Report on the Future Conversations Workshop at CHIIR 2021

Damiano Spina*
RMIT University
damiano.spina@rmit.edu.au

Johanne R. Trippas*
The University of Melbourne
johanne.trippas@unimelb.edu.au

Paul Thomas*
Microsoft
pathom@microsoft.com

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

Katriina Byström† Leigh Clark† Nick Craswell† Mary Czerwinski† David Elsweiler† Alexander Frummet† Souvick Ghosh† Johannes Kiesel† Irene Lopatovska† Daniel McDuff† Selina Meyer† Ahmed Mourad† Paul Owoicho† Sachin Pathiyan Cherumanal† Daniel Russell† Laurianne Sitbon†

Abstract

The Future Conversations workshop at CHIIR'21 looked to the future of search, recommendation, and information interaction to ask: where are the opportunities for conversational interactions? What do we need to do to get there? Furthermore, who stands to benefit?

The workshop was hands-on and interactive. Rather than a series of technical talks, we solicited position statements on opportunities, problems, and solutions in conversational search in all modalities (written, spoken, or multimodal). This paper –co-authored by the organisers and participants of the workshop– summarises the submitted statements and the discussions we had during the two sessions of the workshop. Statements discussed during the workshop are available at https://bit.ly/FutureConversations2021Statements.

1 Introduction

The Future Conversations'21[1] workshop co-located at CHIIR 2021[2] aimed to provide a venue for new ideas, controversial statements, or gaps in the conversational search research. The format

*Organizers
†Authors and participants
[1]https://sites.google.com/view/cair-ws/future-conversations21
[2]https://acm-chiir.github.io/chiir2021/

followed to prepare and run this virtual workshop was rather unusual (Figure 1). The workshop invited researcher's opinions in the form of a short statement submitted beforehand. These statements were then commented on by participants who signed up for the workshop. Authors were asked to submit their statements which were made publicly available on Google Drive to all workshop attendees prior to the workshop. Everyone was encouraged to read the statements and leave comments. These comments were part of the discussion. The aim of this statement elicitation was to start the conversation before the workshop and allow people to formulate responses to the statements.

The workshop was held online at two different time-slots to support participation from different time zones. The workshop ran twice for two hours and was driven by the discussion of the statements by the participants. Nearly all authors of the submitted statements were able to attend both sessions and thus support active discussion.

The workshop included a welcome and brief introduction, then every statement was introduced followed by a discussion. We discussed the statements based on overlapping themes. The order of the statements were reversed for the second workshop to help the diversity of the conversation. Inspired by the model used in the FACTS-IR workshop at SIGIR 2019 [Olteanu et al., 2021], the workshop attendees were informed that participation in the workshop would lead to a SIGIR Forum submission. All who actively participated during the workshop and write-up after the workshop are co-authors of this publication.

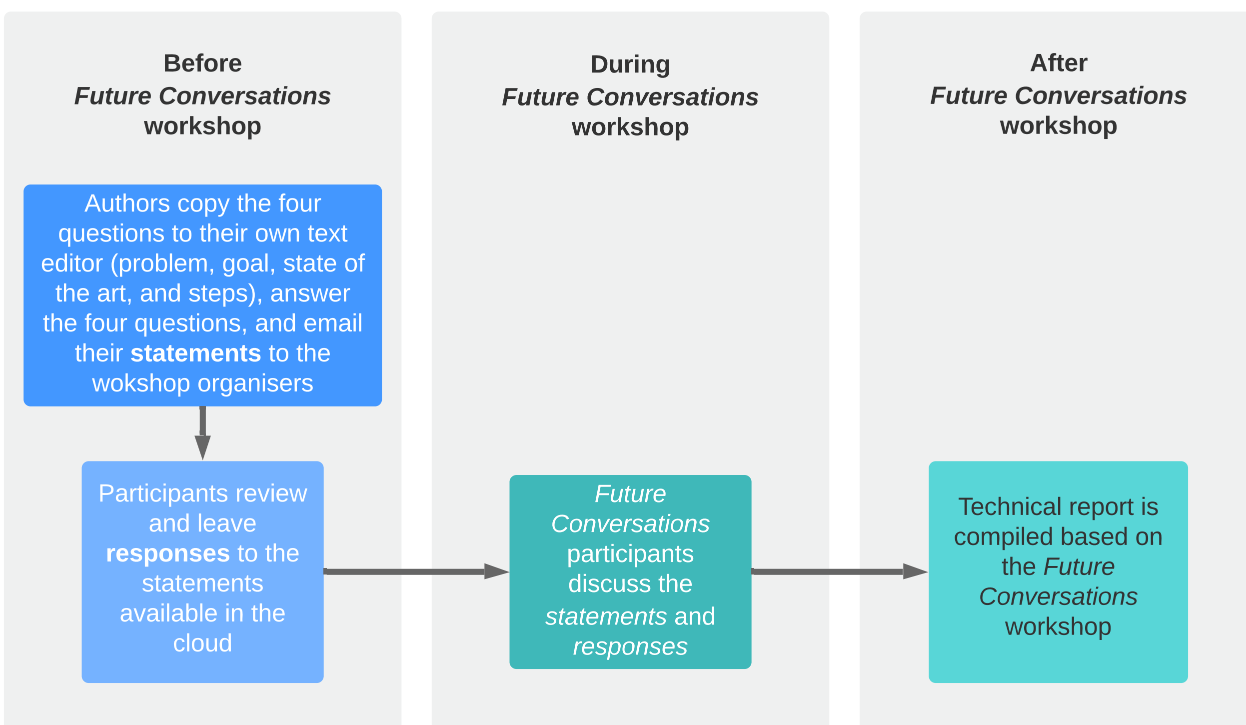

Figure 1: Workflow of the activities carried out before, during, and after the workshop sessions.

The CAIR'20 workshop proposed and accepted for ACM CHIIR 2020 –which was cancelled due to the COVID-19 pandemic– did not have a session at the Virtual CHIIR 2020 event organised in August 2020. Thus, the Future Conversations'21 was proposed as a virtual workshop for ACM CHIIR 2021. The original format planned for CAIR'20 was not suitable and therefore the organisers experimented with the format described above.

