More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings

Moira McGregor[1, 2]
[1]Microsoft Research
1065 La Avenida
Mountain View, CA 94043 US
johntang@microsoft.com

John C. Tang[1]
[2]Mobile Life Research Centre
Stockholm University
SE-164, Kista, Sweden
moira@mobilelifecentre.org

ABSTRACT

Personal assistants using a command-dialogue model of speech recognition, such as Siri and Cortana, have become increasingly powerful and popular for individual use. In this paper we explore whether similar techniques could be used to create a speech-based agent system which, in a group meeting setting, would similarly monitor spoken dialogue, pro-actively detect useful actions, and carry out those actions without specific commands being spoken. Using a low-fi technical probe, we investigated how such a system might perform in the collaborative work setting and how users might respond to it. We recorded and transcribed a varied set of nine meetings from which we generated simulated lists of automated 'action items', which we then asked the meeting participants to review retrospectively. The low rankings given on these discovered items are suggestive of the difficulty in applying personal assistant technology to the group setting, and we document the issues emerging from the study. Through observations, we explored the nature of meetings and the challenges they present for speech agents.

Author Keywords

Automatic Speech Recognition; Meeting Agents; Speech Interaction; Collaborative Workplace Technology

ACM Classification Keywords

H.5.3 Group and Organization Interfaces, Computer-supported cooperative work

INTRODUCTION

Much has been made of the opportunity for speech-based agent systems to assist and aid human activity. In situations such as driving, where manual control is not available, speech provides an alternative way to interact with technology. Similar modal interaction has been deployed in other domains, including museums and information kiosks, call centres, smart-home and assistive care systems where automatic speech recognition (ASR) can be used to detect commands and complete simple actions safely. Alongside development of commercial systems such as Apple Siri and Microsoft Cortana, interest in mobile speech recognition technologies has flourished. These systems have achieved high recognition rates by detecting predefined commands (such as 'send text message'), with users explicitly voicing those commands to a given device. Using similar techniques it is also possible that a system could pro-actively detect and carry out useful actions without an explicit command being spoken. So, for example, a system might detect one person to another saying, "We should meet next week…", and act on that phrase to propose an appropriate calendar entry. One possible scenario for such a system could be its use in collaborative work meetings, a longstanding domain for automated speech recognition ASR applications [6].

Introducing new technology to the workplace environment carries a high risk of rejection, and previous software agents, such as the Microsoft "Clippy", quickly became reviled by many users as annoying and intrusive [26]. In this paper we discuss a corpus of audio-video material of meetings collected to gain access and insight on the setting in which any speech-based 'meeting agent' might be situated, as well as to provide 'training' speech data for future ASR algorithms.

To understand how well a speech agent would be able to 'listen' to talk in meetings and provide meaningful support to the participants during the meeting and after, we created a low-fi technology probe [4] that made use of manual transcripts of the meetings and the Microsoft Cortana data grammar, to emulate what sort of recognition a system could potentially achieve in the context of collaborative work meetings. Could a speech agent detect suitable actions to carry out during a meeting? Some actions would require user input during the meeting itself, whereas others could be completed automatically without user intervention. It is possible that a speech-based meeting agent would thus be able to help with the running of the meeting itself, and also with the utility of the meeting after the event, by producing to-do items for meeting participants, for example.

CSCW '17, February 25-March 01, 2017, Portland, OR, USA

ACM 978-1-4503-4335-0/17/03…$15.00
DOI: http://dx.doi.org/10.1145/2998181.2998335

For the probe, we recorded a varied set of nine workplace meetings, conducted amongst developers and managers in a technology development organisation. Our goal was to emulate how a system like this might potentially be of use. To do this we simulated the system actions by first, having a human transcribe the meeting audio, and second, manually selecting 'actionable items', following the Cortana conversational agent schema to categorise relevant utterances. These two manual steps created what was a 'best case' scenario for a meeting agent: making use of a high quality transcript, along with a human-based understanding of the transcript to recognize and produce action items from the meetings. We then reviewed the lists of 'actionable items' with the meeting participants involved and asked them to rate them in terms of how useful they would have been either during the meeting, or after it.

Our action items were rated as poor by the meeting participants - with a mean score of 2.23 per action item (with 5 being extremely useful, and 1 being completely useless). For our participants, in explaining these low scores, the actions failed to fit with the meeting or gave an incorrect summary of what was actually being discussed, or what the participants intended. We discuss four reasons for this: 1) lack of contextual information recorded in the action, 2) miscategorisation of dialogue, 3) dealing badly with errors in the transcription, and 4) the low overall number of action items detected in the meetings. Addressing these problems provides a number of challenges for ASR research. In particular, the issue of transcript quality is likely to prove a substantial barrier to meeting agent technology in the future [28:p123].

However, the fieldwork and the meeting transcripts themselves provide a complementary resource for addressing and understanding these issues. Drawing on previous CSCW work on meetings, their functioning, and the design of technology for the support of collaborative work, in the second part of the results we explore the nature of meetings themselves. There seem to be dimensions of meetings which are missed by the transcribed records of talk, each of which has design implications for speech-based agents in collaborative workplace meetings. In particular, we discuss three elements of meetings that could present problems for agent technology and designers of future systems: the individual information needs of participants; the importance of social interaction in meetings; and the different perspectives on workplace meetings outcomes.

BACKGROUND

Technology for Meetings

In a meeting, people exchange information, raise issues, express opinions, make suggestions, propose solutions, argue, negotiate and make decisions. Despite their ubiquity and persistence, meetings are understood to be inefficient both anecdotally as well as statistically, with estimates of productivity ranging from 33–47% [13]. Green and Lazarus characterise four aspects of low productivity as follows: 1) Process loss–aspects of group interactions inhibiting good problem solving; 2) Free riding–a participant does not contribute adequately to a meeting; 3) Conformance pressure–public setting of a meeting leaves participants reluctant to express disagreement with majority (or organisational) view; 4) Information Loss–a failure to capture meeting outcomes for future access. Any of these aspects may combine to contribute to an unproductive meeting. Several studies have explored both the technology involved in meeting browsers, as well as the user experience of existing and prototype meeting technologies.

Keeping Records

Given the frequency of meetings, meeting records form an important and rich source of information [9], documenting how the multi-party sequences of talk, gestures and actions work toward a shared goal. Outputs in the form of saved meeting notes, agreements for future action items, assigned tasks, issues resolved, etc., archive relevant information for ensuing meetings. Whittaker et al. identify two record types: public and private [38].. Public records are a contract of decisions and commitments, serving as a shared to-do list, resolving disputes, and recording decisions, actions, and the surrounding context. However, they can be laborious to produce, untimely, inaccurate, and fail to capture the experience. Personal records, in contrast, are produced during or immediately after a meeting often in the form of a cryptic and highly personalized reminding tool.

