Voice Interfaces in Everyday Life

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

ABSTRACT
Voice User Interfaces (VUIs) are becoming ubiquitously available, being embedded both into everyday mobility via smartphones, and into the life of the home via 'assistant' devices. Yet, exactly *how* users of such devices practically thread that use into their everyday social interactions remains underexplored. By collecting and studying audio data from month-long deployments of the Amazon Echo in participants' homes—informed by ethnomethodology and conversation analysis—our study documents the methodical practices of VUI users, and how that use is accomplished in the complex social life of the home. Data we present shows how the device is made accountable to and embedded into conversational settings like family dinners where various simultaneous activities are being achieved. We discuss how the VUI is finely coordinated with the sequential organisation of talk. Finally, we locate implications for the accountability of VUI interaction, request and response design, and raise conceptual challenges to the notion of designing 'conversational' interfaces.

Author Keywords
Amazon Echo; conversational agent; conversational user interface; conversation analysis; intelligent personal assistants; ethnomethodology; collocated interaction

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
Voice interaction has become a feature in many commercial devices such as mobile phones and tablets. More recently, voice has become the primary interface with standalone screenless devices such as the Amazon Echo and Google Home. These interfaces, often referred to as voice user interfaces (or VUIs), conversational agents, or intelligent or virtual personal assistants, are described as embodying the idea of a virtual butler [23] that helps you 'get things done'. Researchers' adoption of such technologies as "*conversational* interfaces" [20] (our emphasis) resonates in many ways with the advertised user experience of such devices: specifically as technologies that it is possible to 'have a conversation' with and 'just ask' questions of. In addition, some VUIs (marketed as 'smartspeakers') are pitched as being especially suited to use in the home for a variety of purposes: to help with cooking, play music, access news and information, or play games with.

Despite the wealth of enabling research in computational linguistics such as natural language processing, dialogue systems, and computational sociolinguistics [21], research that empirically examines the social and interactional issues of VUIs in everyday use is lacking. In other words, with a few exceptions, little is known about the practical accomplishment of interactions with VUIs and the articulation of just how such interactions unfold as embedded in everyday life of VUI users. We believe this absence is significant, since our own study suggests a range of conceptual shifts that might need to be taken into account when designing VUIs for home settings and more broadly.

Our work is in the tradition of HCI and CSCW research that deploys technology to study the situated and emergent lived experience in the home [37]. In this way, we are continuing recent work emerging in CSCW that has begun to examine VUIs in collaborative action [26], for social settings such as meetings [18], and socialising with friends in a café [27]. Our study reports findings from month-long deployments of the Echo with the Alexa voice agent in five households. Audio capture was selectively performed by a separate device, a Conditional Voice Recorder, to collect over 6 hours of verbal exchanges involving the Echo in some way.

Our study draws on the traditions of Ethnomethodology and Conversation Analysis (EMCA) [8,32], as is common in HCI literature (e.g. [24]), to examine the various ways in which the Echo was implicated in talk. In the main part of our study we explore the ways in which the Echo is embedded into the situational exigencies of the home (such as other activities going on during use), and how its users account for the interactional work that use involves. We then look at the sequentially organised ways in which VUI use is achieved in a multiparty conversational setting and conclude by discussing three key issues emerging from our findings: conceptual concerns regarding the framing of

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3174214

interaction with VUIs as 'conversational'; the implications arising from the accountability of requests made by users to VUIs; and finally, the design and role of responses from VUIs as interactional resources for further action.

RELATED WORK

There are three broad areas that our study connects with. Firstly, we set the scene for our paper by covering the development of VUIs. Then we consider the role of conversation analysis in HCI when addressing VUIs specifically, noting clear absences and also formulating the shape of an emerging new area. Finally, we take an orthogonal perspective into account: namely, the methodological challenges associated with the design, deployment, and study of technologies in the home.

Voice User Interfaces

As we have already noted, there are multiple ways of naming machines that people can 'talk to', including conversational agents, or intelligent personal assistants. This is a broad category, however, so we employ *VUI* to indicate our focus on spoken word interactions. In doing so, we are necessarily distinguishing our work from the study of chatbots, such as Facebook M, that involve practices of reading and writing (typing messages and reading responses). We focus specifically on interfaces that are *primarily* voice-based, in which the user talks to the device and the device responds with a synthesised voice. Current commercially-available examples include Alexa as found in the Amazon Echo, and Assistant as found in the Google Home. We also note a slight distinction between this interest and virtual or embodied humans/agents, such as SimSensei [5], that are spoken to but also include a *visual* representation of a human counterpart that audibly and visually responds.

Of course, machines that can be spoken to via 'natural talk' have a considerable heritage in both science fiction (e.g. HAL 9000 from *2001: A Space Odyssey*) and research (cf. [16]). We note that many of these ideas have been realised already, with various systems created to help in domains such as healthcare decisions making [5], guiding visitors in museums [14], and companions for the elderly [40]. Early systems focused on specific sets of tasks such as providing users with weather information through a telephone call [42], however over time they have taken on increasingly complicated forms and functions including the embodiment of anthropomorphic qualities [23]. Recent systems designed for use in the home use an internet connection for speech processing and information retrieval, first explored in portable agents on mobile devices [13].

