, 1983.

[12] D. Case. *Looking for information. A survey of research on information seeking, needs and behavior.* Emerald Group Publishing Limited, 2012.

[13] P. J. Daniels. Cognitive models in information retrieval—an evaluative review. *Journal of documentation*, 42(4):272–304, 1986.

[14] L. Dybkjaer, N. O. Bernsen, and W. Minker. Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, 43(1):33–54, 2004.

[15] B. M. Evans and E. H. Chi. An elaborated model of social search. *Information Processing & Management*, 46(6):656–678, 2010.

[16] I. Guy. Searching by talking: Analysis of voice queries on mobile web search. In *Proc. of SIGIR'16*, pages 35–44. ACM, 2016.

[17] D. J. Harper, I. Koychev, Y. Sun, and I. Pirie. Within-document retrieval: A user-centred evaluation of relevance profiling. *Information Retrieval*, 7(3):265–290, 2004.

[18] P. Ingwersen and K. Järvelin. *The turn: Integration of information seeking and retrieval in context*, volume 18. Springer Science & Business Media, 2005.

[19] D. J. Litman and S. Pan. Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction*, 12(2):111–137, 2002.

[20] G. Marchionini. *Information seeking in electronic environments.* Number 9. Cambridge university press, 1997.

[21] M. McTear, Z. Callejas, and D. Griol. The conversational interface. *New York: Springer*, 10:978–3, 2016.

[22] M. F. McTear. Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, 34(1):90–169, 2002.

[23] M. R. Morris. Collaborative search revisited. In *Proc. of CSCW'13*, pages 1181–1192, 2013.

[24] D. W. Oard. Query by babbling: A research agenda. In *Proc. of CIKM'12*, pages 17–21, 2012.

[25] R. N. Oddy. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1):1–14, 1977.

[26] F. Radlinski and N. Craswell. A theoretical framework for conversational search. In *Proc. of CHIIR'17*, pages 117–126. ACM, 2017.

[27] N. G. Sahib, A. Tombros, and T. Stockman. A comparative analysis of the information-seeking behavior of visually impaired and sighted searchers. *Jour. of the Amer. Soc. for Inf. Sc. and Tech.*, 63(2):377–391, 2012.

[28] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[29] T. Saracevic. The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the Annual Meeting-American Society for Information Science*, volume 34, pages 313–327, 1997.

[30] C. E. Shannon. *The mathematical theory of communication.* University of Illinois press, 1949.

[31] S. Sitter and A. Stein. Modeling the illocutionary aspects of information-seeking dialogues. *Information processing & management*, 28(2):165–180, 1992.

[32] A. Stein and E. Maier. Structuring collaborative information-seeking dialogues. *Knowledge-Based Syst.*, 8(2-3):82–93, 1995.

[33] A. Stein, J. A. Gulla, and U. Thiel. User-tailored planning of mixed initiative information-seeking dialogues. *User Modeling and User-Adapted Interaction*, 9 (1-2):133–166, 1999.

[34] R. S. Taylor. The process of asking questions. *Journal of the Association for Information Science and Technology*, 13(4):391–396, 1962.

[35] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proc. of CHI'04*, pages 415–422, 2004.

[36] P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. MISC: A data set of information-seeking conversations. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017.

[37] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. How do people interact in conversational speech-only search tasks: A preliminary analysis. In *Proc. of CHIIR'17*, pages 325–328. ACM, 2017.

[38] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. A conversational search transcription protocol and analysis. In *Proc of SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, CAIR '17, 2017.

[39] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella. PARADISE: A framework for evaluating spoken dialogue agents. In *Proc. of EACL'97*, pages 271–280. ACL, 1997.

[40] R. W. White. *Interactions with search systems.* Cambridge University Press, 2016.

[41] T. D. Wilson. Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999.