In addition to public and private records, any meeting may be recorded for future reference (in audio, video or transcription formats). However, while technology using audio visual recordings may provide an extremely rich contextual experience of a meeting, subsequently extracting the information disseminated during a meeting without the need to replay the entire recording is still a challenge. We sought to learn if the broader use of current technology might afford new practices around meeting records.

Information Retrieval

An important design consideration is what kind of information is sought about a meeting that has already happened. The top five queries made of a prototype meeting browser which recorded meeting content were: decisions made, participants/speakers present, topics discussed, agenda items, and arguments for decisions [8]. Studies suggest that 60% of queries about missed meetings relate to decisions, highlighting the potential value in using argumentative models of speech [31].

More practically, browsing an entire recording of a meeting using playback facilities helped users answer less than 20% of queries [2]. Nevertheless, post-meeting search activity rose to 25% success with inclusion of contextual data such as Topic search or Speaker ID, to help users navigate the recording [2]. To aid review of a long meeting for specific information, summarisation techniques and 'personalised browsers' are based on specific user requirements [18,23]. In linguistic modeling, progress has been made in identifying and extracting specified categories of utterance; decision points [17], action items [29], subjective statements [32], and detecting agreement and disagreement [11]. Geyer et al. review use of domain-specific indices to navigate meeting records retrospectively [12].

Evaluation of prototype tools

Evaluation of prototype tools developed for the meeting environment provide rich resource for future design of Automated Speech Understanding technology: A number of studies looked at value of meeting notes [19,31]. Kalnikaitė et al. [19] experimented with two novel markup tools used that created marks on the meeting transcript in real time. Both tools enhanced recall without compromising conversational contributions. However, the highlighter tool which required users to annotate spoken text during the meeting, increased the perceived workload of participants. Moreover, the errors in the ASR were found to distract users from the meeting in hand.

Another prototype system gave an importance score for transcribed words, removed unimportant utterances to create a summary 'gist' of meeting content based upon short samples from a corpus of meetings. Participants reported that this gisting helped them understand what had happened in a meeting they did not attend [36]. A prototype meeting system called VROOM was used to tease out both user concerns over technology for meetings, as well as design considerations [34]. Issues highlighted include rigid and inflexible workflows and unreliable technology as well as data loss or corruption and anxiety over employee surveillance and logged data access. Design guidelines stress the value of interviewing real working people for formative research, while the need to use errors as system training feedback and include the context of attendee profiles, time, and space to infer details of meeting location and attendees [5,10].

Speech-based intelligent agents

Speech-based interaction is an established feature in specific domains including hands free command and control in car navigation, where recognition rates of 78% - 87% have been achieved [14]. It has been already introduced in some aspects of flight traffic coordination, although human flight controllers must monitor and accept ASR into electronic strips to attain adequate accuracy [16] In health care, dictation technology is being developed as an aid to healthcare professionals in digitisation of diagnosis notes [40]. Speech-based agent systems have been created for specific contexts including assistive agent Billie, a prototype interactive scheduling assistant [39] and also in guiding visitors around a museum space [35,22]. Luger and Sellen characterise the particular form of command-dialogue assistants like Siri, Cortana, Amazon Echo and GoogleNow as a 'Conversational Agent' [24]. They report on interactional experiences of their everyday use based on interviews, finding user expectations to be 'dramatically out of step' with the actual capabilities of the system.

METHOD

This study uses a low-fi technology probe designed as an interview prompt around use of ASR, for understanding design implications for the particular context [4] of collaborative workplace meetings.

Pilot and design

As both automated speech recognition and intelligent agents are relatively new to the workplace meeting domain, this study was exploratory using a combination of a low-fi technical probe, observations, recordings, email survey and face to face interview techniques. The aim was to maximise the information gathered on the likely impact of these technologies in the workplace, rather than measuring specific variables such as recognition rates which are often used in evaluating speech-based algorithms.

Initial observational fieldwork in four meetings (two teams) allowed us to consider how a technology probe might be created to simulate an 'intent-algorithm': a software algorithm that can automatically select relevant utterances from meeting transcripts. Following similar work to repurpose everyday conversation occurring around mobile devices [25], a probe was created, to track how users might respond to a novel technology [4], in this case a speech-based 'Meeting Agent', which could automatically produce actionable items based on the transcripts of their meetings.

To create the probe, we manually simulated two technical stages of how such an agent would function: first, the ASR of meeting dialogue, followed by extraction of utterances from the transcripts, which contained relevant 'actionable items'. To achieve this, the audio of each meeting was transcribed and thoroughly checked for accuracy. One member of the team then manually highlighted actionable items within the transcripts, following guidelines of the Microsoft Cortana conversational agent data schema. An 'action item' is difficult to define, and we needed to agree how to identify which utterances to extract and present as 'action items' from the transcripts. We looked to the existing Microsoft Cortana data grammar guidelines which specify the mapping between input user utterances and corresponding machine-internal semantic representations. Cortana is designed and developed for the single user command-dialogue model for speech recognition, which is far more simplistic than multi-party meeting dialogue. However, despite these limitations we used the schema as it is well developed and it was being used by the wider research team as the starting point for developing algorithms for speech-based technology for the business domain. There were in total 42 actions available in the Cortana schema, (the schema is continually evolving). Based upon existing assumptions of what an action item is, 10 of these actions were identified as relevant to meeting scenarios: find

Type of meeting	Attendees	Duration	Total Action Items	Interviews
Daily production standup (T1)	10	32m	11	2
Daily production standup (T2)	7	34m	8	2
Weekly status–project leads (T3)	7	64m	10	4
Weekly status–project leads (T3)	7	60m	11	1
Weekly project status–all team (T4)	10	58m	4	2
Weekly project status–all team (T4)	10	56m	10	1
Weekly design review 'crit' (T5)	6	86m	3	2
Triweekly bug work allocation (T6)	2	33m	2	2
Weekly status research team (T7)	15	65m	7	1

Table 1. Meetings recorded with numbers of participants, action items detected in transcripts & post-meeting interviews

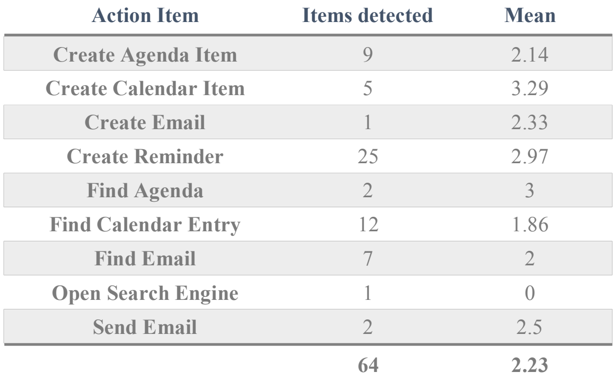
calendar entry, create calendar entry, open agenda, add agenda item, create single reminder, make call, search, find email, send email, and open setting. Those not used included applications such as navigation. With this subset in place, one member of the research team (to create a level of consistency) manually reviewed and annotated the transcripts, picking out utterances with relevant terms and words spoken such as dates, times, names, agenda items and so on [6]. The resulting list of utterances was then presented as a list of 'action items' on paper, for evaluation.

