"*Do Animals Have Accents?*": Talking with Agents in Multi-Party Conversation

Martin Porcheron*, Joel E. Fischer*, and Sarah Sharples†
*The Mixed Reality Laboratory †Human Factors Research Group
School of Computer Science Faculty of Engineering
University of Nottingham, UK University of Nottingham, UK
{martin.porcheron, joel.fischer, sarah.sharples}@nottingham.ac.uk

ABSTRACT
In this paper we unpack the use of conversational agents, or so-called *intelligent personal assistants* (IPAs), in multi-party conversation amongst a group of friends while they are socialising in a café. IPAs such as *Siri* or *Google Now* can be found on a large proportion of personal smartphones and tablets, and are promoted as 'natural language' interfaces. The question we pursue here is how they are actually drawn upon in conversational practice? In our work we examine the use of these IPAs in a mundane and common-place setting and employ an ethnomethodological perspective to draw out the character of the IPA-use in conversation. Additionally, we highlight a number of nuanced practicalities of their use in multi-party settings. By providing a depiction of the nature and methodical practice of their use, we are able to contribute our findings to the design of IPAs.

Author Keywords
conversational agents; intelligent personal assistants; mobile devices; smartphones; multi-party conversation; collocated interaction; ethnomethodology; conversation analysis.

ACM Classification Keywords
H.5.3. Information interfaces and presentation (e.g., HCI): Computer-supported cooperative work; H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Mobile devices are pervasive social objects that permeate all parts of our everyday life [43,49]. Significant research has investigated the ways in which device interaction is socially embedded in the context within which their use occurs. A number of studies have examined the interactional methods through which people in different settings interleave their device use within their daily activities, such as watching television together [37], sitting around the dining table [11], and socialising together in pubs [36,45]. Other work has also considered the implications of everyday device interactions beyond mere smartphones, instead considering the use of smartwatches [35] and smartglasses [8]. In a similar vein, we turn to a novel interaction technique found on many devices of the last five years: speech input, and in particular, the *conversational agents* found on smartphones and tablets. On most existing devices, an agent may be triggered through one of two means: by pressing a physical or on-screen button, or by the utterance of a 'magic phrase' that serves as a conversational opener (e.g. *"Hey Siri"*). The human interlocutor (i.e. 'the user') then *talks* to the agent, and is able to engage in dialogue and ask questions (e.g. about the weather), or give commands (e.g. to call someone); the IPA responds either by speaking back or by displaying a response on the device's screen. In essence, the agent is a *natural language interface* to the device's existing functionality.

We adopt the industry-preferred term *intelligent personal assistants* (IPAs), but not uncritically so; marketing materials suggest that IPAs interact like any person might, and can respond to natural human talk. For example, both *Siri* (Apple Inc.) and *Cortana* (Microsoft Corporation) appear to exude humour in response to general conversational input, questions, and commands. In turn, their responses to a human conversational partner might be seen as sarcastic or entertaining. While IPAs may provide the veneer of conversational intelligence, our study examines just how IPAs are actually used in conversations in order to provide design insights grounded in empirical evidence.

Marketing materials further position the natural language interface of some IPAs as explicitly supporting multi-party environments like the home (e.g. *Amazon Echo*), suggesting any member of the party can 'just talk' to the device. The assumption we find intriguing is that a natural language interface makes device interaction directly observable-reportable [13] (and thereby accountable) to others who are present. Indeed, research has shown that accounting for device use is a critical feature in multi-party settings; for example, individuals interacting with mobile devices via touch employ various methods to account for their device interactions (such as making the screen visible for others, or verbalising what they are doing on the screen) [5,36,50]. Therefore, we examine how IPA use in multi-party settings *actually* occurs, for which we adopt a conversation analytic

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/10.1145/2998181.2998298

approach [40], drawing on ethnomethodology. Through our analysis we uncover members' practical reasoning about an IPA's performance by orienting to the accountable actions of people and IPAs during a multi-party conversation. In this paper we present a number of fragments of data as *vivid exhibits* [2] of the situated activities undertaken by the members in the setting when using an IPA. In turn, we make a number of contributions to the CSCW and HCI communities: we identify how the content of queries to the IPA are formulated, how people talk to the devices, how members orient to the use of the IPAs in the setting, and ultimately and yet most bluntly, we consider the character of what talk looks like with IPAs. Our work concludes with identifying how this character of talk is different from human-to-human talk, how future IPAs could be better tailored to what to may become mundane practice in this setting, and make a number of our conversation analytic findings available to the design of future systems.

BACKGROUND

We now briefly situate this paper with respect to existing literature on mobile devices in collocated interactions, introduce and describe the function of IPAs, and finally provide some related conversation analytic work in order to frame our analytic perspective.

Mobile Devices in Collocated Interactions

Research on collocated interactions has generated numerous examples of specifically designed applications for use in groups, such as photo collage building and sharing [9], ad-hoc brainstorming [25], and even the recording of sports events by the public [12]. Others, instead of creating specific applications for a context or activity, have observed how people consume video on their mobile devices [30], search the internet [3], or use devices as second screens in the living room [37]. It is this last tranche of examples that we wish to draw attention to: technologies designed for a single person are in fact made multi-person through the appropriation by members [24]. Furthermore, the availability of a constant communication channel [34] and the mobile Internet [17], means that devices can and are used in all settings for a large range of tasks [5,45]. We note that the use of mobile devices has not been without complaint, with some claiming

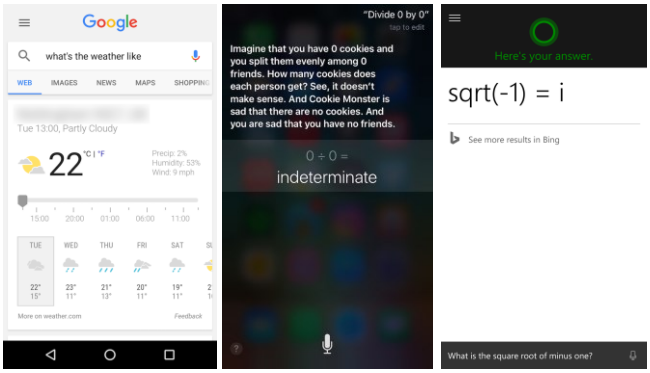


Figure 1. Screenshots of the visual interfaces for Google Now, Siri, and Cortana, taken from their use on smartphones.

technology makes us more distant from each other [47] because of this omnipresence. Others have talked of problematic areas such as attentional orientation, attributing this to the potentially conflicting modalities of interaction between device and conversation [36]. In our work, we believe that a speech interaction may alter this by shifting device interaction into the same modality as conversation, i.e. talk, and we intend to uncover how this interaction will unfold within face-to-face conversation amongst friends.