Talking to Computers

While speech technology research in general has long considered linguistic models and their relation to VUIs, the connection between conversation analysis and VUIs is a limited one. Gilbert et al. [9] argue that the findings from conversation analysis can indeed be employed in the design of human-computer interactions. Such prior work has employed conversation analysis to inform the design and development of computational models of conversation to support 'conversational' agents, with, for example, agents being designed to adopt elements of the turn-taking systematics [33]. There are some contradictions inherent in this approach, however, particularly in conversation analysis' adoption as a way of 'modelling' conversation. It has been pointed out that conversation analysis, which draws upon an ethnomethodological perspective [8], shows that human-to-human talk is in fact not bound or restricted by rules, or formalisable, but instead consists of sequences in which utterances and turn-taking arrangements are locally managed and negotiated in and through their production [3]. Nevertheless, our work does *not* concern itself with questions about whether a computer that 'talks' as-if-it-were-human can be created and sets aside such concerns, instead orienting to an ethnomethodological perspective of unpacking how interaction with VUI is achieved *within* talk-in-action. Specifically, this paper seeks to explicate how people routinely occasion, attend to problems with, and cooperatively manage interactions around the Alexa VUI in everyday home life.

Aside from a short-lived confluence of interests in the 1990s between EMCA-informed researchers, HCI, and AI [9], research on the use of VUIs has only recently returned as a topic of interest within the HCI and CSCW communities. Luger and Sellen [17], through interviews, problematise the limited functionality of existing commercially-available VUIs, stipulating that a "gulf" exists between expectant and actual capabilities of the systems. However, Luger and Sellen do not focus their work on an examination of actual settings of use as they happen. More closely related to our paper here, however, is work by Pelikan and Broth [24], who inspect the interactional organisation of human-robot interactions, identifying the practical competencies through which users of the robots in question adapt their talk to improve the accuracy of spoken word transcription by the robot, demonstrating how the limitations of automatic speech recognition may be partly overcome through methodological innovation on the part of the user. Perhaps most connected to this paper is our earlier study [27], which identifies the characteristics of interaction with VUIs on mobile devices and how such interactions unfold in multi-party social settings. This work examines how requests to VUIs are performed, attended to, and managed within the setting. However, to the best of our knowledge, we find that there are no examinations of the ways in which users of VUIs designed for the home practically and interactionally situate 'utterances' from the device within their ongoing conversational setting, nor how users' own utterances are brought off as directed requests to the device.

Studying Technology in the Home

Our final set of literature in this triangulation relates not to VUIs but to the nature of the intended setting that devices like the Echo or Google Home are targeted at. Places like

the home have posed significant methodological issues for HCI and Ubicomp specifically. In this way, we situate our work within a tradition established in part due to the growth of Ubicomp but becoming widely applicable in HCI too; this approach seeks to understand how interaction occurs and becomes embedded within lived experience [38], and further how these qualities can be exploited for design [4]. Methodological complexities connected with this line of work [37] have perhaps led to researchers experimenting with a range of data capture approaches so as to gather richer understandings of 'embeddedness'. For example, Ferdous et al. [7] incorporated home visits and video capture of technology use in family activities, while Rooksby et al. [31] set up video cameras to capture television watching and the use of mobile devices. However, while such use of video capture in the home offers the opportunity for very rich analyses, because we were interested in looking at VUI use over an extended period of time (coupled with an absence of a routinised, predictable use case for many VUIs), we were led to move away from video towards audio-only. As we describe next, in response to this problem, we developed a device to deploy alongside the Echo to record interactions with the Echo automatically, in the spirit of Pizza et al. [25].

RECORDING VUI INTERACTION IN THE HOME

We recruited five households to take part in a month-long study in order to capture naturalistic use of the Echo in the home. Three of our five households were inhabited by couples, while the other two households were inhabited by families consisting of two parents and two children. The age range of participants spanned late-20s to mid-50s. Each household was given an Echo, configured with a household member's Amazon account, and the Alexa companion app installed on their personal smartphone. Households freely selected the positioning of their Echo and could relocate it as desired. Four of the households placed the Echo in a kitchen or dining area, while one placed it in a living room. To capture Alexa use we deployed a second purpose-built device, a Conditional Voice Recorder (CVR), depicted in Figure 1, that is activated when a proximate Echo is used. The CVR captures audio using a conference microphone but retains only the last minute in a temporary buffer. When the 'wake word' *Alexa* is detected, the CVR saves the prior minute and records one further minute of audio (this period is extended if the wake word is heard in the subsequent minute). The CVR also features a button to turn off audio capture, and two LEDs (blue and red), that indicate when the CVR is 'listening' (blue) and when it is recording (red).

The resulting corpus consists of over 6 hours of recorded data. Within this, we identify over 883 distinct 'request' utterances, i.e. talk that is directed to the Echo in a seeming attempt to get it to 'do something', e.g. answer a trivia question, play particular music, or set a timer. Often these requests formed part of a larger sequence which might encompass various other requests that are temporally or topically related. Our corpus contains 185 of these.

Figure 1. Conditional Voice Recorder (CVR).