Participants and meetings

The first step was to gather a corpus of authentic meeting speech data working with seven mixed teams (T1-T7) in a multinational technology company, (Table 1). These were formal meetings in that they were arranged in advance, occur on a recognised frequency, involve invited attendees who have organisational roles which relate to the meeting objectives and are run by an assigned meeting facilitator [3]. The types of meetings are loosely categorised as:

- Production Standup: project status, short and high frequency
- Management: a broad agenda, weekly and monthly
- Design Review / Crit: collaborative design feedback, weekly
- Work Allocation: short and high frequency, tri weekly

The meetings were recorded using existing, in-room meeting technology resulting in both audio and video material. A backup recording was also made by a single, tripod-mounted camcorder.



Action Item	Items detected	Mean
Create Agenda Item	9	2.14
Create Calendar Item	5	3.29
Create Email	1	2.33
Create Reminder	25	2.97
Find Agenda	2	3
Find Calendar Entry	12	1.86
Find Email	7	2
Open Search Engine	1	0
Send Email	2	2.5
	64	**2.23**

Figure 1. Number of items detected and mean score

All meetings were already scheduled (i.e., none were set up for the study) via calendar invites, which included a link to join the meeting remotely over Skype. In total, nine meetings were recorded resulting in a corpus of 488 minutes of speech data. Two teams were recorded twice over two consecutive meetings, six meetings involved remote attendees and overall there were 57 participant attendees (18 female, 1 transgender, and 38 male). All participants were reimbursed with gift vouchers.

Collecting different perspectives on meetings

Artifacts were collected to represent the varying perspectives on a meeting: (i) the individual view of a regular meeting attendee, (ii) an external observer with no tacit 'domain knowledge', and (iii) simulated output of an automated intent algorithm (our probe). A researcher sat in each meeting as an external observer and took notes. For the view of individual attendees, an email survey was sent immediately after the recording of each meeting and 49 completed surveys were received. The attendees were asked to list the key items that they took from the meeting. They were also asked to describe their role in the meeting, and what they valued most from attending the meeting.

Follow up interviews

A subset of participants from each meeting were interviewed (Table 1), using the simulated action items as a probe to prompt discussion on the use of ASR technology in meetings–in total, 17 interviews were conducted. The work to transcribe the meeting audio and extract the simulated action items took some time to complete, (from 1 to 2 weeks). The time lag between the meeting (and email surveys) and the follow up interviews helped to ensure the participants were stretching their memory to recall the details of the meeting, similar to previous work [19] and similar to how they might engage a speech-based meeting agent. The interviews began with a project overview, then discussion of the participant's role and their usual practice in note taking–referring back to meetings retrospectively. The interview then reviewed the simulated 'action items' presented to them.

Participants were encouraged to 'think aloud' while reviewing all the actions, which were presented in a list on paper, with the actual words spoken in bold print, and an action type annotation attached to each utterance. They were then asked to score the accuracy and usefulness of each item on a scale of 1-5, with 5 as the best. Discussion followed regarding how well the items summarised the meeting, and they were asked to highlight errors and omissions where they occurred. In the latter part of the interview, participants were asked about their experience of technology in meetings in general and, with reference to the items already presented to them, probed on the potential impact of speech-based technology in their meetings.

FINDINGS

Our findings are presented in two parts: first, a summary of the recurring issues experienced by participants when presented with the list of simulated ‘action items’ from their meeting–including sample ‘action items’ presented with participant scores and comments. These are then followed by observations of the meetings, focusing on factors which may affect the success of speech-based meeting technology. These observations are introduced to provoke design ideas on what a speech-based meeting agent may be able to contribute to the collaborative meeting environment.

High versus low-scoring action items

The corpus was made up of nine meetings, extending to 488 minutes of recorded meeting data, which in turn yielded 64 action items. The distribution of action and their scores can be seen in Figure 1. This low number of items was unexpected–particularly given the number of people present in the meetings (as many as 15 attendees in one meeting which ran for 65 minutes and resulted in only 7 action items, see Table 1). In total we collected 100 ratings of 64 action items based upon how accurate and useful they were.

Reviewing participants’ ratings of action items with them allowed us to get an indication of which had worked and which had not. A number of accurately transcribed and categorised action items were successfully inferred from the transcriptions of the meeting (Figure 2). Top scores were given for items in the categories Create Reminder, Find Calendar Item, and Create Agenda Item. Participants responded positively and considered some to be viable output items. There were even instances when an action item was extracted from the transcripts which had been forgotten and not acted upon by the participant themselves:

“…if they ended up being more like, y’know, a list of 20 things that were like this five [referring to a high-scored action item], *which is ‘can I get two slides from you’, and you can do something… like, set a due date, set an owner and things like that, I think that would be pretty useful.”* **P29, meeting attendee**

The high-scoring items referred to future planning and collaboration, such as individual and group reminders for tasks to be completed or adding items to a future agenda.

In contrast, many of the action items were scored low by participants (56% were scored 2, 1, or 0, which was even below the 1-5 scale). These items, which suggest *prospective action* triggered during the meeting, such as launching an internet search application or opening the calendar of the meeting facilitator, were rated poorly, similar to McMillan et al.’s study of individual speech [25]. When asked to explain why, participants considered that these activities (e.g., launching search) were rather trivial, and unhelpful in the meeting setting. Indeed, some suggested that it would hinder rather than help collaboration if the system automatically opened additional items during a meeting where there are already documents opened and shared between a physical and virtual meeting space

Sample Action Item:
SpeakerV "I'm hoping Frank has something that we can start on. So I'll talk to him after this."
System to create a calendar entry for SpeakerV
P20 "This is actually pretty good. I would rate this four even, because yeah, I needed to talk to Frank, who is a peer PM of mine who understands (the server)… So this would have been a nice reminder to show up on my calendar automatically. So this is pretty good."

Figure 2. A viable action item which scored 4 out of 5

Sample Action Item:
SpeakerC “**I sent them to you. I never sent them out.**"
System to create a calendar entry for SpeakerF
1 out of 5, 2 out of 5, 1 out of 5 (3 participant scores)
P25 - No. 9 is frustrating. It makes me wanna figure out what that was. I just don't have the context, but I think it could be very useful, cause it could be something for me to follow up on. I think I am speaker C, so I feel like if I hadn't taken notes like I did in the meeting, I might have missed out.
P26 - It's probably very close to the exact words that were said, but it’s so conversational, that in this format I can't possibly tell what the meaning of that is.

Figure 3. Lack of contextual information

Low-scoring action items

Overall the action items were scored as poor, with a mean score of 2.23 per action (Figure 1). In explaining these low scores, our participants mentioned the items did not fit with the meeting or gave an incorrect summary of what was actually being discussed or intended. We characterise three reasons for the low scores: a lack of contextual information included in the action, a miscategorisation of dialogue and issues with errors in transcription.