2 Statements

Statements and responses were collected from the participants discussing challenges and opportunities to advance knowledge in conversational IR. We provided the participants with clear guidance on how a short statement could look like to support statement submissions which included the following:

A statement consists of about two paragraphs addressing each of the following four points:

1. Problem:
    - Who is it that has a problem, or an opportunity, with conversational IR? Please tell us something about a person, group of people, or a situation. (Examples: blind users; domestic situations; hospitals; web search companies; automotive industry; speakers of minority languages.)
    - What are the problems or opportunities there? Please give as many examples and as much detail as possible – at least a paragraph. (Examples: interfacing with screen readers and other existing tools; tracking social contexts; removing touch points; managing long-form interactions; understanding attention; voice input and output for low-resource languages.)
    - Why is the problem important and timely?
2. Goal:
    - What should the research community be doing to help? Please give as many examples and as much detail as possible.
3. State of the art:
    - A brief overview of where we are at now, in production systems or research.
    - What do we already know? What are the one or two key resources we already have?
4. Steps:
    - List any resources, collaborations, research goals, or anything else that may help along the way.

The following eight statements were submitted:

1. Including patient context and physicians-in-the-loop to create transparent health conversational systems (Johanne R. Trippas, The University of Melbourne)

2. Conversational search in the domain of cooking (Alexander Frummet, University of Regensburg)

3. Arguing with search engines (Johannes Kiesel, Bauhaus-Universität Weimar)

4. Companionship features in intelligent personal assistants (Irene Lopatovska, Pratt Institute)

5. Mixed-initiative for conversational search (Paul Owoicho, University of Glasgow)

6. Using speech in conversational search: considering people with diverse speech patterns (Leigh Clark, Swansea University)

7. Future *fair* conversations (Damiano Spina, RMIT University)

8. Social competence in conversational retrieval (Paul Thomas, Mary Czerwinski, Daniel McDuff, and Nick Craswell, Microsoft)

In the following sections we briefly describe each statement, as well as a summary of the discussions held at the two sessions of the workshop. The original statements submitted are available at https://bit.ly/FutureConversations2021Statements.

2.1 Including Patient Context and Physicians-in-the-Loop to Create Transparent Health Conversational Systems

Johanne R. Trippas, The University of Melbourne

Problem. Seeking advice on search engines for health reasons can help understand which medical services are available or where they are located. Thus, search engines can act as a gateway or platform to guide patients to professional medical services. However, when consulting "Dr Google" for health diagnoses, it can become difficult for searchers to decide whether the found information is credible, accurate, and relevant to their health issue.

Existing issues with searching for health information through search engines include *(i)* user's non-medical query can provide limited information for symptoms, *(ii)* the context of the patient is removed (e.g., health history), *(iii)* no expert judgement from health professionals is included, and *(iv)* people are searching for highly specialised information (i.e., jargon) which may mislead the patient and cause anxiety.

Goal. A conversational system in a health setting must see the human as a holistic "set of data points" (i.e., treatment of the whole person, including mental and social factors, rather than just the symptoms of a disease). The research community can investigate how to *(i)* include the patient's context in a conversational system such as multimedia and multi-device input (i.e., analysing video our sound features) [Deldjoo et al., 2021], *(ii)* create a physician-in-the-loop system to overcome a system's black-and-white symptoms-based decision-making approach.

State of the Art. Many chatbots have been developed to help automate and self-diagnose patients. However it has been suggested that these systems lack the functions of the complete diagnostic process of an in-person visit [Vaidyam et al., 2019]. Furthermore, smart devices have been incorporated, such as Amazon Echo to help identify cardiac arrests through background sounds [Chan et al., 2019]. Other research has suggested that mobile devices' non-intrusive sensors could be utilised to identify illnesses [Quatieri et al., 2020].[3]

Steps. Our research community should increasingly work closely with health professionals.[4] This will help to navigate and understand evidence-based health decision-making complexities and help create systems which include the "patient as a whole". Furthermore, we should aim to create systems which make the conversational search methods transparent to conversational search methods transparent to patients and health professions and communicate the triaging process. More information on the differences in health-service providers (i.e., chatbots or trained physicians) are needed.

2.1.1 Discussion

Our discussion on searching for health information through conversational interactions first focused on how to effectively and efficiently elicit symptoms and guide the user through possible diagnoses. For example, existing websites which offer "symptom checkers"[5] direct the user to the appropriate healthcare action by asking questions about their health issue. These checkers are built on heuristics to conclude "what to do" (i.e., self-care, talking to a health professional, or go to an emergency department), "where to go" (i.e., local healthcare services based on the "what to do" advice), and "more information" about the identified symptoms. We discussed whether conversational systems could learn something on how to elicit information from such checkers.

Nevertheless, when people are in a stressful context (e.g., when a person is unconscious or not breathing), calling an ambulance service to help to diagnose the issue is more desirable. Even though human call takers are often restricted to rigorous protocols as a way of "symptom checking", these call takers can also incorporate and deal with "human" input. For example, conversational systems may have issues with automatic speech recognition when people are distressed, cannot integrate crucial background noises, or fail to locate the callers due to inadequate descriptions of where the patient is located.

Questions around the following topics were discussed:

- How well do symptoms checkers perform versus a search engine [Cross et al., 2021]?
- Will people trust conversational systems on sensitive health topics in time-critical contexts?
- How can a conversational system express comfort in life or death situations?

2.2 Conversational Search in the Domain of Cooking

Alexander Frummet, University of Regensburg

[3] https://news.mit.edu/2020/signs-covid-19-may-be-hidden-speech-signals-0708

[4] https://www.premier.vic.gov.au/cutting-edge-technology-helping-paramedics-save-lives/

[5] https://www.healthdirect.gov.au/symptom-checker

Problem. While current speech-only conversational assistants perform quite well in short few-turn dialogues like setting a timer, getting current weather information, or pointing to recipes on the web, they fail when longer conversational flows occur. Nevertheless, if conversational assistants improve such that they can understand longer conversations, they might be more helpful than simply directing users to a recipe. Having a reliable guide escort you through the cooking process might, for example, nudge more people to cook something for themselves. If systems could guide and assist users during the cooking process, this could have several advantages: Users who cook for themselves tend to live more healthy [Hartmann et al., 2013]. Also, systems may be able to "nudge" users to select healthier recipes when they converse with them [Elsweiler et al., 2017]. Considering the strong increase of obese people worldwide,[6] nudging can be one means among many to mitigate this problem and implement a healthier lifestyle [Wilson et al., 2016].