Intelligent Personal Assistants

In 1960, J. Licklider remarked that "there is a continuing interest in the idea of talking with computing machines" [23]; a quote that is as relevant 57 years later as it was then. For example, work has pursued the ideas of talking machines (i.e. conversational agents) that act as companions for the elderly [48], or virtual museum guides [21]. In this work, however, our concern is with another form of conversational agent, the 'virtual butler', or rather as marketing materials suggest, the 'intelligent personal assistant'. These assistants help people 'get things done' [32] and provide assistance whenever they are called upon for various different tasks. Figure 1 exemplifies three of the most popular commercial IPAs responding to different types of questions. As shown in the dialogue with Siri, responses may contain humour in addition to factualness. Furthermore, in addition to task-oriented questions and commands, some commercially available IPAs also respond to general questions such as "*how are you?*" and "*what's your favourite colour?*", further anthropomorphising the agent.

Early iterations of IPAs were focused around single tasks, such as JUPITER [51] that was capable of providing weather information. The system relied on people making telephone calls to interact with it, with the system engaging in dialogue with the interlocutor by talking back in a conversational manner. As network connectivity and accuracy with automatic speech recognition improved, IPAs, such as InCa [20], were able to operate on portable devices by making use of remote computing power and wireless communication technologies. IPAs are now readily found on many devices such as smartphones, tablets, watches, and televisions. Additionally, although such systems fail to fully mimic human talk, Pelikan and Broth [33] were able to reveal the succinctness of how people *adapt* their talk to an agent's needs and capabilities, making their interactions more successful. Their work focused on a dyadic face-to-face conversation with a humanoid robot, and was able to reveal a number of difficulties individuals face in such talk. In our work, we pivot to considering how this talk unfolds as situated action within multi-party conversation.

A number of pieces of work about IPAs have suggested positive aspects in order to justify their development, such as Jones et al. [18] who describe how a voice-controlled personal assistant could be used to support collaboration amongst those gathered around an interactive smart table. Others such as Luger and Sellen [26], however, paint a more

challenging picture. Through interviews they found that there still exists a "gulf between user expectation and experience" with existing conversational agents. This gulf stems from people's perceptions that such systems should deliver more than they presently do and of issues with communicating system functionality. Innovations to address this gulf include features such as displaying understood text on a screen, voice typing [22] (i.e. live dictation), and the grounding (i.e. affirmation) of spoken input through responses [6,28], although peoples' reported experiences suggests that numerous problems still remain. We believe that by exploring the use of IPAs *in situ*, and by employing a conversation analytic approach, we can provide an understanding that contributes to the design of IPAs.

Multi-party Conversation

Conversation Analysis is an analytic approach related to ethnomethodology [13,39] that concerns itself with the study of everyday social interaction and orients to the sequential and situated action of members [46]. We apply our analytic orientation in order to understand how members structure their interactions with each other in relation to the IPA, and how they accommodate interactions with the IPA. Norman and Thomas [29] remind us that by unpacking this orderly action of members, and revealing their spoken and unspoken action, design in HCI can be informed and tailored by orienting to the interactional sequences employed by members interacting with systems. We also intend to reveal the sequential activity of how members talk with an IPA in multi-party conversation, and through this we can uncover the nuanced interactional accomplishments and problematic interactions that take place.

Multi-party conversation proceeds with much the same organisational practises of dyadic talk: members take *turns* to talk, with each turn consisting of one or more "turn constructional units". A point in talk where the speaker may change is defined as a "transition relevance place" [40]. Furthermore, a number of remarkable and relevant systematic practices do exist for multi-party conversation, such as a preference for answers to questions to be provided by any member as opposed to the selected next speaker simply providing a response [44]. Work also details how the formation of multiple smaller conversations can take place (called "conversational floors" [1,10]), to allow for multiple members to talk at once non-problematically without requiring overlap resolution [41]. During our analysis we were sensitive to this although we do not frame our findings in these terms and instead we let the character of talk emerge as we explicate members' actions.

APPROACH

We now provide a brief description of the setting in which we conducted our observations, details about the participants, and also provide the rationale for our methodological and analytic orientation. The study was approved by the university's School of Computer Science Ethics Committee.

Research Setting

In order to situate our study, we chose a "casual social" setting [36], or "third place" [31], to conduct a number of observations of friends socialising together. This type of space provides a suitable natural environment for us to observe participant behaviours with mobile devices "in the wild" [7] in a similar fashion to that of others [36]. A casual social setting forms an environment in which individuals and groups can socialise with each other, that may be outside of the home or workplace, and that provides a level of comfort and relaxation for those who gather there. In our studies we selected a neighbourhood café that served hot and cold food, cakes, and drinks. The café is in a residential suburb of a city, within a pavilion at a local park and nearby to schools and a university. We arranged suitable times for observations with the café and participants which would allow us to video and audio record the friends talking during a gathering lasting up to ninety minutes. All sessions were recorded on weekday afternoons when the café was open to the public. Video capture was completed by two fixed wide-angle cameras on tripods with an audio recorder placed on the table.

Participants

We recruited groups of friends via email and social media to visit the café together for the purposes of socialising. Prior to the study, participants were asked whether they had previously used a personal assistant on their mobile device, although there was no frequency or expertise required by them in order to take part. We recruited three groups of four friends to go to the café together over a two-month period. Seven participants self-identified as male, and five as female; they aged in range from 22 to 37 (M = 28.75). All went through the process of informed consent and were reimbursed for their time with a shopping voucher each. During the studies, all participants drank various drinks, some ate cake, and one brought some light reading with them to do as they were chatting with their friends.

Methodology

Our study methodology is most aptly described as participant-observer, with a researcher present at the table conversing with the group where relevant. The group of friends met the researcher at the café and were asked to complete a consent form prior to data capture. They were free to move about in the café although primarily sat around a single table as they socialised, drank, and ate cake with each other. For the study, participants were asked to preferably use the personal assistant on their mobile devices instead of typing where possible.

There was no requirement to use a device and there were no tasks set for the friends to perform during the study. We did consider the idea of curating a number of tasks for groups to perform with devices during the sessions, however following a pilot study in which participants were given 'free reign' on what activities to perform during the study, and told to converse as they normally would, we concluded that this was not needed – people still chose to use IPAs. Therefore, we simply asked that they socialise and when the opportunity

arose, they use an IPA. After the study, we asked a number of informal questions to gauge feedback and inform us of personal perspectives on the use devices, however this group interview was used as a debriefing exercise rather than to shape our findings. Thus, our data consisted of recorded video and audio data and some informal interview responses only. To analyse the collected corpus, we employed an ethnomethodological and conversation analytic perspective [14,15,19] as our analytic lens. Through this, we unpack the orderly and situated practice of using IPAs by members. Our analysis required the watching of the collected corpus multiple times, in order to segment and identify relevant fragments of data consisting of IPA use. Fragments were continually watched, with the methodical actions of members within the setting recorded and transcribed.

Our work was oriented to unpacking *the retrospective-prospective character* [13:35-75] of members accomplishing the work of using an IPA in this setting, in and through their ongoing social interaction. This orientation required us to explicate and specifically identify the successful accomplishments that occasioned the use of the IPA. This also included how the device was introduced, the command or query to the IPA formed, the actions (in-talk and body orientation) of members in the setting throughout the activity, and so on. In other words, our orientation, and our analysis, allowed us to gain a comprehensive understanding of the sequential activities performed by members in using the IPA, and this is what we present as our findings.