Approaching the Data and Interactional Phenomena

Taking an ethnomethodological and conversation analytic perspective [8,32], we were interested in how participants organised their actions with and around the Echo. In particular we examined how participants, as conversationalists, analysed moment-by-moment unfolding interactions with and around the device (and with one another, of course). Our hunch was that various conversational methods would come into play and be adapted so as to 'get stuff done' with the Echo. We note that what we are *not* doing in our analysis is seeking to treat the device as a *participant* in conversation. The use of conversation analysis as a tool to unpacking the sequential and embedded practice of situated actions is an established technique in human-computer interaction (e.g. [9]).

The fragments of data we present follow a set of standard transcription conventions [1,11]. For reference, we note where pauses take place ((.) or (1.4) for 0.1 and 1.4 seconds respectively), where talk is LOUD or °quiet°, where it is spoken <faster> than usual, and where an utterance is elong:::ated. Overlapping talk is represented using indentation and [square brackets], ((unspoken actions)) are given in double parentheses. All names have been altered from the original transcript. Talk that appears to primarily be addressed to the VUI is highlighted in **bold** and has a blue background. In turn, the synthesised speech produced by the Echo is identified by the label 'ALE' (i.e. *ALExa*) in transcripts and has a green background. The inclusion of synthesised speech as part of the transcript is not to suggest any conceptual equivalence between participants and the Echo, but is merely a convenient way of presenting device output as it appears temporally in interaction.

The fragments we examine in this paper are taken from one household that we will call the Kent family. We have selected the Kents as the analysis of a single case [34] so as to provide a series of "vivid exhibits" [2] of the broad array of methods we find participants employing across our corpus. That is to say, that through our analysis we identified the ways in which our participants 'used the Echo' and we present the fragments here as *exemplars* of participants' sense-making interactional work; they are *not*, however, the sole example of such interactional methods in

our corpus. Following an ethnomethodological orientation, we take it as given that participants continually work to order their own interactions and rely upon the orderly features of others (so as to analyse what one another is doing and thus 'go on'). This means that we are seeking to exhibit *just some* of the ways in which participants bring the Echo practically into that interactional order (i.e. of, in this case, a family meal). We leave it to future work to validate, refute, or add to our findings. In our examination we also bring to bear not only our own experiences of using the Echo, but also other VUIs such as Google Home and Siri.

Inspired by a similar line of approach to Reeves and Brown [28], we look at the family's interactions in two interconnected ways. Firstly, we examine the ways in which the Echo is made 'at home' and *embedded* into the various activities of home life. Interaction with the device does not take place as a singular indiscriminate event but rather is achieved as a situated action as part of—or rather, *embedded* within—the life of the home. Secondly, we turn to unpacking how the VUI features in *sequential courses of action*, i.e. the orderly production of conversation.

HOW VUI INTERACTION IS MADE 'AT HOME'

There are two parents, Susan and Carl, and two children around ten-years old, Liam and Emma, in the Kent Family. They have been using the Amazon Echo for approximately a week and have developed a reasonable familiarity and competence in its use. Of course, the fragments that we will use as well as our broader dataset does not offer a clear glimpse of long-term appropriation. Rather, what it *does* do by virtue of its capture at the beginnings of use, is to surface some of the initial ways that participants explore the uses of the device and work to (albeit often unsuccessfully) align the Echo to the social setting of the home.

Embedding the Echo into the Activities of the Home

In being present in the home, the Echo comes to be inextricably intertwined in the various ongoing activities that take place there. Our data is replete with sequences of interaction in which participants address the Echo in some manner and incorporate its output into the scene while also engaging in multi-party conversation and completing other activities in the home (as we will see). The Kent family are eating an evening meal all together at the dinner table on Mother's Day. The Echo is placed on the top of a bookcase that is used as a sideboard in the dining room. Our first fragment starts with Susan, the mother, announcing to the others that she would like to play Beat the Intro "in a minute". Beat the Intro is a game available for the Echo that the Kents have previously played together; it involves listening to a few seconds from the start of a song which players must then guess by announcing the song and the artist. The game is a "Skill", a third-party developed installable feature for the Echo. After Susan's announcement, Liam produces an assessment of this ("oh no") and then an elongated "no" as Susan then instructs the Echo to play the game. Carl mentions Mother's Day, while Susan instructs Liam to eat his food. Susan then attempts another instruction to Alexa to "play beat the intro".

```
01  SUS  i'd like to play beat the intro in a minute
02  LIA  [ oh no:: ]
03  SUS  [ alexa   ][ (1.1)  ] beat the in[tro
04  CAR             [ °yeah° ]
05  LIA                                   [°no:::...°
06  CAR  (0.6) it's mother's day? (0.4)
07  SUS  it's (    ) yep (.) listen (.) you need to keep
08       on eating your orange stuff (.) liam
09       (0.7)
10  CAR  and your green stuff
11  SUS  alexa (1.3) alexa (0.5)=
12  CAR                          =°and your brown stuff°
13  SUS  play beat the intro
```

Fragment 1: I'd like to play Beat the Intro in a minute

Our first observation is that addressing the Echo—here located in instructions to "play beat the intro"—is embedded amongst *multiple activities*, or 'courses of action' if you will, that the family are working to accomplish together. For instance, the family are eating dinner together, and they are talking about that eating (lines 07-10 particularly). Requests for compliance from Liam are produced by Carl amongst Susan's initial instruction to the Echo (line 03), where Carl counters Liam's negative response to Susan's preparatory utterance "I'd like to play beat the intro in a minute" with the reminder that "it's mother's day?" (line 06). Activities that we might also gloss broadly as 'parenting' turn on establishing appropriate ways of behaving during mealtimes, particularly for younger members of the family, such as the instruction to Liam to "keep on eating your orange stuff". All the while, we find these other concurrent activities closely geared into the organisation of Susan's further requests to the Echo. For instance, Susan's second instruction commencing on line 11, is interleaved with Carl's continuation of Susan's prior request to Liam to eat his food. Carl provides a series of and-prefaced turns: "and your green stuff" on line 10, and "and your brown stuff" on line 12.