Goal. Creating a usable conversational search system in the domain of cooking remains a considerable challenge. In previous research focusing on cooking, we found that people experience multiple and diverse information needs [4]. Some of the utterances people make to describe these needs, such as queries about ingredient amounts, require less context to be understood. Others, like questions about specific cooking techniques, on the other hand, require more context to be understood. Since it is crucial for the success of preparing a meal to understand which information the user seeks from a system and to provide proper assistance, the research community needs to investigate (1) how context information can be included in spoken conversational search systems and (2) how suitable assistance can be given to the user depending on the type of need prevalent.

State of the Art. Investigating and predicting information needs in a conversational setting is not a new idea. Previous research has focussed on predicting broad and domain agnostic information needs [Shiga et al., 2017; Qu et al., 2019]. Also, efforts have been made to assist users during the cooking process [Martins et al., 2008; Nouri et al., 2020]. These studies, however, were not performed in a spoken conversational search environment.

Steps. To establish the information needs that occur in a cooking context, determine if these can be detected automatically and resolved by providing appropriate assistance, a corpus of free-form spoken conversational search data is available which was collected in an in-situ study [Frummet et al., 2019]. User utterances in this corpus have been annotated with respect to information needs. Furthermore, these utterances were assigned to recipe steps to include more information about the context. Based on this, we need to find means to include context information and assistance strategies to get one step closer to a truly conversational search system in the domain of cooking.

2.2.1 Discussion

The discussion at the workshop focussed on "context tracking" in mixed-initiative conversational search dialogues for cooking. Different contexts were described as external (i.e., the user and their surroundings) and internal/local (i.e., the context within the conversation or task). External

[6] https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

context can include video graphics information to understand the ingredients at hand or eye-tracking data to understand where the user is looking. Such information can be relevant for a better understanding of the utterances and thus inform the next step in the cooking process. The internal context was discussed to personalise the linguistic cues to different users with varying cooking levels/knowledge. Additional challenges include, assigning speech requests to the task that is currently running, the difficulty of explaining the "touch" of an object or ingredient (i.e., how something is supposed to "feel") [Nonaka and Takeuchi, 1995].

Apart from tracking external and internal context in future studies, clarifying questions were recommended to solve context problems as a first step. As a starting point for generating such questions, transcripts of culinary shows where participants ask questions to the chef can be employed.

2.3 Arguing with Search Engines

Johannes Kiesel, Bauhaus-Universität Weimar

Problem. Life is full of choices, each of which has its pros and cons, often touching subjective experiences like morality, faith, or meaning. In such situations –from political controversies to personal struggles– it is often helpful to discuss with someone else. And wouldn't it be great to discuss with someone that is emphatic and sceptical like a mentor, knowledgeable like everyone on the Web combined, yet approachable and patient like a machine?

Goal. This someone is a search engine one can argue with. As the fundamental difference to regular web search, arguing is about understanding different points of view. Among others, this implies that occasionally one may want the search engine to be explicitly biased against one's own opinion.

State of the Art. Current search engine interfaces are ill-equipped to assist in such situations due to their focus on direct answers [Potthast et al., 2020]. Argument search engines [Wachsmuth et al., 2017] exist but do not yet allow for conversations [Kiesel et al., 2020]. While IBM's Project Debater achieved impressive results for formal discussions [Slonim et al., 2021], arguing as proposed here requires a computational understanding of ideologies and personalities that is still in its infancy [El Baff et al., 2020]. Proposals for evaluation criteria of argument search engines exist [Potthast et al., 2019] but likely need adaptation for the arguing scenario.

Steps. To allow for a conversation, more research is needed on grouping [Bar-Haim et al., 2020] and summarizing [Alshomary et al., 2020] arguments as well as on quantifying, explaining, and controlling the various biases of result (sub-)sets [Kulshrestha et al., 2019]. To ensure the conversations' effectiveness and appropriateness, the research community needs to gather datasets of human arguing and develop specific interaction guidelines (e.g., following [Amershi et al., 2019]).

2.3.1 Discussion

The questions raised at the workshop focussed on the envisioned interaction. Are there linguistic markers the system could employ to detect the user's agreement and understanding of specific points of view? How much personalization is necessary? How should the system present the arguments for a fair juxtaposition (see statement on future fair conversations)? The participants suggested studying Reddit's changemyview forum for related conversations. However, one should bear in mind that answers of people and machines are perceived differently (i.e., for what is considered "polite"). Furthermore, the discussion highlighted relevant work on synthetic voices concerning persuasiveness [Dubiel et al., 2020], perceived competence [Doyle et al., 2021], and eliciting trust [Torre et al., 2018].

2.4 Companionship Features in Intelligent Personal Assistants

Irene Lopatovska, Pratt Institute

Problem. Users of Intelligent Personal Assistants (IPAs, e.g., Google Assistant, Apple Siri, or Amazon Alexa) engage with IPAs for various purposes, including listening to music, controlling IoT devices, managing and obtaining information [Canbek and Mutlu, 2016]. However, a large number of interactions seem to be driven by users' curiosity about IPAs personality [Lopatovska and Oropeza, 2018; Lopatovska et al., 2019]. Utterances like "Where do you live?", "Do I look fat?" and others signal users' expectations of IPA ability to handle personal conversations.

Current IPAs are not designed to support long-term interactions, and often lack appropriate content to handle utterances on the personal topics [Lopatovska, 2020]. Yet the need for personal conversations is present, especially in the time of social distancing and isolation leading to a social deprivation of face-to-face relationships [Loades et al., 2020]. This may be particularly true for adolescents and elderly populations who are known to have a higher likelihood of feeling lonely [Ellis et al., 2020; Von Soest et al., 2020].

We propose to start experimenting with the IPA's "companionship" features: pro-actively checking on users' moods, listening, conversing, offering supportive and affirmative messages.

Goal. In contexts where IPAs are highly anthropomorphised and users might be in need of a companion or a conversationalist, I propose developing IPA apps that would aim to listen, provide helpful feedback, encouragement and otherwise support the emotional well-being of its users.

State of the Art. There are currently numerous apps aimed at addressing users' well-being, ranging from assisting in breathing and meditation, to mindfulness activities (not to mention fitness applications). However, most apps are designed to guide user behaviour and treat users as passive instruction recipients, instead of engaging in an active dialog with the user.

Steps. The tech community can help by collaborating with psychologists/behavioural scientists to design IPA programs that are modelled on human-to-human interactions. Conversational solutions to loneliness, anxiety and other behavioural problems should be investigated for their potential to inform the design of "companion" programs in IPAs. Designing programs that "do no harm" in dealing with emotional/behavioural issues will not be easy, and we can start by making small steps on the path to full IPA "companions" (e.g., before designing long conversations, we can test effects of shorter responses that rely on humour, conversational "distractors" from sad thoughts, or uplifting content.