FINDINGS

In our work, we orient to the sequentiality of using IPAs by considering and observing the naturally accountable character of their use. This reveals to us the nature of how the members' actions were occasioned in and through interaction, and sequentially, what this methodical and situated practice brought about. We provide *vivid exhibits* [2] of the accomplishment of using IPAs to exemplify the orderly practice of members and present a rich picture of how members' use of their mobile devices unfolds. In particular, we intend to reveal (1) how members perform a command or query with their device, and (2) how members orient to and appropriately deal with the query and the IPA's response to the query. This provides the basis for understanding the specific character of talk with an IPA in multi-party conversation and allows us to discuss throughout the findings the nature of 'what it is' to talk to an IPA.

Our data is presented as a series of fragments of talk, mostly given using the transcription notation by Heath et al. [15], with a number of specific modifications including the addition of non-verbal actions performed by members given in double parentheses. Summarily: we note where talk is `LOUD` or `°quiet°`, paused between words `(.)` or utterances `(0.4)`, where a member **`talked to an IPA`**, and where sounds are `elong:::ated`. Additionally, we show where two members' talk is overlapping by using square brackets (`[ ]`) and indentation, and `((actions))` are given within double parentheses. Names of members, which are altered along with other identifiable information, are given by their initial letter within fragments and the researcher is identified as `R`. Any utterance or action by a member's device/IPA is given separately on the right hand side of the transcript. The transcripts presented here are redacted or simplified in some places for brevity.

We have also provided some counting throughout our findings in order to furnish readers with an understanding of the commonality for which we saw different aspects of IPA use, however we stress that our findings should be considered *entirely* qualitative and that we do not seek to make quantitative judgements. The numbers provided in this paper allow readers to understand the shape of corpus and interpret the qualitative findings only. In totality, our corpus consists of 123 utterances to conversational agents by members, across 40 distinct episodes of data from a corpus consisting of 3.6 hours of video data. We now present data from three distinct fragments, each with the following corresponding findings: the first fragment introduces the impetus of repeated queries and of the silence that may be produced following query performance; the second fragment demonstrates the importance of visual contact within the interaction and of accountability of talking with an IPA; and the third reveals the collaboration that may occur with refinement of a query.

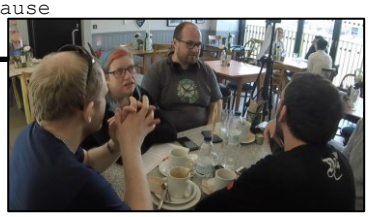
Fragment 1: *"Do Animals Have Accents?"*

To begin with, we present the transcript given in Fragment 1 that describes an interaction amongst the four friends Lilly, Gary, Karl, Antonius, and the researcher. This transcript furnishes us with numerous noteworthy observations which we will first draw out before further unpacking them in order to identify how members perform and orient to the utterance of a query. The group, which consists of members from the UK, Romania, and Austria, have been discussing the different onomatopoeic sounds that various animals make and how these sounds vary by country and language.

There are presently two conversation floors taking place: in the floor we focus on, Karl asks Lilly about animal accents before recounting scenes from a television show to Lilly (omitted from the transcript in lines 07–20), and in the other floor Antonius is recalling his recollection of the sounds different animals make when uttered in Austrian. Just before Karl begins to recount his story, Lilly picks up her smartphone (line 05) and begins to type with the onscreen keyboard throughout the story. After the story, both Lilly and Karl laugh and then Karl orients to and engages with the other floor. At this point, Lilly moves her smartphone closer to her mouth and asks her IPA "`do animals have accents?`" (line 24). This question was not specifically asked in talk, but arises as a result of the topic that all the members have focused on in both floors at some point.

In the fragment, following Lilly's query, we see a short pause (line 25) before Gary shifts his gaze to Lilly and responds to

		Members	IPAs
01	K:	do cats acth- (0.5) can you work out whether it's French because	
02		because its talking in a- doing a French cat impression	
03	L:	I::::: think some animals you can	
04		(1.9)	
05	L:	((picks up phone from table and taps on screen))	
06		(1.4)	
		...	
21		(4.0)	
22	L:	er:::m: ((holding phone in front of her at chest level))	
23		(3.7)	<init by button press>
24	L:	((moves phone in front of face)) **do animals have accents?**	
25		(2.1)	
26	G:	((shifts gaze to L)) yes they do actually! I think I've read something	
27	L:	I think I have [too↓]	
28	K:	[you missed mine- my racist joke	
29	G:	[yeas! cows! I- I read about cows that they have different	
30		accents around the world	
31	K:	cars?	
32	G:	[cows]	
33	L:	[°cows°]	
34	K:	I thought you said cars	
35	G:	calves maybe as well (.) who knows?=	
36	L:	=**DO: ANIMALS HAVE ACCENTS!**	<init by button press>
37		(2.4)	
38	L:	°rubbish°=	
39	K:	=parrots presumably do=	
40	L:	=can you ask it? ((holds phone out in front of K's face))	
41	R:	((retrieves phone out of pocket))	
42	K:	**DO: ANIMALS HAVE ACCENTS!**	<init by button press>
43		(0.9)	
44	L:	no:!	
45	R:		(audible) Sorry I'm-
46	R:	((R touches screen to stop utterance))	<init by button press>
47	R:	**do animals have accents?**	
48	R:		(audible) Ok I've
49	L:	**do: animals have accents?**	found this on the web
50	R:	(sigh)	
51	G:	do [they?]	
52	L:	[Ah (.)]it's working now!	
53	R:	((touches top search result on device screen))	



Fragment 1. Discussion about animal sounds and variances across different countries and languages.

her question, as shown in the second photo within the fragment, even though her question was aimed at her IPA. Lilly more abruptly re-utters her query a short while later (line 36); another pause in talk then occurs (line 37). A few seconds later her quiet utterance of "rubbish" (line 38) suggests failure of the device to perform as expected, and this is confirmed momentarily later as she passes the responsibility of uttering the query by holding the device in front of Karl and questioning whether he could "ask it" (line 40). At this moment, the researcher retrieves his phone from his pocket to perform the query (line 41). The final attempt by Lilly (line 49) yields search results, as does the researcher's. Following the fragment both begin to share information retrieved from webpages to the other members of the group. This fragment is an indicative of what IPA use looks like – that is, there are a number of grossly observable features that take place: there is selection of speakers, repetition of queries, pauses in talk, body co-orientation and so on. We will now explicate these distinct actions in order to understand the practice that unfolds.