Lack of contextual information

When presented with many of the actions, our participants expressed frustration that they were only getting to read a snippet of discussion, some of which had extended across a number of utterances. Selecting and presenting a single utterance as a proxy of an action item, often resulted in inaccurate action identification and attribution to the wrong people, since there was regularly confusion about who was the speaker or recipient of an instruction or comment (see Figure 3). In many cases, it was suggested that including more of what was said directly before and after the utterance extracted would help arrive at a comprehensible action item.

An important element in understanding the action items was the social structure of the meeting. While trying to make sense of the action items and checking the completeness of the content, a number of participants asked simply, “Who was in the room?”

“I wonder if there is a way to… know exactly who the attendees were. So one of the things you start to pick out is you have Barry, Chuck and Beth. Those were the main presenters for this meeting,

but then you have Tina and Adam, and a couple of other people scattered through–Colin... So knowing all of the attendees, and their roles is kind of a high-level thing for the meeting itself." **P33, meeting facilitator**

A speech-based meeting agent could draw upon organisational information regarding what roles people fulfill or who collaborates frequently with whom to augment the transcription of what is said in a meeting with details of speaker name, role, current projects and more. This social context might help users who review the output after the event, in the reconstruction of the meeting and in making the output more readily comprehensible.

Miscategorisation of Items

The probe labeled each identified item using the categories of the Cortana grammar as described earlier. Along with a lack of contextual dialogue, resulting in somewhat incomplete action item detail, participants found some of the low-scoring items had been assigned to the wrong category. This was misleading and could be heard in the think aloud to disrupt participants in their efforts to make sense of the utterances presented to them (see Figure 4).

In some instances, our participants could work around this as the simulated action items were listed together. However, as a meeting agent system develops, the categorised information would potentially become more separated and the information contained within each item could become effectively 'silo-ed'. This need for improved accuracy in categorisation highlights the requirement for flexibility–making it easy for users to amend category, and add to detail of each action item. Nevertheless, any additional work in fixing categories and augmenting critical details of action items to make them meaningful for the user serves to increase the user burden and reduce their confidence in the reliability of the technology. It also increases the likelihood of users reverting to prior practices of taking personal notes, or asking other attendees for a summary of what happened in a meeting they missed:

"It's making sure that I don't have to do as much work. I guess missed information would be one, either it missing information, or two, it not capturing anything at all. Cause it would be so sad if you pick it up and it's like, "I didn't capture anything." **P28, meeting attendee**

Sample Action Item:
SpeakerA **"So I wanted to take more time to, we only have five minutes but just to get your suggestions on what else we can do, like what other technical debt we have. I'm sure there are. But…"**
System to create an agenda item
1 out of 5
P41 "So the sixth one (action item) is I think the first one which is actually an actionable tp-do for us… Like the action was to send team manager a mail with our ideas – which is not actually captured by the action here. Wrong category – should be a task reminder for whole team."

Figure 4. Miscategorisation of items

Sample Action Item:
SpeakerF **"Yeah. I'll talk to Jill and see. Yeah. Yeah. I updated the deck though so"**
System to create a reminder for SpeakerF (Lucy)
4 out of 5
P21–"Jill? I'm not sure who Jill is. This is probably when they were working on updating the slide deck because we were presenting a deck with the capacity to some higher up people, and some folks were working on that to get the information/… [P21 continues trying to make sense of action item] P21–I'll talk to Jill and see, and see what? I don't know. I'm not sure if the Jill is an error or not. Oh, it might be Phil, Phil is Lucy's manager."

Figure 5. Errors in transcription

In addition to the risk of this technology providing excessive false positives and poor categorisation, which would give the user extra work to sift the wheat from the chaff, the worst-case scenario would be a system that missed important action items altogether:

"It missed some of the actions for me. Yeah. And it didn't capture anything for Brian, I think." **P22, meeting attendee**

In this meeting, the interviewee had had two distinct action items discussed. Our probe detected only one of them–in addition the interviewee noted that our system entirely missed actions that were assigned to his colleague, Brian. These omissions were attributed to neither action items being discussed at length during the meeting, raising further concern that items may be overlooked by the system.

Errors in Transcription

Participants regularly found the transcription of utterances from their meetings difficult to understand for a number of reasons. The following example (see Figure 5) shows how a small error in the transcription–mistaking the male name Phil for female name Jill–makes the item confusing, misleading and time wasting. Yet this instance highlights the extracted utterances could be better understood with provision of accurate speaker identification, attribution and recipient disambiguation (diarisation) throughout, as well as organisation relational details–had it been presented along with the transcription. Had Speaker F been clearly identified in the transcription and associated with her manager, Phil, together this information could have improved the comprehensibility of the transcription, and the participant may have been able repair the name error and made sense of this action item with greater ease. Nevertheless, a recurring reason for low scores was the difficulty participants had in making sense of the raw utterances of spoken dialogue regardless of the accuracy of the transcription involved.

There were a number of factors, which contributed to the variability of the transcript quality including the challenge of recording multi-party audio. Difficulties arise because conversation is produced through interaction with multiple others, alongside unpredictable environmental incidents and disruptions, which can affect the audible sound level on a practical level dependent upon how far the speaker is from the microphone. More than this, multi-party conversation can also unexpectedly change direction in what is discussed, and at an even more granular level it can affect what is said from one word to another since co-present individuals not only listen to each other, but also take visual and audio cues as to their recipient's or audience's state. Interaction goes back and forth like this to achieve mutual understanding, and as quickly as one topic is raised, another may replace it if it is more pressing to the flow of the present interaction [3]. Harper refers to this blending of verbal and social cues used in the act of conversation as 'structural patternings' [15]. Consequently, the transcribed dialogue may be difficult to make sense of when taken out of context, as seen in the following quote from a participant who reads the transcription of his own status update in a meeting from some weeks earlier. He can hardly be certain that the transcribed dialogue are his own words, and that the action item being called out is actually for him.

P22: *This is definitely–I'm pretty sure this is something I said. [laughter]*
Researcher: *Do you remember saying this?*
P22: *When you see it written–it's stuff that I was working on, but I don't remember saying it. [reading] That was more of a status part. Yep, this was definitely an action item that I did, and followed up with. So that is–let's see. [reading] It sounds like gibberish when you transcribe what we say. [laughter]* **P22, meeting attendee**

Observations of workplace meetings

Clearly there was more to the meetings than our relatively crude agent was picking up. We wanted to see if we could go into more depth on what was happening in our meetings. Drawing on the video recordings of collaborative workplace meetings, as well as upon observations of their functioning made during the fieldwork, allowed us to look at the nature of meetings themselves and the challenges they bring as genera of communication for speech agents.