2.4.1 Discussion

The discussion of IPA-companion focused on several potential directions for developing companion persona as well as specific means of manifesting "companionship". It was suggested that IPA persona manifests itself through a) content that IPA presents to the users (e.g. humour, facts, information that might contradict the user), and b) aspects of the presentation style (e.g. voice). While human-like presentation style (e.g. feminine/masculine sounding voices) might be expected and appreciated by users, designers should consider challenging these expectations that often reinforce undesirable stereotypes (e.g., Microsoft is aiming to make their Cortana IPA appear clearly non-human). IPA styles that communicate "machineness", "otherware" or other clearly non-human personas should be explored, as well as contexts where persona is not desirable (search engines might not need persona manifestations). Participants suggested additional directions for IPA-companion research, including investigation of companion robots in home or care-giving settings, strategies for "nudging" and proactively engaging users, opportunities and threats of developing virtual therapists.

2.5 Mixed-Initiative for Conversational Search

Paul Owoicho, University of Glasgow

Problem. Despite the rapid adoption of digital assistants [Fernandes and Oliveira, 2021], their use for multi-turn conversational search remains limited [Aliannejadi et al., 2020]. While users can carry out short, few-turn "conversations" with their digital assistants, these systems often fail to satisfy the information needs of users over longer-turn dialogues [Zhang et al., 2018]. Conversely, today's consumers increasingly expect convenience and seamless user experiences when using digital platforms [Parise et al., 2016]. This divide between user expectations and system capabilities presents an opportunity for researchers to lay the groundwork for the next generation of conversational search systems that are capable of meeting a user's complex, and evolving information needs. More specifically, it highlights the need for exploring mixed-initiative interaction –where the human and the system work together as collaborators– to help us strive towards even more efficient, effective, and engaging conversational search systems.

To this end, we propose working on creating dialogue search systems that can do more than just returning search results. Given a user's query, these systems should be able to choose from a set of actions that allow them to: ask a clarifying question to eliminate ambiguities, suggest other relevant topics for the user to explore, or even incorporate explicit user feedback to modify the results it presents to the user.

Goal. If we imagine initiative to be a spectrum of human-computer interaction styles (i.e human-led, system-led), mixed-initiative interaction can be described as a flexible combination of the approaches at the opposite ends of the spectrum [Radlinski and Craswell, 2017]. Before we can build mixed-initiative conversational search systems, however, the research community needs to develop a better understanding of what initiative is and how it works. To accomplish this, we need to collect conversational datasets, and create metrics to (1) study, quantify and evaluate the use of initiative, and (2) determine what the right 'mix' of human-initiative and system initiative is for effective conversational search.

State of the Art. The interaction style in current dialogue systems is either primarily driven by human initiative or by system initiative. For example, when a user interacts with Google Assistant about a topic, the user drives the conversation by issuing queries while the system simply retrieves relevant results. There are also several datasets to facilitate the study of initiative in information-seeking conversations and chatbot systems [Dalton et al., 2020; Logacheva et al., 2018; Thomas et al., 2017; Trippas et al., 2017], however, these are not well-suited for mixed-initiative research due to poor ASR transcription quality, wildly varying initiative policies, and a subjective notion of conversation success. Recently, the ConversationShape metric was proposed for measuring the asymmetry of information exchange between participants in a dialogue [Vakulenko et al., 2020]. However, we argue that the metric does not inform on the effectiveness of different initiative policies.

Steps. Future data collection efforts should embrace robust speech transcription protocols and have clear interaction policies for participants to follow. We also recommend the development of metrics that allow us to determine how and when different initiative policies are effective as well as their overall impact on the success of a conversation. These will form the basis for modelling mixed-initiative search systems.

2.5.1 Discussion

The discussion on mixed-initiatives in conversational search explored the need and qualities of a suitable metric which could measure initiative for both the user and system. Several user studies have been conducted to look into the patterns of searcher-intermediary interactions, but the studies differ in the way the conversational agent was implemented. For example, the agent could be a human or a prototypical system. The human intermediary could also be implemented as a Wizard. Therefore, the interactions range from being human-human (including rich dialogues, context, and casual) to human-system (more simplistic). Developing a metric would allow researchers to assess the initiative and its effect on the task success and user satisfaction.

There are some existing models [Zhang et al., 2018] which provide some idea about the initiative in the conversation. However, we need to develop metrics which measure the amount and nature of the initiative in addition to the back-and-forth question-answering. In a laboratory-based experimental setting, the users are more likely to claim that the search session was a success. Success and satisfaction is also hard to measure over short interactions.

When the system mimics humans, and interacts with the user, it also raises the question of ethics and moral obligations. For the system to take initiative, it must prioritise if the initiative is necessary. As seen with prior research, users are reluctant to relinquish control of the search and conversation unless the partner is a human. Therefore, should the system do anything which does not align with the user's search direction. For example, the user may want to look into a particular document while the agent may feel that better information exists outside of that document. Should the agent search for newer documents or respect the wishes of the user? Human-human conversations are also more casual. For a system to mimic a human partner, it has to find the right balance between casual and functional, and incorporate initiative in a non-intruding way.

2.6 Using Speech in Conversational Search: Considering People with Diverse Speech Patterns

Leigh Clark, Swansea University

Problem. Using speech as a modality of finding information is a common way of using conversational agents like Google Assistant, Amazon Alexa, or Siri. However, users with diverse speech patterns (e.g. stammering/stuttering) may be inhibited from interacting fully with these systems, or may be excluded entirely. Stammering (or stuttering) is often characterised by disruptions to speech production including repetition, prolongation or hesitation of particular sounds or words. Approximately 8% of children will stammer at some point and for up to 3% of adults it will be a lifelong condition.[7]

Goal. It is crucial for researchers and practitioners to understand user interactions for people who stammer (and those with other diverse speech patterns). By understanding the current positive and negative aspects of interactions, we can develop inclusive speech-based interfaces. This includes technical aspects and proposing design heuristics for these systems [Clark et al., 2020]. We can also explore whether specific interaction contexts like conversational search are impacted differently to others for specific user demographics.

State of the Art. There are ongoing technical challenges in automatic speech recognition (ASR) such as endpoint detection (identifying when a speaker has finished speaking) [Li et al., 2002]. In terms of user interactions, we can learn from existing work on making speech interfaces accessible [Brewer et al., 2018] and directing further attention to people with diverse speech patterns like stammering.