Repetition of Queries

Our first consideration is to characterise the practice of how a member talks in turn with an IPA. In the fragment Lilly uses her smartphone on multiple instances to perform the query "do animals have accents?" (lines 24, 36, 49), each time with more impetus in her voice. Failing to get a satisfactory response, she selects Karl to talk with an IPA and perform the same query (line 40): "can you ask it?". The researcher also self-selects to perform the same query (line 47). With each repeated query, Lilly accounts for the device's failure to appropriately respond to her initial query: either the device has misheard, or it not heard at all and so another attempt is required to complete the task at hand. Thus, we note that members address a problematic interaction with an IPA through the further production of talk: they repeat their query. Specifically, we state that members repeat queries if a query 'goes wrong'; this may seem like an obvious fact but one we feel is worth stressing. In our observations we counted 31 queries (25%) that were identical in lexical terms to a prior query, although lexicality

is only half of the story. Consider in the fragment where Lilly repeats her query multiple times (lines 24, 36, 49). Although identical in language, the production of talk differs in each one: in the first she uses a general conversational tone; her utterance is consistent with the ongoing conversation. With her second performance, however, she performs the query louder and emphasises key sounds; to members within the setting she demonstrates her frustration with the device – its failure to interpret her words requires her to try again. Therefore, our analysis reveals that the failure of the device's IPA to adequately respond to query occasions the necessity to repeat the query, possibly with greater impetus.

Mutual Production of Silence

Once a query is performed with an IPA, a number of practical actions are undertaken by members as they accommodate the utterance within conversation. Talking to an IPA is naturally accountable in addition to being occasioned in and through the social interaction of members in the setting. The accountability of action is premised on the fact that members of a setting can *observe* and *report* the action [13], and this is feature of talk-in-interaction. Thus, talking to an IPA immediately makes audible what is being undertaken to all within earshot. A member's device interaction is made directly accountable through talk, unlike interactions on touch screens where the device user may have to provide explicit accounts to make the action accountable [36]. This fragment provides interesting markers to consider what specifically follows talking to an IPA; in particular, this transcript reveals that members' talk to IPAs may be sequentially followed by pauses in talk (lines 25, 37, 43), perhaps suggesting anticipation of an answer from the IPA.

Our data shows that routinely, as a practice, talking to an IPA in turn occasions the *mutual production of silence* by the co-present members as they re-orient to the accountable use of the IPA, and in turn focus on the device or the interlocutor. They do not pause their interaction or 'sit in silence' however, their embodied actions of gaze and body co-orientation furnish others with how they are focusing their attention, as they turn to device interaction. In effect, performing a query brings about a *lapse* [16] in the conversation: neither the member who was performing the query selects to talk next, nor does any other member. IPAs function by assuming a pause in talk specifies the completion of a query, thus a pause by the interlocutor is necessary. However, as other members await a result, they themselves do not self-select in commencing a turn. Therefore, we note that the activity of performing a query with an IPA may prescribe a lapse in talk and the mutual production of silence.

Accountability of the Device Interaction

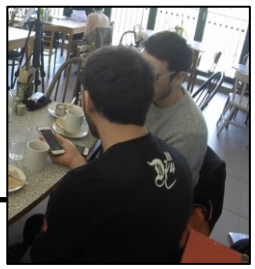
The fragment reveals that as Lilly performs her utterance to her IPA she in turn proffers a conversational topic to the floor (line 24). Her query is audible and accountable to all members within the multi-party conversation and is one to which any member can attend to. Her actions were to select her IPA to respond, but any member, as with multi-party conversation, can intervene and respond if they so choose to do so. The preference in multi-party conversation is for the member who was asked a question to provide an answer, but there also exists a second-order organisation for an answer to be provided by any member over the selected speaker to support the progressivity of talk [44]. This organisational practice is present in our fragment as Gary answers her question with "`yes they do actually`" (line 26), choosing to provide an answer rather than wait for a response.

Finally, Gary's actions reveal to us that members not only orient to an IPA or device but that members may orient and respond accordingly to the query performed. Moreover, although member's talk to an IPA is accountable within multi-party conversation, an IPA may be a muted conversational partner. This is because whether it makes sounds or not is dependent upon both the manufacturer and the owner of the device (and their configuration of the device). In this fragment, for example, Lilly's smartphone does not make an audible response to her queries, although the researcher's device does make sounds. Remarkably, however, the accountability of an IPA's response is not wholly restricted to IPAs that make audible responses or devices which are positioned so as to be visible to co-present others. Instead, how the interlocutor accountably attends to the performance of the query demonstrably provides a (limited) account to other members of the IPA's response. This is exemplified in Lily's repetitions of her query, occasioned by the failure of the IPA to respond in the desired manner. Furthermore, this fragment reveals how members also react to an IPA's performance, which as we see as Lilly purports the notion of failure by muttering "`rubbish`" following her second attempt, and "`no`" following her third. These utterances are not necessarily directed at any party, the group, or the device, but they make available to co-present others the failure of the device to meet her expectations. Therefore, we note that although an IPA may not audibly make its actions available to the setting, members themselves naturally account for the performance of the IPA in and through talk, either by repeating their queries, or through commenting on the device's failure with rhetoric. Thus in the case of a repeated query, the member makes the device's failure to respond accordingly observable-reportable.

Fragment 2: "*My Mother is Mama*"

We now consider the short sequence in Fragment 2, which is from the same session. In this exhibit Gary asks his IPA to call his mother, who is listed under the name of *mama* in his smartphone's address book. The conversation takes place in a separate floor consisting of just Gary and the researcher, who are both sitting next to each other. The other members of the setting are conversing while the two discuss Gary's interactions with his device. Gary ponders, by asking the researcher, whether if he asks his device to call his mother, the device will recognise the name in his contact list (the contact's name is spelt 'mama' in Romanian); we join the action as he attempts to accomplish this.

		Members	IPAs
01	R:	'cos you can also	
02		tell people who they-	
03		like you can say like	
04	G:	**hey Siri=**	
05	R:	=my mother is this	
06		person (0.8)	
07		**hey Siri**	
08		(1.0)	
09	R:	I'd press the button	
10		(1.2)	
11	G:	**hey Siri**	
12		(2.4)	
13	G:	**call my mother**	
14		((both G and R	
15		watch screen))	
16		(5.9)	
17	G:		(screen) what is
18	R:	((points to screen))	your mother's name?
19		yeah but then	
20		(0.9)	
21	G:	**my mother is mama**	
22	G:		(screen) I can't
23			find anyone called
24			mamma



Fragment 2. Short fragment of a member responding to device's request for further information.

The fragment starts just after Gary picks up his phone up from the table and returns his gaze to the researcher (line 01). Without shifting his gaze, Gary lifts his phone and says "hey Siri" (line 04) and then moves the device back to chest height between him and the researcher. After a second, he glances down at his device; his smartphone's screen remains off and so he lifts his device again and re-utters "hey Siri" (line 07). He holds his phone in a position that the researcher can see, although this time his gaze remains on the device awaiting a result. The researcher offers implicit advice based on his personal experience (line 09), although a moment later Gary (successfully) retries "hey Siri" (line 11) and then asks his IPA to call his mother (line 13). He then holds the device between the two of them again, as can be seen in the image within Fragment 2. After nearly six seconds of both partners watching the screen between them, the device seeks further information of the name of his mother in his address book — Gary provides this (line 21) although this fails as the device searches for contacts named "mamma" and does seemingly does not look for synonyms such as "mama". The use of Siri is abandoned shortly thereafter.