These initial observations offer a consonance with prior studies of technology use in the home and how such technologies get drawn into the organisation of home life as resources for action (e.g. see Rooksby et al. [31]). Empirical accounts such as these present a more nuanced perspective to the conceptualisation of such technologies like the Echo as disruptive to established moral order by drawing attention away from interaction with co-present others [39]. Rather, we see here how homes are inherently multi-activity settings in which devices get recruited into and are regulated through the ongoing cooperative and collocated activities that take place in the home [29,31,38].

It is also important to note the design features of VUIs which tend to permit this meshing with activities in the home. Specifically, devices like the Echo provide 'always-on' 'always-listening' capabilities (not without posing considerable ethical and privacy conundrums, however while recognising the importance of this topic, we note that such matters are not part of this particular paper). This leads to the continuous availability of address via the wake word.

Thus, occasioning use of the Echo, and to proceed to interact with it, requires little in the way of movement or much coordinated action from other members (although as we will see later, it is even subtler than this regarding the production of silence). This means that Echo use may be initiated with relative ease through everyday talk, in the hurly-burly of other ongoing activities. Such is the incipient availability of the device that we *rarely* see the kind of action as can incidentally be seen on line 01, where a preparatory account is provided by Susan.

Echo Use and the 'Politics' of Control

It should come as no surprise that the regulation of VUI use—who can address the Echo, when, and how—is achieved by participants in various conversational ways. Our initial Fragment 1 furnishes us with insight into the ways that control of the Echo comes to be managed as a socially organised matter in what we could gloss here as the 'politics of the home'. Specifically, we draw attention again to lines 01-06 in Fragment 1, and the ways in which addressing the Echo, the selection of activities it provides (to play Beat the Intro), and the implications of that for the assembled family (that it will involve a collective engagement in a game at the table) take place around participants' orientation to the 'regulative work' of the specifics of this particular family gathering. So, for instance, we see this regulative work constituted in Carl's reminder of it being Mother's Day, directed at Liam, whose negative response was occasioned by Susan's instruction to the Echo. Carl's reminder here constitutes an analysis of Susan's rights: i.e. that it is her turn to address the Echo and also her right to formulate the instruction and its implications for the seated family.

Deepening this point, we now turn to our second fragment, where addressing the Echo is regulated in a different way. This fragment is from a longer sequence of interaction from a few minutes after the family have finished playing Beat the Intro together, and are now trying to play a different quiz Skill, Quiz Master. The Kent family are having trouble recalling the name of the Skill, so Susan has used her smartphone to look it up (omitted from this transcript). As we join the fragment, Emma takes advantage of this opening, while Susan is busy, to perform a request to the Echo to "resume music". This instruction is part of a broader in-joke at the table in which the children attempt to instruct the Echo to play music the parents do not necessarily appreciate or wish to hear. Susan attempts to talk to Alexa, but the music starts playing. This is then followed by some laughter, after which Susan completes her instruction to "open quiz master".

01	EMM	**alexa**
02	SUS	no hold on a minute=
03	EMM	=**resume** [**RESUME music=**]
04	SUS	[**alexa alexa**] =oh:
05	ALE	((music starts playing))
06	EMM	((laughs))
07	SUS	**alechsah! (1.3) open (.2) quiz master**

Fragment 2: Alexa … RESUME Music

We see here something of a 'competition' between Emma and Susan to address the Echo. As we mentioned earlier, the Echo is designed to be readily available for address at any point, meaning that participants effectively have 'equal access'. This leads to the emergence of various conversational methods to regulate and manage that access, as we see here. Emma initiates her instruction to the Echo in line 01, which is only partially in flight as it is interpolated by a next turn from Susan instructing Emma to "hold on a minute". While Emma does not speak over Susan she nevertheless closely latches a continuation of her instruction to the Echo in line 03, i.e. Susan's instruction does not lead to a course change, which Susan appears to analyse as such through her overlapping talk with Emma in line 04. Emma's continuation involves a repeated element ("resume") and a raising of volume during the overlap with Susan's instruction. This sense of a participant managing another's utterances to the device further exemplifies how VUI control becomes regulated as a social situated matter in and through interaction among the members of the setting. Our point here is that control of the Echo is not somehow separate from the setting, but rather is deeply embedded in its social order, as produced by its participants and their analyses of that social order.

Accounting for the Echo in Interaction

Our final point about the embeddedness of VUI use for its users is the way in which it must be brought into the accountability of social settings. By 'accountability' we mean to say that people routinely attempt to produce social actions in such a way that they appear as account-*able* to others and the situation. This is a continual matter of concern for members of society to the extent that where there are possible deficiencies in the accountability of social actions, members routinely work to offer up accounts of what it is they are doing ('explanations', perhaps).