Steps. Appropriate user engagement is critical to first understanding interactions before further developments to speech interface design can be proposed and implemented. This includes engaging with other research disciplines from human-computer interaction, information retrieval, speech technology, human factors and speech therapy. Additionally, engaging with people outside of academia like charities would provide a user community-focused platform for impactful research.

2.6.1 Discussion

Our discussion first considered the boundaries with stammering and other diverse speech patterns. We considered there may be overlaps in these different types of speech and, consequently, overlaps

[7] https://stamma.org/news-features/stammering-population

in potential avenues to improve speech interface interactions. In making these interactions more inclusive for people who stammer, we discussed how much of this is related to ASR and how much can be addressed by other elements of interface design. Additionally, the scenario in which people engage with speech interfaces may incur different levels of interaction success. Retrieving information over multiple turns, for example, may be more challenging with current systems in contrast to single turn question-answer scenarios.

With these challenges and open-questions, it was clear that an engagement with people who stammer is crucial to understand these challenges in more detail. Indeed, engaging with and for communities was highlighted as an important factor in designing more inclusive futures. Using participatory and co-design approaches was discussed as a method for taking this further.

2.7 Future *Fair* Conversations

Damiano Spina, RMIT University

Problem. Some questions posed to a conversational system may not have one single correct answer, but multiple, diverse answers. For instance, a spoken conversational search system [Trippas et al., 2020] receiving the question "Should Australia have a universal basic income program?" should provide enough explanations about the pros and cons of having a universal basic income, rather than limiting to give a Yes/No answer. In scenarios where satisfying users' information needs involves communicating multiple pieces of information –and presumably, in more than one turn– the creation or reinforcement of systematic cognitive bias is likely to happen. We know that the way information is exposed to users has a significant impact in how information is perceived [Castillo, 2019; Allan et al., 2018]. One could argue that this is even more acute in speech-only interfaces (e.g., smart speakers or intelligent assistants), due the fact that the cognitive load of listening to information limits the capability of exposing diverse information.

Goal. The ultimate goal is to design audio-only conversational communication strategies that would minimise the creation or reinforcement of undesirable cognitive biases. For instance, strategies that include proactive utterances from the system that would encourage users to explore more answers. Would exposing users to more answers prevent the reinforcement of undesirable cognitive biases? If not, what would be the communication strategy to use? What are the strategies that make an audio-only interaction fair? What does a *fair* conversational interaction sound like?

State of the Art. The problem summarised above is highly related to two of the grand challenges identified at the Third Strategic Workshop in Information Retrieval in Lorne (SWIRL III) in 2018 [Allan et al., 2018]: (i) Fairness, Accountability, Confidentiality and Transparency (FACT) in IR [Olteanu et al., 2021], and (ii) Conversational Search [Anand et al., 2020]. However, little work has been done on the problem of fairness exposure of answers for information-based conversational search assistants. Gerritse et al. identified different conversational strategies to deal with biases in Personalised Knowledge Graphs (PKGs) [Gerritse et al., 2020]. Thomas et al. found that it is possible to measure differences in style when participants interact with a speech-only conversational agent [Thomas et al., 2020]. Would different fairness-aware communication strategies lead to changes in style? And, if so, what is the relationship between style and changes in cognitive biases?

Steps. As in other conversational information access challenges, the creation of resources to help explore the effect of conversation strategies in cognitive biases is one of the main steps we should take. This includes: user studies (e..g, using a Wizard-of-Oz methodology), the creation of test collections (e.g., complex non-factoid question answering involving arguments and questions with multiple correct answers), and also new instrumentation to quantify and measure cognitive biases during speech-only conversations between humans and agents (e.g., incorporating signals from wearable devices). Collaborations with other disciplines (e.g., cognitive science) is also crucial if we want to better understand what would a "fair audio-only information seeking interaction" sound like.

2.7.1 Discussion

One of the main topics discussed during the workshop was the notion of fair conversations in relation to trust, provenance, and source of information. Naming the source of information is necessary, but probably not sufficient. Some participants linked the problem of presenting reliable information to the work performed by librarians, who often point to reliable sources: even when the question reveals a bias, a librarian would point to authoritative sources and share a dominant view. Intelligent assistants may go beyond that, as they could provide quantifiable statistics in terms of number of sources with accurate or inaccurate information.

The discussion was also centred in determining what *fair* is: which sources are fair? Who decides that, i.e., is the intelligent assistant accountable for the editorial power? The notion of parity of *exposure* also was discussed, in particular, when opinions or arguments are not necessarily equally reliable (e.g., "the Earth is flat"). In relation to conversational style, persuasiveness of synthetic voices may play an important role: previous work has shown that more persuasive synthetic voices have an impact in terms of perceived truthfulness in users interacting with goal-oriented intelligent assistants [Dubiel et al., 2020].

Perhaps one way for avoiding bias is to acknowledge it, e.g., making the user aware about potential bias in the information presented by the intelligent assistant.

2.8 Social Competence in Conversational Retrieval

Paul Thomas, Mary Czerwinski, Daniel McDuff, and Nick Craswell, Microsoft

Problem. Conversational IR systems are rapidly getting better at understanding searcher utterances, tracking context, and finding and presenting relevant information. However, conversation is much more than an exchange of facts: humans start learning conversational norms in utero, and unlike web user interfaces, conversational exchanges are laden with social signals. As humans, we cannot help but treat computers as social agents [Nass et al., 1994], so as developers and researchers we must be aware of social conventions.

Examples include:

- *Politeness norms*, which constrain what we say to whom and when. We have not yet understood what it is for an IR agent to be "polite", nor formalised any design rules to help.

- *Non-verbal behaviours*, which can carry as much information as the words used [Scherer and Ceschi, 2000] and which are important for feeling an agent is "helpful" [McDuff et al., 2017], are not well studied in IR.

- *Style and alignment* are important for feelings of empathy and efficiency, in talking with people [Tannen, 1987; Brennan, 1996] or machines [Branigan et al., 2010; Dubiel et al., 2020; Pickering and Garrod, 2004].

Conversational agents are rapidly getting more useful, and it is time to think about going beyond "mere" correctness to the full gamut of conversational phenomena.

Goal. An overall goal is simple to state, if hard to get to: a conversational IR agent should not only be competent at retrieving information, formulating coherent utterances, and managing an exchange of information; it should also be competent in at least the basic social nuances of conversation [Reeves, 2010].