Multimodality of Feedback

In this fragment, Gary retrieves his device from the table, which in retrospect we see as an opening to his use of the IPA. He then lifts the device to his mouth, but keeps the screen facing him, such that the bottom of the device is closest to his lips. His accountable performances of "hey Siri" (line 04) reveals his reasoning about the functionality of the device, of where the microphone is situated, and the ability of the device to 'hear' one voice in a 'sea' of many. He then holds the smartphone between him and the researcher, accordingly *sustaining* his device use [36] and attending to the norms of social practice: he does not isolate himself or avoid interaction with the researcher, with whom he is talking. Additionally, he continues to use gaze and body co-orientation, and moreover, he makes visible his device screen, embedding the device and his device interaction within their conversation. Members may make use of an IPA's magic phrase, as Gary does in this fragment, although we hasten to note that of the 40 extended episodes in our corpus, only 12 featured the use of a phrase to trigger an IPA. Consider the sequence of Gary's performance of this magic phrase: we see repeated pauses after his utterances to the IPA (lines 06, 08, 12, 16) as Gary provides the utterance and waits for the device to respond by looking at the screen. Members typically pause following the completion of a magic phrase until the device provides a visual acknowledgement that it is 'listening' (in only one instance did a member immediately follow the phrase with their query). Our findings show that, far from shifting the modality of the interaction *from* visual and touch *to* speech, members still rely on the visual feedback from devices through glances at the screen *in addition* to speech as a direct consequence of the design decisions made with the IPAs.

Body Co-orientation

In this fragment we also uncover that members make the IPA's actions available to others through body co-orientation and the positioning of the device. The image within Fragment 2 shows Gary making his IPA's (re)actions visible to the researcher by holding his mobile phone in such a position that both parties in the conversation can orient to. This practice is employed by members as they make accountable the IPA's response through different methodical actions. This practice turns upon the pertinence of visibility to- and practicality of- their situated action. Summarily, Gary does not audibly report the failure of the device, he does this through his repetitions and sharing of the screen. Therefore, in tying these findings with those from Fragment 1 together, we note that although an IPA's response may not necessarily be accountable to the members of the multi-party setting, their conversational counterpart may offer this account through their own actions by making the device visible, by accountably responding to the device, or through the member's production of rhetorical talk.

Fragment 3: "*When does the sun go down?*"

We now move on to Fragment 3, consisting of four friends: Arthur, Harry, Sally, Julia, and the researcher. The friends are meeting late afternoon during winter and the sun is shining on to Harry's eyes. He holds his hands in front of his eyes although refuses to move because he will "...be fine in like three minutes" (line 01). The members joke about this experience, and that this forms part of their study (lines 08–12). A lull happens in talk for a second and then Julia begins to remove the cover from her iPad, which she has on the table (line 12); she waits for the group laughter to die down, presses the home button (line 14), and begins her utterance as Harry finishes remarking that the sun has now moved (line 15). At this point, all members lean in towards

		Members	IPAs
01	H:	i'll be fine in like three minutes ((holds hands in front of eyes))	
02	R:	keeps coming back as well like-	
03	S:	as soon as you change it comes back	
03	J:	yeah yeaha	
04		(0.3)	
05	R:	there's actually just someone out there with a light!	
06	All:	((laugh))	
07	R:	every time you-	
08	S:	[this like]	
09	A:	[it's all part of the study] this is what we're really being	
10	S:	this like (0.5) deception	
11		(1.1)	
12	R:	how can we blind someone subtly!	<J removes cover from
13	S:	((laughs))	device but leaves open>
14	J:		<init by button press>
15	H:	there we go!	
16	J:	**what's the time of sunset?**	
17		(1.3)	
18	All:	((gaze at the tablet))	(device stops listening)
19		(3.0)	
20	J:	ok!	(device displays clock)
21	A:	((leans in to look))	
22	S:	that's [a] fucking analogue clock it pisses me off!	
23	H:	[today?]	
24	H:	ilunno (0.6) 24 hour=	
25	J:	<no no no!> it misunderstood actually (0.8) understood what's the	
26		[time]	
27	H:	[time] now	
28	J:	so-	
29	A:	soaoah yeah↑	
30	J	shall I ask (1.6) um:=	
31	H:	=what time will the [sun set?]	
32	J:	[((holds down button))]	(audible chime)
33		(4.0)	
34	J:		(screen) go ahead I'm
35		(0.3)	listening...
36	J:	**when does the sun go down?**	
37	J:		(audible chime)

Fragment 3. Members apply their reason to formulate a query that will result in success.

the device, as shown in the image, and wait for the result. After a few moments, the IPA returns the time for the local area as an analogue clock. A number of comments on this are passed: Sally comments on the presentation of the time (line 22) and Harry questions if that is for the present day (line 23). Julia then interrupts the talk and retorts that she has realised the device has "misunderstood actually" (line 25) and that the IPA is presenting the current time, not the time of sunset.

Refining a Query

In unpacking this fragment, it is revealed how members practically reason about how an IPA responds to a query and attend to the IPA's response. In this exhibit, Julia realises the misunderstanding on the IPA's behalf, and makes this accountable to all (line 25). In doing so, she provides an explanation for the problem source – or rather, starts to – as she realises it "understood what's the-" (line 25) and Harry, who seemed to question the answer (line 23) completes her sentence with "time now" (line 27). Through the ongoing interaction, members collaboratively reason that the response was not as expected and that this must be because the interpretation of the query by the IPA was wrong – which finds agreement (line 29) and leads to a proposal to ask a different question (line 31). In turn, the members collaboratively find words to return a successful result. In this, Harry proposes a slightly different question (line 31) although ultimately Julia asks "when does the sun go down?" (line 36), to which the IPA provides an accepted answer. We can see the how members *refine* the query by applying practical reasoning to the IPA's response by reformulating and refining the query. In the fragment, Julia interprets the result from the IPA as incorrect (line 25), but then reasons about the response, and then asks the IPA the same question with a different lexical construction (line 36). In this, she does not just retry or repeat the *same* query however, she in fact *refines* it to solicit a successful answer. Refinement can be seen as a subset of repeating, where a member may still seek to identify the same information but with a new query in order to retrieve a satisfactory result. As with repetition, this too was also a common practice by members; a total of 22 queries (out of 123) were posed to IPAs where lexically they were different, but the purpose remained the same. While in this case the original

interlocutor refined the query, on other occasions other members may also have performed a refined version of the original query on their own device. We posit a distinction in the occasioning of refinement and repetition. Refinement occurs as members attend to an IPA *misunderstanding* their query, e.g. as Julia informed us (line 25). Repetitions, on the other hand, are performed in response to members perceiving the IPA to have *misheard* the query (e.g. members speak slower, louder or more accentuated, but with the same word construction). Typically, each episode of interaction with IPAs was eventually successful and in summary we would say that, if at first a member did not succeed, as in the old adage, they tried, tried, and tried again.

Collaborative Device Interaction

Finally, in this fragment we also see a cooperative, or even collaborative, orientation to the device in use. The members collectively reorganise their body orientation around the device interaction, they pause their talk, they gaze at the tablet, and they attend to the answer as soon as it is provided — i.e. they work together in a team-like manner to complete the query. In this entire sequence, the query is accountably occasioned in and through the conversation about sunshine. The other members then witness the query being performed (line 16), and the failure of the device to respond appropriately is made accountable by making the screen visible to all members. This in turn allows for members to collaboratively reason about the grounds of failure (lines 25–29). In attending to the failure, the members then construct a further query which leads to a satisfactory result. Given the naturally accountable practice of performing a query with an IPA through speech, it appears that the practice of refining a query lends itself to supporting a collaborative activity for the copresent members.