The observations remind us that what is remembered of, valued from and achieved in meetings is diverse and variable. This variability may be accounted for by the different purposes those meetings serve for different individuals attending. While it may be tempting to see the output of group meetings to be clean, orderly and functional, in reality, organisational studies suggest that a rich diversity of activities occur: peer relations, negotiations, team motivation, resolving conflict, establishing information networks, disseminating information, making decisions amidst ambiguity, and allocating resources [27].

In particular we discuss now three elements of meetings which a speech-based meeting agent may struggle to address: 1) the individual information needs of participants; 2) the different perspectives on workplace meeting outcomes; and 3) the importance of social interaction in meetings.

Individual information needs of meeting participants

In an earlier study of 'computing tools' to support knowledge work, Kidd defined three types of office worker: 1) the knowledge worker, 2) communication worker and 3) clerical worker [21]. These types differed with respect to how they deal with information and how they consequently managed documents. A clerical worker was characterised as one who handles or manages the output - or documents - of someone else. The information they deal with is therefore 'extrinsic' to them, for example an HR employee is required to implement established company policies. The working practices of a clerical worker tend to be structured, and their output is more predictable as a result. Consistency and predictability of output is the desired goal for a clerical worker and opportunities to introduce computerised processes are more apparent within this category of office worker. On the other hand, knowledge workers, which include our participants, have an altogether more 'embodied' experience with information–they are changed by the information. To paraphrase Kidd, a knowledge worker is changed by the information they process, and their value as an employee is to understand a body of knowledge and to generate new–potentially unique–information, which is directly relevant and valuable for their organisation [21]. In this way, the output of a knowledge worker is 'intrinsic' information and as such, a more unpredictable entity. How each meeting attendee makes sense or use of information shared in a meeting is intrinsic to them and their specific needs, and therefore what they take away from a meeting is also unique, varied and unpredictable. ASR technologies in the workplace may fare better in domain of more predictable clerical work, where replicability is the aim.

Different perspectives on workplace meetings outcomes

To get an understanding of the inter-variability of what our participants thought was important or valuable about the meeting they had just attended, the post-meeting email survey asked each one to list the key points or notes that they took away from the meeting. The personal notes (Figure 6) gathered from four attendees after meeting T3 show that while there is some content overlap, the list also reveals a high degree of inter-variability of perspective among participants. Moreover, the notes are often idiosyncratic and written in a form understood only by the author, or possibly others in their work domain. The notes could be best described as memory aids rather than an explicit record of events and discussions in the meeting [38]. Indeed one participant (P28) gave just a list of key words to indicate what took place (Figure 6).

Participant ID	Personal Notes
P24	• My take-away was to continue to drive consistency of our development efforts across the org
P26	• Jimmy shared the framework of a model used to think about what internal and external factors influence MAU/DAU (Monthly/Daily Active Users), how we obtain actionable information from various signals and feedback that information to attempt to influence the direction and stability of the product • Org event moved to 6/25. National park hike • Discussion on how various teams in department are approaching development work. We're going to look into best practices across our organization • Further discussion on data analytics F-team including the understanding the Consumer and Enterprise are completely separate
P27	• Follow up on Data Analytics team. Set expectations with Lead IC to ensure charter is clear and scope is consistent with expectations
P28	• Vin Smith (senior mgr) visit next week • Consumer Model–Jimmy • New org name • Org Morale event @ National Park • Our team adopting mature SDLS-like process for development • Data Analytics F-team •

Figure 6. Participants' personal notes for meeting T3

Comparing these personal notes created by participants with the action items identified by our simulated system in the transcript of the same meeting (Figure 7), there is clearly a considerable mismatch between what the individuals took away from the meeting and what might be extracted by a speech-based agent.

ID	Utterances extracted from transcript	Action Item Type
C	"Vincent would be here. I think at ten AM but he's got just like a bit of time. I think he was thinking of coming early. And Allie was like there's no point in coming too early because no one will be here. So then I don't know, we'll have to check if it'll actually happen. So I don't know a ton of details but there might be an all hands Tuesday morning"	Find agenda item
C	"Anyway moving on. Org name coming, don't know what it is. Our org event will be the twenty sixth of this month going to"	Create calendar item
C	"Okay. And then Joe off today. So he's definitely not demoing his new tool which is off the agenda. What? Well you're ready."	Find calendar entry
F	"I totally should have read the agenda. My bad."	Find agenda item
B	"So, um, we'll wait until next week. I'll bring it up again with, uh, Joe, or at least have him connect with you"	Create reminder
E	"So, couple of items. One was I'll be meeting with Fritz when I'm in Redmond this week. And I wanted to get him started off with the, uh, kicking off the core team. So he can start having these discussions with the smaller core team and they can figure out exactly what the goals are, and fine tune, et cetera."	Find calendar entry
D	"So, our next senior leads meeting will be in a month, right."	Find calendar entry
B	"Twenty six oh five where is that?"	Open search
D	"So, I'm gonna be gone for the rest of the week."	Find calendar entry
D	"Ohh okay, when he's coming into the office"	Find calendar entry

Figure 7. Simulated action items from meeting T3

A number of teams were involved in daily production 'standup' meetings (Table 1), which follow a 'Scrum' format; they are time limited, and each member of the team takes a turn to describe what they've done yesterday, will do today and describe any 'blockers' or impediments. The content of these meetings is heavily 'informational' in nature–participants exchange ideas on solutions, tips and advice. The general management meetings in the corpus similarly focused upon dissemination of project and company related information. In contrast, the two smallest meetings–the work allocation and the design review sessions–were less informational. These small, 'monofocal' meetings were task and decision oriented [3], dealing as they did with decisions regarding allocation of product support, and product design concept creation and critique.

The transcripts for the different meeting types (production, management, monofocal) show recognizable domain-specific artifacts [12] in the keywords, objects, and processes used. As teams work together over time they develop indices as points of reference within transcripts. The Cortana grammar for action items categories used by the probe was badly mismatched to the meeting transcripts. Categories could be adapted and trained to be better aligned to the domain-specific artifacts.

To understand what individual participants valued from meetings, each was asked in the post-meeting email survey to describe what they considered to be the most important aspect of the meeting they had just attended. The 49 responses revealed some common themes, which could be grouped into five categories (Figure 8).

Three of the categories refer to information discovery which cannot be reliably sourced elsewhere–i.e., receiving news from senior management, giving and receiving help from others and finally hearing what their colleagues think and say. Mintzberg suggests that managers play a key role in, "securing 'soft' external information (much of it available only to them because of their status) and in passing it along to their subordinates" [27]. The responses reflect this and the need to synchronise one's own status across work colleagues. There is some overlap between the categories, which might be described as 'shared state'. What is notable, is the lack of reference to action items or other clerical interactions. Rather, emphasis is on the value of social and informational aspects of meetings. This is consistent with the low overall number of items detected.