Consider in Fragment 1 how Susan offers one such (prospective) account for a subsequent action, i.e. "I'd like to play beat the intro in a minute" in line 01. Susan's utterance here prepares that account as a 'frame', we could say, for the ways in which her subsequent instruction is to be made sense of by co-present others. Susan's account for her possible future action displays a sensitivity to how that action might be treated by the rest of the family (she also produces it as a preference, "I'd like", rather than a definite "we're going to"). There is also a broader sense in which all kinds of interaction with the Echo are treated as accountable to the situation. For instance, in Fragment 2, the beginning of Emma's instruction on line 01 ("Alexa") leads to Susan's rapid analysis of Emma's address to the Echo as presumptive, out-of-turn, and temporally problematic (i.e. "no hold on a minute", line 02). In other words, talk directed to the Echo is accountable to the coherence of the ongoing conversation, and equally the situation in which the conversation unfolds. Generally speaking, addressing VUIs involves the production of utterances in circumstances that frequently feature other

participants, meaning that such utterances are treated in similar kinds of ways to the ways that all social actions are treated: as accountable to the situation they are in.

HOW VUI INTERACTION FEATURES IN TALK'S SEQUENTIAL ORGANISATION

We have seen how the Echo comes to be enmeshed in the multi-activity of the home, the organisation and regulation of device control, and the accountability of utterances. Yet a significant element of VUI interaction is how it is made to fit into the orderly, *sequential* organisation of talk, i.e. how interaction with a VUI device is accomplished in a turn-by-turn, moment-by-moment unfolding manner. We will now start to unpack the details how the Echo comes to be made 'at home' in the sequential organisation of talk.

First, a word on what we mean by sequentiality and how conversation analysis treats it. Schegloff argues that sequentially is "any kind of organization which concerns the relative positioning of utterances or actions [...] turn-taking [in conversation] is a type of sequential organization because it concerns the relative ordering of speakers" [35]. In other words, conversation, such as those that involve addressing and listening to VUI input or output must necessarily integrate device 'utterances' into the sequential order of talk. Importantly, sequentiality differs from mere temporal ordering (although it can take advantage of it, of course), not only in that it encompasses actions that occur temporally in tandem (such as overlapped talk), but that the sequential coherence of conversation is a continuous achievement by conversationalists, who are seeking to assemble the sense of those actions which are often outside a basic temporal order. For instance, a speaker might answer a question several turns subsequent to it being posed in a conversation (which might be accounted for by a speaker in various ways, e.g. prefacing "before I answer your question…" to their turn).

In the following sections, we examine two key methodical accomplishments of action with the Echo in turn. Firstly, *addressing the VUI*, i.e. how *input* to the device is achieved. Secondly, we look at how participants *deal with responses from the VUI*, i.e. what is 'done' interactionally, sequentially with its *output*, or even the absence of output. We are deliberately using 'input' and 'output' here to ensure that description of human-VUI interaction reflects the ways that participants seem to treat the device so as to avoid anthropomorphic characterisations or conflation.

Addressing the Echo in a Conversational Setting

Addressing the Echo involves producing utterances that are formatted in such a way so as to be detected as *requests* by the device. These requests may emerge across several turns-at-talk (e.g. see Fragment 2 for a complex example) even when there is only one user present. Requests typically involve two kinds of formulations: as a 'question' (e.g. "what is the weather?"), or as an instruction (e.g. "play beat the intro"). In producing such requests as a matter of addressing the Echo, we find that participants (as competent conversationalists) bring the device into the sequential organisation of talk in at least three connected ways (there are no doubt more). Firstly, we look at how requests are produced in ways that fit into and themselves adapt some of the basic turn-taking 'mechanisms' of talk [33]. Secondly, we see how request production often involves the co-production of silence (i.e. the withholding or suspending of turn-taking) so as to aid the participant producing the request. Thirdly, requests are sometimes not the sole domain of one participant but rather sit within collaborative aspects of the sequential order of talk.

To help us exhibit these features, we now introduce Fragment 3 below, taken a few moments after Fragment 1. In this fragment, Carl takes up Susan's thus far failed attempt to start the game Beat the Intro. Susan complains that the Echo does not work for her, but after several seconds, Carl's request also appears to have failed. Emma remarks "she didn't like that". Emma then produces a revised version of the request during which Carl questions whether the game really is called "beat the intro". The Echo responds incorrectly and asks a question, and Emma closes the sequence by responding negatively. Carl expresses a sense of exasperation with "we played it the other night!", and finally Susan attempts the instruction again, which is met by further silence from the device (4.5 seconds).

```
01  CAR  ale[xa (1.0)     ] bea:t: the (.) intro
02  SUS     [ ((laughs)) ]
03  SUS  it does it for you
04       (5.0)
05  EMM  nope (.) she didn like tha:::::t
06  EMM  alexsa [ (1.3)            ] play beat the intro::
07  CAR          [ is it called
08                 beat the intro? ]
09       (2.1)
10  ALE  you want to hear a station for b b intro
11       [ (0.5) ] right?
12  EMM  [ °no:° ]
13  EMM  (1.1) no: (.) i don't alex:a (0.5) no!
14  ALE  (1.3) alrig↑ht
15       (0.7)
16  CAR  we played it the other ni:ght! the game we
17       played the [ other night ((laughs)) ]
18  SUS             [ yeaherr:: alexa         ] skills (.)
19       beat the intro
20       (4.5)
21  SUS  °uh::↓:°
22  EMM  she didn like tha:↓:t
23  SUS  alechSA::::::
```

Fragment 3: Alexa … Play Beat the Intro

We will return to this fragment in the sections below.