State of the Art. We already know a great deal about human-to-human conversation, including anthropomorphism, "personality", social norms, and social signalling beyond simple question-answering. There are some good examples outside of IR, for example, in conversational recommendation systems [Walker et al., 1997] or in-car systems [Stier et al., 2020].

We can also develop guidelines for conversational IR systems, based on social norms and what we learn from the work above. For example, Gnewuch et al. [2017] have design rules based on Grice's well-known maxims for conversation [Grice, 1989].

Steps. We suggest some concrete steps:

1. IR researchers should be aware of social phenomena that may apply to their conversational systems; what it would mean if these phenomena were seen in this new context; and how their designs might change as a result.

2. Agents should be designed with effective expression of emotion, as well as of fact. This might be through facial expressions and body gestures, through style matching speech, or through careful selection of text.

3. We should consider how to build an agent which can track, and adapt, to conversational style.

4. We should consider when "embodied" agents might be useful, to provide a richer set of social cues.

5. IR researchers should collect and use corpora which include data on aspects such as tone and rate of speech, lexical variation, facial and body gestures, and other channels, beyond just text.

6. Guidelines such as Gnewuch et al.'s should be tested in an IR setting, and as a community we need to develop something similar for our own systems.

We should also consider the possible drawbacks of applying this research here. Are there opportunity costs or particular risks with regards to ethics, privacy and security?

2.8.1 Discussion

A major question about this agenda is where we should start—there is a lot known about conversation, more than we can possibly incorporate at once. Four ideas were floated as good places to start:

1. Embodiment (e.g., a "physical" agent which presents e.g., a "head" or "body", and can gesture accordingly);
2. Tracking conversational style: pitch, speed, and lexical variation for instance;
3. Knowing which style might be important when presenting each piece of information –which will, of course, depend on task and personal context;
4. Politeness rules, so agents are less likely to appear rude, and how these work in different scenarios and contexts.

Workshop participants discussed embodiment in some detail. If an agent has a human appearance, we may need to provide more human-like features in other aspects. What do we need in terms of linguistic or prosodic features in this case? Do we need more than we would for a simpler agent which does not present as human-like? Is the extra sympathy worth the effort? Do theories of politeness and face, or other theories, need to be re-imagined for this kind of conversation?

There was also some comment about the persona an IR system could provide, once it is a sufficiently competent interlocutor. Would (could) a system's persona confirm our mental model of search? When is a persona helpful, and when is the task better served by a simpler interaction? Are there tasks where persuasion would be more useful, and would a more human-like persona help here?

3 Conclusions and Lessons Learned

The Future Conversations workshop at CHIIR 2021 explored new ideas, controversial statements, or gaps in the conversational search research. We elicited and discussed short opinion pieces before and during the workshop. As can be seen, this workshop format resulted in the rich description of a promising research agenda on this topic. We now present the high-level overview of these discussions.

The CHIIR community is highly motivated to understand, design, and research conversational search. Overall re-occurring topics included

- Which are the tasks these systems can fulfil?
- How do we include "context" in future systems?

- What is the impact of personalisation on conversational interactions?
- How can we detect which presentation mode is suitable for particular tasks, contexts, abilities?
- Can these systems be pro-active, take initiative, and support mixed-initiative?
- How do we overcome trust issues when people are searching over conversations?
- What is the impact of different “output” and interaction styles?
- How can systems be transparent about the curating of results and their editorial input on existing pieces of information?
- How can these systems help to overcome the limitations of Automatic Speech Recognition (ASR)?

To accommodate the online format of ACM CHIIR 2021, the Future Conversations workshop was held at two different time-slots. Running the workshop twice was demanding for authors who attended twice. Nevertheless, discussing the topic twice with different groups of people was valuable. Different discussions resulted in each session, even though the same topics were discussed on the basis of statements submitted before the workshop.

We provided clear instructions and guidance on how the create the statements (Section 2). We believe that this was beneficial for the authors. In total, we received eight statements, and all of them were discussed during the workshop sessions. Furthermore, the instructions were written in such a way that it was easy to understand which steps could be taken for expanding on the conversational information access, adding to the value of the workshop. Sharing the statements prior to the workshop enabled other researchers to comment on the ideas. Although the comments were not explicitly used during the sessions, they helped starting the discussions –as most of the commentators were also participants of the workshop. We had also planned to compile a technical report during the workshop, however, the discussions during the two-hour workshop were given priority. Instead, we contacted the attendees after the workshop to help shape the statements and add notes on the discussions from the workshop.

Lastly, we diverted to Google Drive to share the statements and comments. Originally we investigated using OpenReview.[8] However, OpenReview did not intuitively support functions such as leaving comments on statements.

This technical report is an evidence of the benefits of running a virtual workshop in this format can have. We believe this format can be successfully adopted for future venues.

References

Mohammad Aliannejadi, Manajit Chakraborty, Esteban Andrés Ríssola, and Fabio Crestani. Harnessing evolution of multi-turn conversations for effective answer retrieval. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR ’20, pages 33–42, 2020.

[8] http://www.openreview.net/

James Allan, Jaime Arguello, Leif Azzopardi, Peter Bailey, Tim Baldwin, Krisztian Balog, Hannah Bast, Nick Belkin, Klaus Berberich, Bodo von Billerbeck, Jamie Callan, Rob Capra, Mark Carman, Ben Carterette, Charles L. A. Clarke, Kevyn Collins-Thompson, Nick Craswell, W. Bruce Croft, J. Shane Culpepper, Jeff Dalton, Gianluca Demartini, Fernado Diaz, Laura Dietz, Susan Dumais, Carsten Eickhoff, Nicola Ferro, Norbert Fuhr, Shlomo Geva, Claudia Hauff, David Hawking, Hideo Joho, Gareth Jones, Jaap Kamps, Noriko Kando, Diane Kelly, Jaewon Kim, Julia Kiseleva, Yiqun Liu, Xiaolu Lu, Stefano Mizzaro, Alistair Moffat, Jian-Yun Nie, Alexandra Olteanu, Iadh Ounis, Filip Radlinski, Maarten de Rijke, Mark Sanderson, Falk Scholer, Laurianne Sitbon, Mark Smucker, Ian Soboroff, Damiano Spina, Torsten Suel, James Thom, Paul Thomas, Andrew Trotman, Ellen Voorhees, Arjen P. de Vries, Emine Yilmaz, and Guido Zuccon. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in Lorne (SWIRL 2018). *SIGIR Forum*, 52(1):34–90, August 2018.