MACHINERY OF INTERACTION

We now move from discussing our findings in terms of fragments of particular methodical accomplishment, and instead reveal the resulting "matter of interactions as products of a machinery" [38].

Performing a query is done by **selecting the interlocutor to perform the query from the members in the setting** through the procedurally organised practice of self-selection, as occurs when Lilly chooses to ask her device whether animals have accents (Fragment 1) or when Julia self-selects in order to determine the time of sunset (Fragment 3), for example. Alternatively, selecting may be done though interaction with one member selecting another to perform the query in and through talk (e.g. Fragment 1). Once a member is selected, the member begins by **retrieving the device and opening talk with the IPA**. This is accomplished by using a magic phrase to enable the IPA (e.g. Fragment 2), or pressing the digital (e.g. Fragment 1) or physical button (e.g. Fragment 3) on the device. The member then undertakes the actions of **(re-)formulating and uttering the query towards the device's microphone**, with the query typically consisting of a series of keywords, a command (e.g. Fragment 2), or a question (e.g. Fragment 3) formed individually (e.g. Fragment 2) or collaboratively by members through talk (e.g. Fragment 3).

Responding to the query performance occurs by **mutually producing silence** in the setting as members orient to the device, the interlocutor, or the query (e.g. Fragment 2), or by **continuing conversation amongst the other members** in accordance with standard multi-party conversational practice (e.g. Fragment 1). Members undertake the routine of accounting for the IPA by **sharing visibility of the device** (e.g. by positioning the device between them as in Fragment 2) or by **explaining or rhetorically responding to the IPA's response** (e.g. exclaiming at the IPA's failure to hear the utterance in Fragment 1). Interlocutors attend to failures by **refining queries in situations where the IPA has misunderstood** (e.g. in Fragment 3 when the IPA has not understood the question posed and returns an 'incorrect' answer) or by **repeating queries if the IPA has mis–or not–heard** (e.g. as occurs in Fragment 1 when the device does not hear the question posed).

DISCUSSION

We now discuss our findings and the uncovered machinery both in terms of the existing literature and what our findings mean for design. Our work examined how a highly promoted and recently popularised interaction paradigm actually unfolds in everyday interaction. We chose a setting that is common for people to socialise, relax, and use their mobile devices as part of their everyday routine. In this sense, our study was about exploring the use of the technology in a 'real-world' (i.e. non-laboratory) setting that we knew would be challenging for IPAs. Yet, studying how interactional and technological problems are accommodated in and through interaction can provide us with insights for design.

Repeating and Refining

Our data is replete with exchanges in which repetitions or refinements are a problematic source within interaction that members routinely attend to (53 out of 123). Our work shows how members individually or collaboratively inspect and interpret the on-screen output of the IPA in order to understand the failure to complete a query. In the case of failures, members *repeat (*31 out of 123), or in some cases, *refine* their query (22 out of 123), but very few times do they abandon the query. Repetitions and refinements happened in close succession, usually within a few seconds. Regarding the question how design might respond to this finding, the most obvious solution that industry probably is already working on is to explore more meaningful feedback provided by the IPA. This could help the interlocutor 'to find the right words', for example, by providing the grounds upon which the query failed, or by suggesting how to refine the query. Design inspiration might also be drawn from auto-completion features such as *Google Instant* in order to support query formulation and refinement without the need for members to *recall* or reason about terms which would be more likely to result in a successful query.

Supporting conversational repair is important, and future systems must also consider the operative language used in verbal correction. For example, as a human acknowledges that an IPA has misunderstood a word, or that they themselves have misspoken, they may say "*oh no, I meant…*". We believe that IPAs could listen for spoken repair phrases to proactively trigger a repair sequence, in addition to the interlocutor's use of repeated or refined queries. This would reduce the effort for a human interlocutor by no longer necessitating a restart of the dialogic interaction with the device. Additionally, our findings reveal a difficulty for IPAs to understand synonyms and homonyms in talk. We concede that it would be unrealistic to expect IPAs to demonstrate a perfect understanding at all times; humans are unable to achieve this themselves. However, through repair in talk we are able to identify any misunderstandings and accordingly correct them [42]. Yet IPAs presently provide limited functionality for this; if a device has not understood a phrase, it could ask people "could you ask your query using different words?" or, perhaps when a word is not recognised, "could you spell that?", alleviating some of the identified problems. Such an implementation could serve as a learning opportunity for software. This would also provide a naturally accountable response from the IPA that would also support multi-party conversational practices, as identified in our work.

More complex speech recognition approaches have also been taken in the literature, such as an idea explored by McMillan et al. [27] to improve the relevancy and performance of IPAs. In their work they use the continuous speech stream to inform and enhance IPAs such that when they are called upon, they will have collected contextually relevant information. Extending this approach, we would also suggest that this contextual relevance could be gathered from prior failed queries, as a utility to both improve accuracy in understand interlocutor's intent during successive queries, albeit at the potential expense of privacy. This could also be used to improve performance of IPAs through learning various contextually relevant meanings of queries.

IPAs as Humanlike Conversational Partners

IPAs are generally anthropomorphised, given names (e.g. Siri), and endowed with humanistic interactional traits such as humour. However, their ability to support conversation is limited, they generally operate through turns-at-talk by repeatedly cycling through the simplest unexpanded units of conversation: adjacency pairings. In some instances, these become expanded sequences through insert expansions as the IPA engages in the routine of "other-initiated repair" [42] to seek further information from interlocutors. This is something that is a standard occurrence in human-to-human talk in order to repair mishearing or misinterpreting. Therefore, our analysis was able to explicate the humanlike orientation to conversational practice that IPAs possess. We also saw members routinely ask questions (42 out of 123 queries to IPAs) and give commands (26 out of 123) to IPAs, suggesting that there is a perception by members to treat them as humanlike, although Luger and Sellen [26] found that this was typically when in private and that in public settings people preferred the use of keywords. Pelikan and Broth [33] also found that people engage in *recipient design* [40] when conversing with an artificial conversational partner as they do with human partners. We feel our findings corroborate this as we identified that members routinely reason about a response from an IPA and attend to, either individually or collaboratively, reformulation of their query. Our findings highlighted the *mutual production of silence* in talk with IPAs; these were periods of silence that become occasioned as multiple members orient to an IPA or mobile device after a query is performed. This activity saw members systematically 'pause' talk (but remain interactionally active through non-verbal means) as they accommodate the IPA's untimely response in talk, similar to the way people may orient to a question in a dinner party, for example.