Building Requests into Conversational Turn-Taking

As in any conversation, the Kent family members display attention to the ongoing sequential organisation of the conversation. This sensitivity enables them to locate moments in unfolding talk where a next-turn may be possible. One of the key features conversationalists orient to is the turn-constructional units (TCU) of talk, i.e. a hearably, situationally, 'complete' part of an utterance that leads to a possible transition relevance place (TRP) where another speaker *might* opt to take their turn [33]. For instance, to use Sacks et al.'s example [33:702], a reception desk might ask a caller "what is your last name Lorraine?",

where a TCU is "what is your last name" since for the caller this part of the utterance is possibly complete (as an adequate question directed to the only other party 'present'). In this example, a TRP lies just after "name" is uttered, indicating a site for possible speaker transition.

For users of the Echo, we noticed that the ways of addressing the device provide for certain conversationally specific TCUs and therefore TRPs. Consider for example Carl's questioning of the name of the Skill ("is it called beat the intro?") in Fragment 3 (lines 07–08) and just how he inserts it sequentially into Emma's utterance. Carl produces this question precisely in the 1.3 second gap between Emma's production of the wake word "Alexa" and subsequent request to the device "play beat the intro". Consider also the request performed by Susan on line 03 of Fragment 1, where she utters "Alexa (1.1) play beat the intro" while Carl quietly says "yeah" during the 1.1 second pause. Carl's "yeah" provides a counter to Liam's rejection of Susan's preparatory utterance in line 01, and, importantly, this "yeah" is positioned at the precise moment after Susan's production of "Alexa" — Carl appears to be orienting to this regular pause. Similarly, in Fragment 2, we see in lines 01-03 how Susan also takes the turn from Emma after she utters "Alexa".

In other words, the wake word "Alexa", in the analytic work of Echo users, seems to be routinely oriented to as a TCU, i.e. a 'complete' utterance that may possibly lead to a turn transition. The syntactically formulaic nature of input production to the Echo and other VUIs, i.e. that of *wakeword-gap-request*, enables competent device users to project this gap, to constructively minimise silence, and to therefore offer the possibility of taking advantage of the gap to self-select and take a turn-at-talk. Often this also leads to the original requester interacting with the Echo then selecting to resume talk following this interweaved turn [15:301–304,33]. Further, a preference for minimising overlap in talk [36] also seems to be in operation as the request is made. For instance, in Fragment 3, line 06, Emma continues seamlessly with her request. In Fragments 1 and 2 we see similar examples including even more closely latched talk. It may be that this practice of resumption occurs in order to minimise overlap to improve the transcription accuracy of the Echo.

Mutually Producing Silence During a Request

Request production is collaboratively achieved in various subtle ways. One key form involves suspending turn-taking during moments of address to the Echo. We see this at various points in our fragments, for instance in Fragment 3, at lines 18-20 during which Susan initiates a request, where the laughter in the room subsides noticeably as she produces the wake word. This kind of silence production, this withholding of turns and suspending of taking a turn (for a further 4.5 seconds in the example from Fragment 3), is one way participants do collaboration around request production. As VUI devices generally may struggle to differentiate different voices during automatic transcription, reducing background noise (i.e. other talk) seems to be a technique employed by users to improve accuracy (note we are not claiming that understanding the underlying mechanism is either known about by users nor even relevant). Prior work has established a similar preference for group silence in conversation following the performance of a request to a VUI [27]. Going further, we note that this sequentially subsequent suspension of turn-taking also offers space for increased hearability of a possible, expected, projected response from the Echo.

Other Kinds of Sequential Collaboration in Request Production

We found that Echo users often perform other kinds of collaborative action in order to produce requests. For instance, Fragment 3 shows Carl, Emma, and then Susan taking turns to address the device. The desired outcome is repeatedly not achieved (i.e. starting the Beat the Intro game), so the family alter their requests in subsequent turns. Request alteration here seems to occur in a twofold manner; first, by altering *prosody*, for example in the pronunciation of the wake word (e.g. lines 06, 13, and 18), and second, by semantic variation of the command word (e.g. none in line 01, "play" in line 06, "skills" in line 18). This again echoes prior work that demonstrates how collaborative action with VUIs is replete with repetitions and rephrasings [27].

Dealing with Responses

Having examined participants' requests *to* the Echo we must now turn to responses *from* the device, delivered as computationally synthesised speech. Just as with requests, we broadly find that conversationalists attempt to enfold Alexa-generated responses into the sequential organisation of talk. In this next section, we look the ways in which participants address the Echo in turns-at-talk, orient to the response from the device, and, if necessary, deal with the response if trouble has occurred. Such 'trouble' arises routinely in interaction with a VUI, and is well represented in the majority of sequences within our corpus. Next, we explicate just three ways (there may be more) that participants attend to VUI responses: orienting to silence in response, responses as suggestive of troubles, and repairing troublesome interactions.