Milad Alshomary, Nick Düsterhus, and Henning Wachsmuth. Extractive snippet generation for arguments. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, pages 1969–1972. ACM, 2020.

Saleema Amershi, Daniel S. Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi T. Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. Guidelines for human-AI interaction. In Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos, editors, *Conference on Human Factors in Computing Systems*, CHI '19. ACM, 2019.

Avishek Anand, Lawrence Cavedon, Hideo Joho, Mark Sanderson, and Benno Stein. Conversational Search (Dagstuhl Seminar 19461). *Dagstuhl Reports*, 9(11):34–83, 2020.

Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. From arguments to key points: Towards automatic argument summarization. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *58th Annual Meeting of the Association for Computational Linguistics*, ACL '20, pages 4029–4039. ACL, 2020.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. Linguistic alignment between people and computers. *J. Pragmatics*, 42:2355–2368, 2010.

Susan E. Brennan. Lexical entrainment in spontaneous dialog. In *Proceedings of the International Symposium on Spoken Dialogue*, 1996.

Robin N Brewer, Leah Findlater, Joseph'Jofish' Kaye, Walter Lasecki, Cosmin Munteanu, and Astrid Weber. Accessible voice interfaces. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18, pages 441–446, 2018.

Nil G. Canbek and Mehmet E. Mutlu. On the track of artificial intelligence: Learning with intelligent personal assistants. *Journal of Human Sciences*, 2016.

Carlos Castillo. Fairness and transparency in ranking. *SIGIR Forum*, 52(2):64–71, January 2019.

Justin Chan, Thomas Rea, Shyamnath Gollakota, and Jacob E Sunshine. Contactless cardiac arrest detection using smart devices. *NPJ digital medicine*, 2(1):1–8, 2019.

Leigh Clark, Benjamin R Cowan, Abi Roper, Stephen Lindsay, and Owen Sheers. Speech diversity and speech interfaces: considering an inclusive future through stammering. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20, pages 1–3, 2020.

Sebastian Cross, Ahmed Mourad, Guido Zuccon, and Bevan Koopman. Search engines vs. symptom checkers: A comparison of their effectiveness for online health advice. In *Proceedings of the Web Conference 2021*, TheWebConf '21, pages 206–216. ACM, 2021.

Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. TREC CAsT 2019: The conversational assistance track overview. *arXiv preprint arXiv:2003.13624*, 2020.

Yashar Deldjoo, Johanne R. Trippas, and Hamed Zamani. Towards multi-modal conversational information seeking. In *Proceedings of the ACM Conference on Research and Development in Information Retrieval*, SIGIR '21. ACM, 2021.

Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. *CoRR*, abs/2102.02094, 2021.

Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. Persuasive synthetic speech: Voice perception and user behaviour. In María Inés Torres, Stephan Schlögl, Leigh Clark, and Martin Porcheron, editors, *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20, pages 6:1–6:9. ACM, 2020.

Roxanne El Baff, Khalid Al Khatib, Benno Stein, and Henning Wachsmuth. Persuasiveness of news editorials depending on ideology and personality. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 29–40. ACL, 2020.

Wendy E Ellis, Tara M Dumas, and Lindsey M Forbes. Physically isolated but socially connected: Psychological adjustment and stress among adolescents during the initial COVID-19 crisis. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 52 (3):177, 2020.

David Elsweiler, Christoph Trattner, and Morgan Harvey. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 575–584, 2017.

Teresa Fernandes and Elisabete Oliveira. Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research*, 122:180–191, 2021.

Alexander Frummet, David Elsweiler, and B. Ludwig. Detecting domain-specific information needs in conversational search dialogues. In *NL4AI@AI*IA*, 2019.

Emma J. Gerritse, Faegheh Hasibi, and Arjen P. de Vries. Bias in conversational search: The double-edged sword of the personalized knowledge graph. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, ICTIR '20, pages 133–136. ACM, 2020.

Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. Towards designing cooperative and social conversational agents for customer service. In *Proceedings of the 38th International Conference on Information Systems*, ICIS '17, 2017.

Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1989.

Christina Hartmann, Simone Dohle, and Michael Siegrist. Importance of cooking skills for balanced food choices. *Appetite*, 65:125–131, 2013.

Johannes Kiesel, Kevin Lang, Henning Wachsmuth, Eva Hornecker, and Benno Stein. Investigating expectations for voice-based and conversational argument search on the Web. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, pages 53–62. ACM, 2020.

Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal*, 22(1-2):188–227, 2019.

Qi Li, Jinsong Zheng, Augustine Tsai, and Qiru Zhou. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. *IEEE Transactions on Speech and Audio Processing*, 10(3):146–157, 2002.

Maria Elizabeth Loades, Eleanor Chatburn, Nina Higson-Sweeney, Shirley Reynolds, Roz Shafran, Amberly Brigden, Catherine Linney, Megan Niamh McManus, Catherine Borwick, and Esther Crawley. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(11):1218–1239, 2020.

Varvara Logacheva, Mikhail Burtsev, Valentin Malykh, Vadim Polulyakh, and Aleksandr Seliverstov. Convai dataset of topic-oriented human-to-chatbot dialogues. In *The NIPS'17 Competition: Building Intelligent Systems*, pages 47–57. Springer, 2018.

Irene Lopatovska. Classification of humorous interactions with intelligent personal assistants. *Journal of Librarianship and Information Science*, 52(3):931–942, 2020.

Irene Lopatovska and Heyrling Oropeza. User interactions with "Alexa" in public academic space. *Proceedings of the Association for Information Science and Technology*, 55(1):309–318, 2018.

Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4): 984–997, 2019.

Filipe M. Martins, J. Pardal, Luís Franqueira, Pedro Arez, and N. Mamede. Starting to cook a tutoring dialogue system. *2008 IEEE Spoken Language Technology Workshop*, pages 145–148, 2008.

Daniel McDuff, Paul Thomas, Mary Czerwinski, and Nick Craswell. Multimodal analysis of vocal collaborative search: A public corpus and results. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI '17, pages 456–463. ACM, 2017.

Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 72–78. ACM, 1994.

Ikujiro Nonaka and Hirotaka Takeuchi. *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford University Press, 1995.

Elnaz Nouri, Robert Sim, Adam Fourney, and Ryen W. White. Step-wise recommendation for complex task support. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, pages 203–212. ACM, 2020.