Importance of social interaction in meetings

Broadly, all the meetings were mechanisms for improved inter-personal and organisational communication. Managers shared company news with their teams who reciprocated with thoughts and feedback, workers helped their colleagues to solve technical problems, shared their current status and heard what the next step in their project would be, as well as shared everyday troubles and successes. When asked if they would refer back to a full transcription of the meeting, participants commented that this would be difficult, time-consuming and ineffective, adding that the shared notes prepared by an appointed person (usually the meeting facilitator) would be better. This indicates that the facilitator's notes provide interpretation, or encoding, rather than simple external storage of information.

Category	Total Responses	Sample Response
1. Status update and information share	[21]	"Quick status on work items for this sprint from team members. Awareness of any blockers that might be blocking/slowing the team down."
2. Find out what management are planning and doing	[7]	"Update from manager about what's happening across the team, what the higher ups in the management chain are thinking, what new projects are on the horizon, etc."
3. Give and receive help and advice	[7]	"I usually get some reminder about a task or deliverable that I had deprioritized, and that is helpful."
4. Hear what people think.	[6]	"At a high level, I think I most value the chance to engage with peers who have the same challenge I do in leading teams and achieving through others. We do this ad hoc, bit its good to have time set aside each week to reinforce this, share ideas and sync on shared goals."
5. Make decisions	[2]	"Consistency (in bug allocation) is important!"

Figure 8. What do you value in meetings?

Alongside the low instance of note-taking during the meetings, participants said they seldom ever referred back to the notes. Indeed, the interpretative activity of producing summarizations in the form of notes provides adequate information processing for meeting facilitator P24:

"Oh. For me I think it's easier to jot down what I believe are the salient points during the meeting than it would be to look at the transcript and have to pull them out. I don't know how often I'd go back (to the transcript); even just my notes, I don't go back to them that often." **P24, meeting facilitator**

Earlier studies, suggest that although 64% of managers keep meeting notes for considerable length of time, as many as 44% said the notes were rarely, if ever, consulted [20,21]. This infrequency of re-visiting meeting notes suggests that the social interaction within the meeting is more important to attendees, than output in the form of notes or action items. Participants are motivated by encoding information during the meeting, rather than collecting data to external storage for future reference.

Analysis of the activities reveals strong similarities in the role of chairperson or facilitator across this diverse collection of workplace meetings [37]. However formally it is recognised, the role of chair has functional importance for opening and closing meetings, managing the progress of the meeting via activities tracking the agenda, and steering topic in general. Regardless of the context or content of a meeting, some form of agenda management occurs [1] and in the next snippet (Figure 9) the meeting chair makes use of the agenda to wrap up discussions which had gone off into unnecessarily speculative detail, by providing a 'gloss' or summary of the discussion so far, (line 4). These repeated utterances are designed and delivered to wrap up current topic discussion, and they signal readiness to move on and provide us with a view on how meeting discourse is interactionally arranged, as well as the key importance of the facilitator and agenda. These subtle interactions facilitate progress to the next topic, but they also help to build agreement and record points of information. These social cues serve to make a meeting recognisable for participants, and provide an interactional boundary [3].

```
1.  B: Anyway.
2.  Chair: So that's why it's like foundation.
3.  D: Yes.
4.  Chair: Anyway moving on. Org name coming,
5.  don't know what it is. Our org event will be
6.  the twenty sixth of this month going to.
7.  B: bup burra ba! Coyote National Park.
8.  E: Ahh, nice.
9.  Chair: [So the]
10. B:     [It's just easiest]
11. Chair: We will provide bus transportation,
12. leave at ten and aim to be back between four
13. thirty and five.
```

Figure 9. Facilitator manages meeting progress

Meetings were brought to an end by initiation of closing statements and gestures, but the discourse and interactions did not necessarily finish concisely, and meetings regularly broke up into smaller, informal exchanges that took place in or directly nearby the meeting venue. The audio recordings sometimes captured these discussions, and the post-meeting exchanges were often important and consequential and moved from the general informational nature of the meetings towards decisional talk, in smaller groups.

While our data here reveal stable, recognisable meeting interactions–these are neither automated nor deterministic. The specific content and meaning is produced each time, within its own context and the unpredictable nature of meetings makes it hard for the current state of ASR technology to support and extract action items. Nevertheless, by this structured orderliness a group of people work their way through a meeting and information emerges, agreement is sought and tasks arise. While they are difficult to design support for, these are the complex social interactions and outcomes, which make organisations.

DISCUSSION

The expansion in functionality and popularity of commercial systems like Siri and Cortana, has raised the prospect of introducing ASR to form a system agent that could proactively support work in the business environment, already a site of technology supported collaborative work. Automated transcription and mining of all words spoken in a meeting, offers the apparent promise of increasing productivity by giving access to everything said in a meeting and thus resolving the 'information loss' issue [13]. We created a probe to investigate the connection between what action items a speech-based agent might produce based upon the talk heard in a meeting, and what the participants in reality took away from the same meeting. The probe was created by using human speech recognition, and the action items were selected from the transcripts by a human. Nevertheless, the output of the simulation–the action items–were rated poorly by participants and the system proved a failure. Why did these items fail so badly? The results suggest that ASR alone is not the problem. Even with the best possible transcription, it is not feasible to understand and elicit accurate action items, or points of reference, from transcribed dialogue alone. It would appear that there are no simple rules, and documenting action items is too complex for a speech-based system alone.

To begin to grasp what constitutes that complexity, our observations of meetings introduce important dimensions of meetings which were missed by the transcribed records of talk, each of which has design implications for speech-based agents in collaborative workplace meetings.

First, the low instance of action items in the meetings was surprising for us. Yet, this low number combined with the varied views on what participants valued from their meetings, gave a clear indication that for those attending the meetings, the action items were not the main attraction. Our initial assumptions about action items were inaccurate in three ways: 1) the frequency of action items in collaborative workplace meetings, 2) what constitutes an action item and 3) the perceived value to participants of those action items.

Contextual Access

Making sense of the action items when viewed after the meeting and out of context, proved difficult due to a combination of factors. A lack of contextual information, which could help participants to make sense of the extracted utterance, was often cited as the problem, along with errors in the transcription. These two issues could be alleviated with the inclusion of contextual information–including ready expansion of the extracted utterance to give the user the preceding and following text spoken in the discourse. Additionally, the extracted utterances could be better understood with provision of accurate speaker identification and attribution (diarisation) throughout, as well as organisation relational details–to help recreate the context of the meeting. Nonetheless, the struggle to make sense of many of the action items confirms that while meeting interactions and discourse are sequential and meaningful when experienced in situ, they are difficult to fathom when reviewed out of context, and the process of retrospective sense-making is readily disrupted by simple transcription errors.