Orienting to Silence in Response

Like moments of silence in everyday talk, where such silence is often treated as a trouble source (e.g. a long pause that follows someone asking a question may be heard as a negative response), Echo silences in place of expected moments of response may be met with a similar kind of analysis by participants [41]. Consider Fragment 3, where silences of 4.5-5 seconds ensue after requests from Carl (lines 01-04) and, later, Susan (lines 18-20). The silence that follows is treated as troublesome in these moments, which we can see in Emma's remark of "she didn't like that" after both moments. We also note that participants' sensitivity to delays in response can lead to other ways of attempting to resolve trouble. For example, Carl questions

whether the Skill is called "beat the intro" (line 07), offering an explicit candidate for the source of trouble, i.e. that his previous request might have been using an incorrect name of the Skill.

We note that the kind of sensitivity to silence displayed by participants here is different to that in everyday talk. There is an expected temporal delay in the device's response since the Echo must remotely compute a response, introducing latency of usually at least one second (in our corpus), however on occasion this response-time can be shorter or longer. But here we see how silence is treated as a non-response at some point and variously a failure of some kind. This connects with some of the points made previously: that participants often mutually produce silence to allow for VUI request production, and they often co-produce silence in projecting a response.

Responses as Suggestive of Trouble

Before we examine how the participants in our study sought to remedy problems, we need to look at a related issue: how responses themselves were treated at suggestive of trouble. Whereas in VUIs found on touchscreen devices (e.g. smartphones or tablets) voice-to-text transcription is often displayed on the screen, users of screenless devices have to rely solely on the auditory response (although they may find more clues as to what went wrong in the companion app supplied with most screenless devices). We find that there is a significant mismatch sometimes between the ways in which designed responses from the Echo appear to integrate indicators of the *form* of trouble, and actually how participants dealt with them. Although it is tempting for simplicity's sake to call certain Echo responses 'error messages', this would not be correct as these responses are not always the result of a computational error, e.g. they may be due to the VUI device mistranscribing the request. Nevertheless, these responses are an important resource for diagnosing and resolving the trouble.

Our next fragment, below, provides one such exhibit of how responses from Alexa may be dealt with. This fragment begins after the family have played Beat the Intro. Emma asks Susan to perform the request, a "normal quiz" (in contrast with Beat the Intro). Susan then directs an instruction to Alexa: "set us a family quiz". The first response from Alexa "I can't find the answer to the question I heard" leads to Emma producing a similar instruction to Susan's. Another similar response is provided by the Echo, leading to Liam joining in with his own instruction request. After more difficulties and some laughter, Carl twice attempts a similar kind of instruction ("enable family quiz") and gets a response from Alexa in the form of a question about enabling "Neil Family Quiz".

```
01  EMM  can you ask for a normal quiz?
02  SUS  alexa (0.7) set us a family quiz
03       (2.5)
04  ALE  sorry (.) i can't find the answer to the
05       question i heard
06       (0.4)
07  EMM  ALech-sa: (1.0) set: (0.5) a family quiz
08       (2.3)
09  ALE  sorry (.) i don't have the answer to that
10       question
11  SUS  °well°
12  LIA  alexa (0.9) [ ↑PLease set (0.4) a family quiz ]
13  E+C              [ ((laugh))                       ]
14       (1.6)
15  ALE  i wasn't able to understand [ the question i
16                                                heard ]
17  EMM                              [ ((laughs))      ]
18  LIA  ((makes high pitch noise))
19  CAR  alechsa! (0.8) family quiz
20  SUS  come on there's some theres some quizzes here we
21       could have a quiz (  )
22  CAR  enable family quiz
23       (2.1)
24  ALE  did you want to enable neil family quiz?
25  EMM  ((laughs))
26  SUS  YES!
```

Fragment 4. Set us a Family Quiz

Interestingly, the initial response from the device in lines 04-05 can be seen to imply a question-answer sequence ("I can't find the answer to the question I heard"), even though participants appear to orient to the sequence as a matter of instruction: we can see this in Susan's transformation of Emma's question to her, i.e. "can you ask for a normal quiz?", which becomes "set us a family quiz". The VUI miscategorises the instruction as a question (technically it *overspecifies* the request). This may be problematic in that the user may in turn orient to the Echo's miscategorisation rather than to the source of trouble. However, this seems to be largely ignored by the family, who take it in turns to repeatedly rephrase the request as slight variations of the first: omitting the "us" (line 07), adding "please" (line 12), and omitting the command verb "set" (line 19). In a sense, the device's responses to this point seem to be ineffective resources for the participants to resolve the trouble and get the device to work.

Repairing Troublesome Interaction

Developing the final point in the last section, we consider here more of what participants actually do in repairing troublesome interaction with the Echo. To begin, consider the interaction in Fragment 4 from a hypothetical VUI designer's point of view: it is likely that the Echo does not recognise "set" as a command to invoke the desired Skill. Each response produced by the Echo in Fragment 4 is met with a rephrased request by different members of the family in turn. The first two responses from the Echo (lines 04-05 and 09-10) are not treated by the family members as occasioning a need to significantly alter their instructions to the device: instead they respond with quite minor variations of the original instruction from Susan (line 02), and notably retain the word "set". The third response from the Echo (lines 15-16) is somewhat different, referring to a problem of 'understanding': "I wasn't able to understand". This leads to overlapped laughter from Emma[1]. Carl then produces a minimal version of the earlier instructions (line 19), but he seems to treat this as problematic since he

[1] While the response contains the aforementioned misspecification of the request as a "question" this may not have been heard over the laughter, and in any case, it is not seemingly oriented to.

quickly issues another instruction in which he changes the command verb to "enable". This finally leads to a response indicating progress (line 24), and thus, repair of the trouble. This again demonstrates practices of reformulating, or rephrasing requests as a feature of voice interaction [12,27]. Participants also repeat requests, altering prosody to attempt to get the device to work (in many situations, both greater impetus and a rephrasing is used in successive requests to the device), but we have not explored this here.