Alexandra Olteanu, Jean Garcia-Gathright, Maarten de Rijke, Michael D. Ekstrand, Adam Roegiest, Aldo Lipani, Alex Beutel, Alexandra Olteanu, Ana Lucic, Ana-Andreea Stoica, Anubrata Das, Asia Biega, Bart Voorn, Claudia Hauff, Damiano Spina, David Lewis, Douglas W. Oard, Emine Yilmaz, Faegheh Hasibi, Gabriella Kazai, Graham McDonald, Hinda Haned, Iadh Ounis, Ilse van der Linden, Jean Garcia-Gathright, Joris Baan, Kamuela N. Lau, Krisztian Balog, Maarten de Rijke, Mahmoud Sayed, Maria Panteli, Mark Sanderson, Matthew Lease, Michael D. Ekstrand, Preethi Lahoti, and Toshihiro Kamishima. FACTS-IR: Fairness, accountability, confidentiality, transparency, and safety in information retrieval. *SIGIR Forum*, 53(2):20–43, March 2021.

Salvatore Parise, Patricia J Guinan, and Ron Kafka. Solving the crisis of immediacy: How digital technology can transform the customer experience. *Business Horizons*, 59(4):411–420, 2016.

Martin J. Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brian Sciences*, 27(2):169–225, 2004.

Martin Potthast, Lukas Gienapp, Florian Euchner, Nick Heilenkötter, Nico Weidmann, Henning Wachsmuth, Benno Stein, and Matthias Hagen. Argument Search: Assessing Argument Relevance. In *42nd International ACM Conference on Research and Development in Information Retrieval*, SIGIR '19. ACM, 2019.

Martin Potthast, Matthias Hagen, and Benno Stein. The Dilemma of the Direct Answer. *SIGIR Forum*, 54(1), June 2020.

Chen Qu, Liu Yang, W. Bruce Croft, Yongfeng Zhang, Johanne R. Trippas, and Minghui Qiu. User intent prediction in information-seeking conversations. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, CHIIR '19, pages 25–33, 2019.

Thomas F Quatieri, Tanya Talkar, and Jeffrey S Palmer. A framework for biomarkers of COVID-19 based on coordination of speech-production subsystems. *IEEE Open Journal of Engineering in Medicine and Biology*, 1:203–206, 2020.

Filip Radlinski and Nick Craswell. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*, CHIIR '17, pages 117–126, 2017.

Byron Reeves. People do like people: The benefits of interactive online characters. *Madison Avenue J*, April 2010.

Klaus R Scherer and Grazia Ceschi. Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and social psychology bulletin*, 26 (3):327–339, 2000.

Sosuke Shiga, Hideo Joho, Roi Blanco, Johanne R. Trippas, and Mark Sanderson. Modelling information needs in collaborative search conversations. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 715–724. ACM, 2017.

Noam Slonim, Yonatan Bilu, Carlos Alzate, Roy Bar-Haim, Ben Bogin, Francesca Bonin, Leshem Choshen, Edo Cohen-Karlik, Lena Dankin, Lilach Edelstein, Liat Ein-Dor, Roni Friedman-Melamed, Assaf Gavron, Ariel Gera, Martin Gleize, Shai Gretz, Dan Gutfreund, Alon Halfon, Daniel Hershcovich, Ron Hoory, Yufang Hou, Shay Hummel, Michal Jacovi, Charles Jochim, Yoav Kantor, Yoav Katz, David Konopnicki, Zvi Kons, Lili Kotlerman, Dalia Krieger, Dan Lahav, Tamar Lavee, Ran Levy, Naftali Liberman, Yosi Mass, Amir Menczel, Shachar Mirkin, Guy Moshkowich, Shila Ofek-Koifman, Matan Orbach, Ella Rabinovich, Ruty Rinott, Slava Shechtman, Dafna Sheinwald, Eyal Shnarch, Ilya Shnayderman, Aya Soffer, Artem Spector, Benjamin Sznajder, Assaf Toledo, Orith Toledo-Ronen, Elad Venezian, and Ranit Aharonov. An autonomous debating system. *Nature*, 591(7850):379–384, Mar 2021.

Daniela Stier, Katherine Munro, Ulrich Heid, and Wolfgang Minker. Towards situation-adaptive in-vehicle voice output. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20. ACM, 2020.

Deborah Tannen. Conversational style. In Hans W. Dechert and Manfred Raupach, editors, *Psycholinguistic models of production*. Ablex, 1987.

Paul Thomas, Daniel McDuff, Mary Czerwinski, and Nick Craswell. MISC: A data set of information-seeking conversations. In *Proceedings of the 1st International Workshop on Conversational Approaches to Information Retrieval*, CAIR '17, 2017.

Paul Thomas, Daniel McDuff, Mary Czerwinski, and Nick Craswell. Expressions of style in information seeking conversation with an agent. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, pages 1171–1180. ACM, 2020.

Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanatto. Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In Amber L. Story, editor, *Technology, Mind, and Society Conference*, pages 40:1–40:6. ACM, 2018.

Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. How do people interact in conversational speech-only search tasks: A preliminary analysis. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, pages 325–328, 2017.

Johanne R. Trippas, Damiano Spina, Paul Thomas, Mark Sanderson, Hideo Joho, and Lawrence Cavedon. Towards a model for spoken conversational search. *Information Processing & Management*, 57(2):102162, 2020.

Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S Kashavan, and John Blake Torous. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7):456–464, 2019.

Svitlana Vakulenko, Evangelos Kanoulas, and Maarten de Rijke. An analysis of mixed initiative and collaboration in information-seeking dialogues. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, pages 2085–2088. ACM, 2020.

Tilmann Von Soest, Maike Luhmann, and Denis Gerstorf. The development of loneliness through adolescence and young adulthood: Its nature, correlates, and midlife outcomes. *Developmental psychology*, 2020.

Henning Wachsmuth, Martin Potthast, Khalid Al-Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. Building an argument search engine for the Web. In Kevin Ashley, Claire Cardie, Nancy Green, Iryna Gurevych, Ivan Habernal, Diane Litman, Georgios Petasis, Chris Reed, Noam Slonim, and Vern Walker, editors, *4th Workshop on Argument Mining (ArgMining) at EMNLP*, pages 49–59. ACL, 2017.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. PARADISE: A framework for evaluating spoken dialogue agents. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, ACL/EACL '97, pages 271–280. ACL, 1997.

Amy L. Wilson, Elizabeth Buckley, Jonathan D. Buckley, and Svetlana Bogomolova. Nudging healthier food and beverage choices through salience and priming. evidence from a systematic review. *Food Quality and Preference*, 51:47–64, 2016.

Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. Towards conversational search and recommendation: System ask, user respond. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 177–186, 2018.