Impact of Misrepresentation on Collaboration

What was not said explicitly in meetings was omitted or overlooked by our system. This raises the potential of ASR technology to misrepresent what was discussed in a meeting, which in turn may encourage 'grand-standing' by those attendees who want to ensure their contribution to a meeting is recorded. The findings also show that errors in transcription is still a massive obstacle to smooth uptake of ASR technology in collaborative settings. Users already regularly adopt workarounds for meeting technology, like video conferencing, to work adequately to get the job done. As users learn what the technology can recognise and interpret in the meeting setting, will this result in the formalisation and reduction of talk and behaviour in meetings? The aspiration of this new technology is to combine the live meeting discourse with external information from the 'office graph', including emails, calendar items along with broader organisational information to provide 'intelligent' suggestions to fill informational gaps. Nevertheless, a frequent concern expressed in interview was that this technology could result, inadvertently, in the 'indexing' of each attendee's contribution to the meetings, and participants were keen to understand who would have access to the meeting transcriptions. More broadly, there was a concern that documenting meeting discourse may stifle open and free discussion within group meetings [33].

Flexible and Transparent Information

One of the difficulties of processing content via an automated agent is the problem of when and how to categorise the information. In the case of some items in our study, they were correctly identified as potential actionable items however, they were classified into the wrong category. While information is in the process of being shared and informing different recipients, it should not be categorised. To do so runs a high risk of reducing the value of the meeting. The same information can mean different things to different 'knowledge workers' and when transcripts of meeting discourse are categorised, this effectively locks the information into one bucket, where it may remained unused. As Kidd describes, information transforms the recipient uniquely, 'the mark is on the worker', as they make sense of the information within their particular context [21]. For as long as meeting information is being processed by recipients, what is required are technologies which are flexible and generative since the same information can be meaningful and used in different ways by different individuals.

Domain-Specific Artifacts

The meeting transcripts revealed recurring use of jargon and words that allowed meeting members to deal quickly in a way that they, as a group, understand. Developing unique vocabularies, which include the key terminology used within a meeting for different domains, departments, and even teams, would accelerate recognition of the talk by a speech-based agent. Getting the terms right for the audience was scored more highly than prospectively launching actions within a meeting, similar to McMillan et al, who found "both people and activities have identifying keyword clusters."

Supporting the Facilitator

The material shows that the role of meeting leader or facilitator has not been diminished by advances in technology for the workplace [21]. Our participants would look to their meeting facilitator to provide notes which they expected to include interpretation and filtering of discussions that took place in a meeting. Moreover, they were not confident of the reliability of ASR in terms of accuracy and categorisation, as Luger and Sellen found with users of conversational agents [24]. On the other hand, command dialogue with an intelligent meeting agent was considered acceptable, and would be welcomed by some participants who expressed keen interest in a technology which would behave in the meeting setting like Cortana or Siri, and respond to spoken commands at points initiated by the participants themselves, who identified this as suitable means of handling administrative action items, such as "let's put the next design review in the diary now", or "send a reminder to produce two summary slides by Tuesday".

Designing to support the meeting leader or facilitator seems the more fruitful way to introduce meeting agent technology to collaborative workplace settings. Other, participatory methods of summarising and documenting meeting outputs have explored the use of short co-created summary videos [7] to record what was discussed, agreed, or assigned in meetings.

Study limitations

While the probe here focused on how speech input to the simulated system might be output for use during or after the meeting, the form of their delivery was not discussed. The timing and manner of feedback could be critical to their perceived utility. In addition, the Cortana schema used here while well developed, is highly constrained. Finally, the participants were recruited voluntarily from one very large organisation; there was a notable absence of internal functional teams like HR. Other domains may be differently disposed to adoption of speech-based technology.

CONCLUSION

Our observations of the varied set of collaborative workplace meetings here show us that items of value to participants are endogenous to the interaction of the meeting, and as such are rarely separable entities. Instead they rely upon the interpretative skill of a meeting facilitator/leader to summarise and communicate items effectively. Meetings are complex, generative interactions between multiple participants, rather than passive acts of production of simple data for storage. Productive work is being done by the provider and recipient of information in a meeting: both unique and optimised to the moment and context of production.

To support meetings with technology we need to understand their complexity better. We give an initial step towards that understanding here, first by exploring how a speech-based agent might perform by conducting a probe to extract action items from collaborative meeting transcripts. We then highlight that these items represent only an extremely small part of workplace meeting interaction, and through observation we outline the diversity in individual informational needs, the varying perspectives on meeting outcomes as well as the importance of social interaction within meetings. Future work could look to document the diversity of meeting domains, as well as participatory roles. With these domain-specific data, the work to extend and refine the vocabulary and classification of automated speech recognition algorithms could progress.

ACKNOWLEDGEMENTS

We thank the larger group that was developing the speech technology and meeting agent, including Andreas Stolcke and Dilek Hakkani-Tur, for their technical perspective on this project. We also thank the anonymous participants and their teams who let us record and analyse their meetings.

REFERENCES

1. J. O. Angouri and Meredith Marra. 2010. Corporate meetings as genre: a study of the role of the chair in corporate meeting talk. *Text & talk* 30, 6: 615–636. https://doi.org/http://dx.doi.org/10.1515/text.2010.030
2. Satanjeev Banerjee, Carolyn Rose, and Alexander I. Rudnicky. 2005. The necessity of a meeting recording and playback system, and the benefit of topic–level annotations to meeting browsing. In *Human-Computer Interaction-INTERACT 2005*. Springer, 643–656. http://link.springer.com/chapter/10.1007/11555261_52
3. Deirdre Boden. 1994. *Business of Talk*. Wiley.
4. Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI Interprets the Probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 1077–1086. https://doi.org/http://dx.doi.org/10.1145/1240624.1240789
5. Hsinchun Chen, A. Houston, J. Nunamaker, and J. Yen. 1996. Toward intelligent meeting agents. *Computer* 29, 8: 62–70. https://doi.org/http://dx.doi.org/10.1109/2.532047
6. Yun-Nung Chen, Dilek Hakkani-Tür, and Xiaodong He. 2015. Detecting actionable items in meetings by convolutional deep structured semantic models. In *Proceedings of ASRU*. https://doi.org/http://dx.doi.org/10.1109/asru.2015.7404819
7. Brendon Clark. 2016. One-Shot Video | Interactive Institute. https://www.tii.se/one-shot-video
8. A. H. M. Cremers, B. Hilhorst, and APOS Vermeeren. 2005. What was discussed by whom, how, when and where? Personalized browsing of annotated multimedia meeting recordings. *Proceedings of HCI*: 1–10.