Therefore, our data reveals how the sequence of talk has some characteristics of conversation, and that talk with IPAs has the hallmarks of everyday talk between people. Yet we must also remark on the actual performance of utterances to IPAs by members and how this is distinctly different to how one would talk to another human, even if it consists of the same lexical construction. To illustrate this, we recall Fragment 1 with the query "*Do animals have accents?*", in which this question was posed repeatedly to an IPA. In this example, Lilly asks the question calmly at first, she raises her voice and employs more impetus a second time, she then asks another member to "ask it", and finally she succeeds on her fourth attempt. Imagine, if you will, this sequence of actions unfolding with a human counterpart instead of the IPA: the instinctive and common-sense response would probably be that talking to someone by raising one's voice, and asking another to "ask it*",* and by another member repeating the question would be considered rude. Indeed, the failure of the IPA to adequately respond to the member could also be considered rude and inattentive to the conversation. This development of events uncovers how talking to an IPA is reminiscent of conversation but that the production of talk to an IPA is fundamentally different because the recipient is not a human. Our data reveals that members may refer to an IPA as an "it" irrespective of the IPA's spoken voice being imbued with gender and this fundamentally reveals that through the veneer of humanlike interaction, members still treat an IPA as an *agent*, or a machine, and not human.

Whether work should be done to make machines talk more like a human is contentious, with some arguing that the *unformalisability* of conversation suggests efforts to create a true humanlike conversational partner are futile [4]. However, our purpose here has not been to discuss whether a machine could transcend from human*like* to human-realistic talk. Instead, we intended to reveal the nature of talk with existing IPAs, and to highlight nuanced interactional troubles that could be addressed in design.

IPA Use in Multi-Party Conversation

Our final discussion point is of how IPA use in multi-party conversation unfolds and the contributions of interacting through voice in a multi-party, face-to-face conversation. An expectation that we had going into this study was that using speech would alleviate members of the necessary accounting practices found with interaction on touch screens, such as by explaining what a device was used for, or by sharing the screen [36]. Contravening this assumption, our findings actually show that members still provided verbal accounts for device use, particularly as they attend to failures of the IPAs. Furthermore, members still shared their device's screens with each other – in part because the IPAs studied rely on a touchscreen for interaction. Our observance of members' interactions with IPAs in a casual social setting also drew out the technical limitations of the devices, such as difficulty in the device 'hearing' what was spoken to it in a bustling multi-party public setting, however, we are sure as technology improves these limitations will be eased.

We also found that speaking to an IPA intrinsically makes available the device interaction to all members in the setting, thus providing an opportunity for any member to engage with the device interaction and the interlocutor. This natural accountability has the effect of *democratising* the device use by allowing any member to engage without invitation, and to intervene or collaborate with the unfolding device interaction. Moreover, talking to an IPA provides a mechanism through which all within the setting can interpret and reason about not only about the actions of the member who performed the query, but also to reason about the query. In turn, each member can display practical reasoning in situations where a query failed, essentially transforming a single-person interaction with a mobile device into collaborative multi-person interaction. Existing research on mobile device use in collocated settings has long explored ways of supporting collaboration (e.g. [25]), and our findings suggest that a speech-based dialogue interface could be a viable contender for this practice.

CONCLUSION

In conclusion, our work was oriented to the sequentiality of using intelligent personal assistants (IPAs) in multi-party conversation in order to reveal the character of their use. In doing so, we revealed the methodical, but interactionally problematic, features of interacting with IPAs in the setting. This included the repeating and refining of queries to IPAs, and the mutual production of silence by members. We also uncovered how members routinely organise their queries to IPAs by individually or collaboratively formulating their queries. Our worked showed how the performance of queries and verbal responses to the IPA were naturally accountable, but that members still relied on familiar practices such as sharing screens or providing verbal accounts to members within the setting.

In examining IPA use in a social setting, we found that the natural accountability of voice interaction provides a collaborative mechanism for any member to orient to and engage with the device interaction. Our work showed that while IPAs exude humanistic traits in talk, members' production of talk, and indeed the character of talk that unfolded, seemed at odds with their talk to other human interlocutors. Finally, we discussed a number of actionable ideas for design such as supporting 'word finding' for refining queries, IPAs learning from repeated queries, and of using speech input for collaborative collocated interactions.

ACKNOWLEDGEMENTS

We would like to thank Homemade, Nottingham, for generously providing the venue and supporting our work. Martin Porcheron and Sarah Sharples are supported by EPSRC grants EP/G037574/1 and EP/G065802/1; Joel E. Fischer is supported by EPSRC grant EP/N014243/1.

Data access statement: transcripts of the fragments used in this paper are available at https://doi.org/10.17639/nott.61.

REFERENCES

1. Paul M Aoki, Margaret H Szymanski, Luke D. Plurkowski, James D Thornton, Allison Woodruff, and Weilie Yi. 2006. Where's the "Party" in "Multi-Party"? Analyzing the Structure of Small-Group Sociable Talk. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (CSCW '06), 393. https://doi.org/10.1145/1180875.1180934
2. Liam Bannon, John Bowers, Peter Carstensen, John A Hughes, Kari Kuutii, James Pycock, Tom Rodden, Kjeld Schmidt, Dan Shapiro, Wes Sharrock, and Stephen Viller. 1993. Informing CSCW System Requirements. In *COMIC Deliverable 2.1*.
3. Barry Brown, Moira McGregor, and Donald McMillan. 2015. Searchable Objects: Search in Everyday Conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 508–517. https://doi.org/10.1145/2675133.2675206
4. Graham Button, Jeff Coulter, John R E Lee, and Wes Sharrock. 1995. *Computers, Minds and Conduct*. Polity Press.
5. Karen Church, Antony Cousin, and Nuria Oliver. 2012. I Wanted to Settle a Bet! - Understanding Why and How People Use Mobile Search in Social Settings. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '12), 393. https://doi.org/10.1145/2371574.2371635
6. Susan E. Clark, Herbert H.; Brennan, Herbert H. Clark, and Susan E. Brennan. 1991. Grounding in Communication. *Perspectives on Socially Shared Cognition*: 127–149. https://doi.org/10.1037/10096-006
7. Andy Crabtree, Steve Benford, Chris Greenhalgh, Paul Tennent, Matthew Chalmers, and Barry Brown. 2006.

Supporting Ethnographic Studies of Ubiquitous Computing in the Wild. In *Proceedings of the 6th ACM Conference on Designing Interactive Systems* (DIS '06), 60. https://doi.org/10.1145/1142405.1142417

8. Brian Lystgaard Due. 2015. Challenges with Google Glass in Social Interaction. In *Proceedings of the 4th Participatory Innovation Conference 2015*, 440–448.
9. Abigail Durrant, Duncan Rowland, David S. Kirk, Steve Benford, Joel E. Fischer, and Derek McAuley. 2011. Automics: Souvenir Generating Photoware for Theme Parks. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (CHI '11), 1767. https://doi.org/10.1145/1978942.1979199
10. Carole Edelsky. 1981. Who's got the floor? *Language in Society* 10, 3: 383–421. https://doi.org/10.1017/S004740450000885X
11. Hasan Shahid Ferdous, Bernd Ploderer, Hilary Davis, Frank Vetere, Kenton O'Hara, Geremy Farr-Wharton, and Rob Comber. 2016. TableTalk: Integrating Personal Devices and Content for Commensal Experiences at the Family Dinner Table Hasan. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '16), 132–143. https://doi.org/10.1145/2971648.2971715
12. Martin D. Flintham, Raphael Velt, Max L. Wilson, Edward J. Anstead, Steve Benford, Anthony Brown, Timothy Pearce, Dominic Price, and James Sprinks. 2015. Run Spot Run: Capturing and Tagging Footage of a Race by Crowds of Spectators. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 747–756. https://doi.org/10.1145/2702123.2702463
13. Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall.
14. Charles Goodwin and John Heritage. 1990. Conversation Analysis. *Annual Review of Anthropology* 19, 1: 283–307. https://doi.org/10.1146/annurev.an.19.100190.001435
15. Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. SAGE.
16. Elliott M. Hoey. 2015. Lapses: How People Arrive at, and Deal With, Discontinuities in Talk. *Research on Language and Social Interaction* 48, 4: 430–453. https://doi.org/10.1080/08351813.2015.1090116
17. Lee Humphreys, Thilo Von Pape, and Veronika Karnowski. 2013. Evolving Mobile Media: Uses and Conceptualizations of the Mobile Internet. *Journal of Computer-Mediated Communication* 18, 4: 491–507. https://doi.org/10.1111/jcc4.12019
18. Alistair Jones, Jean-Paul A. Barthes, Claude Moulin, and Dominique Lenne. 2014. A rich multi-agent architecture for collaborative multi-touch multi-user devices. In *2014 IEEE International Conference on Systems, Man, and Cybernetics* (SMC '14), 1107–1112. https://doi.org/10.1109/SMC.2014.6974062
19. Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences* 4, 1: 39–103. https://doi.org/10.1207/s15327809jls0401_2
20. Mohammed Waleed Kadous and Claude Sammut. 2004. InCa: A Mobile Conversational Agent. *PRICAI 2004: Trends in Artificial Intelligence* 3157: 644–653. https://doi.org/10.1007/978-3-540-28633-2_68
21. 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 *Lecture Notes in Computer Science*, 329–343. https://doi.org/10.1007/11550617_28
22. Anuj Kumar, Tim Paek, and Bongshin Lee. 2012. Voice Typing: A New Speech Interaction Model for Dictation on Touchscreen Devices. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems* (CHI '12), 2277. https://doi.org/10.1145/2207676.2208386
23. J. C. R. Licklider. 1960. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics* HFE-1, 1: 4–11. https://doi.org/10.1109/THFE2.1960.4503259
24. Andrés Lucero, Matt Jones, Tero Jokela, and Simon Robinson. 2013. Mobile Collocated Interactions: Taking an Offline Break Together. *interactions* 20, 2: 26. https://doi.org/10.1145/2427076.2427083
25. Andrés Lucero, Jaakko Keränen, and Hannu Korhonen. 2010. Collaborative Use of Mobile Phones for Brainstorming. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '10), 337. https://doi.org/10.1145/1851600.1851659
26. 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/10.1145/2858036.2858288
27. 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/10.1145/2702123.2702532
28. Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface*. Springer

International Publishing. https://doi.org/10.1007/978-3-319-32967-3

29. Michael Norman and Peter Thomas. 1990. The Very Idea: Informing HCI Design from Conversation Analysis. In *Computers and Conversation* (1st Editio), Paul Luff, David Frohlich and Nigel Gilber (eds.). Academic Press, 51–65.
30. Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. 2007. Consuming Video on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 857. https://doi.org/10.1145/1240624.1240754
31. Ray Oldenburg. 1989. *The Great Good Place: Cafes, Coffee Shops, Community Centers, General Stores, Bars, Hangouts, and How They Get You Through the Day*. De Capo Press.
32. Sabine Payr. 2013. Virtual butlers and real people: styles and practices in long-term use of a companion. In *Your Virtual Butler*, Robert Trappl (ed.). Springer-Verlag Berlin, Heidelberg, 134–178. https://doi.org/10.1007/978-3-642-37346-6_11
33. Hannah R.M. Pelikan and Mathias Broth. 2016. Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 4921–4932. https://doi.org/10.1145/2858036.2858478
34. Oscar Peters and Somaya ben Allouch. 2005. Always connected: a longitudinal field study of mobile communication. *Telematics and Informatics* 22, 3: 239–256. https://doi.org/10.1016/j.tele.2004.11.002
35. Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5456–5469. https://doi.org/10.1145/2858036.2858522
36. Martin Porcheron, Joel E Fischer, and Sarah C. Sharples. 2016. Using Mobile Phones in Pub Talk. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1647–1659. https://doi.org/10.1145/2818048.2820014
37. John Rooksby, Timothy E Smith, Alistair Morrison, Mattias Rost, and Matthew Chalmers. 2015. Configuring Attention in the Multiscreen Living Room. In *Proceedings of the 14th European Conference on Computer Supported Cooperative Work* (ECSCW '15), 243–261. https://doi.org/10.1007/978-3-319-20499-4_13
38. Harvey Sacks. 1984. Notes on methodology. In *Structures of Social Action Studies in Conversation Analysis*, J Maxwell Atkinson and John Heritage (eds.). Cambridge University Press, 21–27.
39. Harvey Sacks. 1995. *Harvey Sacks: Lectures on conversation*. Blackwell Publishing, Oxford.
40. Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4: 696–735. https://doi.org/10.1353/lan.1974.0010
41. Emanuel A Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29, 1: 1–63.
42. Emanuel A Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53, 2: 361. https://doi.org/10.2307/413107
43. L Srivastava. 2005. Mobile phones and the evolution of social behaviour. *Behaviour & Information Technology* 24, 2: 111–129. https://doi.org/10.1080/01449290512331321910
44. Tanya Stivers and Jeffrey D. Robinson. 2006. A preference for progressivity in interaction. *Language in Society* 35, 3: 367–392. https://doi.org/10.1017/S0047404506060179
45. Norman Makoto Su and Lulu Wang. 2015. From Third to Surveilled Place. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 1659–1668. https://doi.org/10.1145/2702123.2702574
46. Lucy A. Suchman. 1985. *Plans and Situated Actions: The Problem of Human Machine Communication.* https://doi.org/10.2307/2073874
47. Sherry Turkle. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
48. Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L. Sidner, and Timothy Bickmore. 2012. Designing Relational Agents as Long Term Social Companions for Older Adults. In *Intelligent Virtual Agents*. 289–302. https://doi.org/10.1007/978-3-642-33197-8_30
49. Ran Wei and Ven-Hwei Lo. 2006. Staying connected while on the move: Cell phone use and social connectedness. *New Media & Society* 8, 1: 53–72. https://doi.org/10.1177/1461444806059870
50. Alexandra Weilenmann and Catrine Larsson. 2002. Local use and sharing of mobile phones. In *Wireless World*. Springer, London, 92–107. https://doi.org/10.1007/978-1-4471-0665-4_7
51. Victor Zue, Stephanie Seneff, J.R. Glass, Joseph Polifroni, Christine Pao, T.J. Hazen, and Lee Hetherington. 2000. JUPlTER: a telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing* 8, 1: 85–96. https://doi.org/10.1109/89.817460