http://scholar.google.com/scholar?cluster=12768619403359757807&hl=en&oi=scholarr

9. Richard L. Daft and Robert H. Lengel. 1983. *Information Richness. A New Approach to Managerial Behavior and Organization Design*.
10. Patrick Ehlen, Matthew Purver, John Niekrasz, Kari Lee, and Stanley Peters. 2008. Meeting Adjourned: Off-line Learning Interfaces for Automatic Meeting Understanding. In *Proceedings of the 13th International Conference on Intelligent User Interfaces* (IUI '08), 276–284. https://doi.org/http://dx.doi.org/10.1145/1378773.1378810
11. Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics* (ACL '04). https://doi.org/http://dx.doi.org/10.3115/1218955.1219040
12. Werner Geyer, Heather Richter, and Gregory D. Abowd. 2005. Towards a Smarter Meeting Record--Capture and Access of Meetings Revisited. *Multimedia Tools and Applications* 27, 3: 393–410. https://doi.org/http://dx.doi.org/10.1007/s11042-005-3815-0
13. Walter A. Green and Harold Lazarus. 1991. Are Today′s Executives Meeting with Success? *Journal of Management Development* 10, 1: 14–25. https://doi.org/http://dx.doi.org/10.1108/02621719110139034
14. S.W. Hamerich. 2007. Towards advanced speech driven navigation systems for cars. 247–250. https://doi.org/http://dx.doi.org/10.1049/cp:20070376
15. Richard Harper. 2010. *Texture: Human Expression in the Age of Communications Overload*. The MIT Press. http://dl.acm.org/citation.cfm?id=1941863
16. Hartmut Helmke, Jürgen Rataj, Thorsten Mühlhausen, Oliver Ohneiser, Heiko Ehr, Matthias Kleinert, Y. Oualil, and M. Schulder. 2015. Assistant-based speech recognition for ATM applications. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)", Lisbon, Portugal*. http://www.atmseminar.org/seminarContent/seminar11/papers/363_Helmke_0120151059-Final-Paper-4-28-15.pdf
17. Pei-Yun Hsueh and Johanna Moore. 2007. What decisions have you made: Automatic decision detection in conversational speech. In *In NAACL/HLT*. http://www.research.ed.ac.uk/portal/files/7771732/N07_1004.pdf
18. Pei-Yun Hsueh and Johanna D. Moore. 2009. Improving Meeting Summarization by Focusing on User Needs: A Task-oriented Evaluation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (IUI '09), 17–26. https://doi.org/http://dx.doi.org/10.1145/1502650.1502657
19. Vaiva Kalnikaitė, Patrick Ehlen, and Steve Whittaker. 2012. Markup as you talk: establishing effective memory cues while still contributing to a meeting. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 349–358. https://doi.org/http://dx.doi.org/10.1145/2145204.2145260
20. Fawzia Khan. 1993. *A survey of note-taking practices*. Hewlett-Packard Laboratories.
21. Alison Kidd. 1994. The marks are on the knowledge worker. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 186–191. https://doi.org/http://dx.doi.org/10.1145/191666.191740
22. Stefan Kopp, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents*, 329–343. https://doi.org/http://dx.doi.org/10.1007/11550617_28
23. Agnes Lisowska, Andrei Popescu-Belis, and Susan Armstrong. 2004. User query analysis for the specification and evaluation of a dialogue processing and retrieval system. http://archive-ouverte.unige.ch/unige:2264
24. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5286–5297. https://doi.org/http://dx.doi.org/10.1145/2858036.2858288
25. Donald McMillan, Antoine Loriette, and Barry Brown. 2015. Repurposing Conversation: Experiments with the Continuous Speech Stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3953–3962. https://doi.org/http://dx.doi.org/10.1145/2702123.2702532
26. Robinson Meyer. 2015. Even Early Focus Groups Hated Clippy. *The Atlantic*. http://www.theatlantic.com/technology/archive/2015/06/clippy-the-microsoft-office-assistant-is-the-patriarchys-fault/396653/
27. Henry Mintzberg. 1975. The manager's job: folklore and fact. *Harvard Business Review* 53, 4: 49–61. https://ezp.sub.su.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=3867274&site=ehost-live&scope=site
28. Roger K. Moore. 2013. Spoken language processing: where do we go from here? In *Your Virtual Butler*, Robert Trappl (ed.). Springer-Verlag, Berlin, Heidelberg, 119–133. http://dl.acm.org/citation.cfm?id=2554494.2554508

29. Gabriel Murray and Steve Renals. 2008. Detecting action items in meetings. In *Machine Learning for Multimodal Interaction*. Springer, 208–213. http://link.springer.com/chapter/10.1007/978-3-540-85853-9_19
30. Gabriel Murray and Steve Renals. 2008. Detecting Action Items in Meetings. In *Machine Learning for Multimodal Interaction*, Andrei Popescu-Belis and Rainer Stiefelhagen (eds.). Springer Berlin Heidelberg, 208–213. http://dx.doi.org/10.1007/978-3-540-85853-9_19
31. Mukesh Nathan, Mercan Topkara, Jennifer Lai, Shimei Pan, Steven Wood, Jeff Boston, and Loren Terveen. 2012. In Case You Missed It: Benefits of Attendee-shared Annotations for Non-attendees of Remote Meetings. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (CSCW '12), 339–348. https://doi.org/http://dx.doi.org/10.1145/2145204.2145259
32. Stephan Raaijmakers, Khiet Truong, and Theresa Wilson. 2008. Multimodal Subjectivity Analysis of Multiparty Conversation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP '08), 466–474. https://doi.org/http://dx.doi.org/10.3115/1613715.1613774
33. Felix Stalder and Christine Mayer. 2009. The Second Index. Search Engines, Personalization and Surveillance (Deep Search) | n.n. -- notes & nodes on society, technology and the space of the possible. http://felix.openflows.com/node/113
34. Phil Thompson, Anne James, and Antonios Nanos. 2013. V-ROOM: Virtual meeting system trial. 563–569. https://doi.org/http://dx.doi.org/10.1109/CSCWD.2013.6581023
35. David Traum, Priti Aggarwal, Ron Artstein, Susan Foutz, Jillian Gerten, Athanasios Katsamanis, Anton Leuski, Dan Noren, and William Swartout. 2012. Ada and Grace: Direct interaction with museum visitors. In *Intelligent Virtual Agents*, 245–251. https://doi.org/http://dx.doi.org/10.1007/978-3-642-33197-8_25
36. Simon Tucker, Ofer Bergman, Anand Ramamoorthy, and Steve Whittaker. 2010. Catchup: a useful application of time-travel in meetings. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 99–102. https://doi.org/http://dx.doi.org/10.1145/1718918.1718937
37. Stephen Viller. 1991. The Group Facilitator: A CSCW Perspective. 81–95. https://doi.org/http://dx.doi.org/10.1007/978-94-011-3506-1_6
38. Steve Whittaker, Rachel Laban, and Simon Tucker. 2006. Analysing Meeting Records: An Ethnographic Study and Technological Implications. In *Machine Learning for Multimodal Interaction*, Steve Renals and Samy Bengio (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 101–113. http://link.springer.com/10.1007/11677482_9
39. Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. 2013. Virtual agents as daily assistants for elderly or cognitively impaired people. In *Intelligent virtual agents*, 79–91. https://doi.org/http://dx.doi.org/10.1007/978-3-642-40415-3_7
40. Julián Zapata and Andreas Søeborg Kirkedal. 2015. Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows. In *Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania*, 201–210.