DISCUSSION

The presentation of our findings focuses on the practical achievements of VUI users, and thus itself forms the main contribution of this paper. Here, however, we move on to reflect upon what the implications of this study might be for HCI. Our points are broadly conceptual in character—we are loath to nail down strong practical implications and frame what follows as opening discussion for both designers and researchers.

The Misnomer of 'Conversational Interfaces'

Although our fragments have clearly shown a VUI device being made a part of everyday conversation, we reject the notion that such devices and interfaces are *conversational* in nature and that interaction with the interface is a *conversation*. We take the stance that 'conversational interaction' is a misnomer for this kind of human-computer interaction, and confuses interaction with a device *within* conversation with an *actual* conversation. Although participants featuring in our data certainly do recognisably employ methods of talk to accomplish various activities with the Echo, it is hard to make a case based on our data that responses from the device have a similar *status* to the conversation into which they are embedded.

In our opinion, the term 'conversational interaction' is unhelpful as it fails to distinguish between the *interactional embeddedness of VUIs* and conversation. Consider, for example, the adjacency pair (e.g. greetings, question-answer, or offer-acceptance), an 'atomic' organisational structure in talk that is employed in many of our everyday interactions [32]. In adjacency pairs the second pair part (e.g. answer) is sequentially and implicatively tied to the first pair part (e.g. question). What this means is that there is *no* distinguishing independent feature of a first pair part that definitively ensures that it is indeed, say, a question; instead the question-character of a first pair part is only endowed with that character in light of how a second pair part *treats* the first pair part (i.e. we could say 'answers make the question'). Interaction with a VUI may be seen to unfold in the same way that adjacency pairs in conversation do, yet importantly it is *pre-configured* to be this way *by design* rather than being a process that unfolds interactionally as described above. For example, early on McTear defined spoken dialogue systems as "computer systems with which humans interact on a turn-by-turn basis" [19]. While we do not disagree with this definition, the chosen terminology makes it easy to confuse *input-output* on a *turn-by-turn basis* with *turns-at-talk*. Turns, as well as adjacency pairs, however, are categorically different in that they are the building blocks that are simultaneously shaped by and renew the context of human-to-human interaction [10]. Our data shows the ways in which VUI interaction is fundamentally different from human interaction, demonstrably so in the ways in which responses from the device do not necessarily coherently follow the input. As we saw in Fragment 4, responses from Echo may categorise an instruction as a question. While it is possible to do this in everyday conversation (e.g. on being seemingly instructed to do something, one can respond "are you asking me, or telling me?"), users of the Echo seem to routinely treat this as problematic and troublesome output that needs fixing in some way or another, rather than as a response that recasts their own utterance as a question (which can be something conversationalists do).

Without the device able to 'understand' logical models of talk (and we would not want to start such a discussion here on whether a machine could, such a discussion has been extensively covered elsewhere, e.g. [3]), here we merely seek to sensitise the HCI community to treat the term 'conversational interaction' (and its derivatives) with suspicion, much in the same way that others have questioned the use of terms like 'natural' in designing interfaces that employ embodied action. O'Hara et al. [22] state that while narratives that frame interaction paradigms as allowing people to "act and communicate in ways they are naturally predisposed to" can serve a number of purposes (e.g. marketing and communicating to a wider audience), they also find the framing problematic. They go on to argue that the narrative of 'natural' interfaces situates the locus in the interface alone, ignoring the fundamentally in situ and embodied features that constitute interaction. In this work, our pragmatic response to these concerns was to explicate the members' concern of 'getting this thing to work', and to dispense with notions of 'talking to a computer'. Thus, our data shows how interactions with the VUI become *embedded in* turns-at-talk, that the device itself is fundamentally not treated as a conversationalist, and that the voice interaction is replete with categorically different features than conversation.

The cross-disciplinary perspective of conversation we have adopted here has the view that there are no predefined rules for which talk between people must follow, but that such 'rules' are established as achievements in and through interaction by conversationalists, situationally and moment-by-moment. Yet, VUIs presently operate on a different plane, adhering only to predetermined structure. This may aid their use in proffering predictability of the interaction for users, but it is distinctly non-conversational as human interlocutors treat it. As such, we believe this perspective projects a shift from treating design tasks for VUIs from conversation design to that of *request/response design*.

Accountability of Making Requests

Our second point concerns request design: what the VUI designers intend users to say to a device, and how they broadly conceptualise the necessary utterances within interaction design. Specifically, we argue that request design fundamentally needs to consider the *projective accountability* of requests to the contextual circumstances that the designer wishes a user to produce. The embedded nature of interaction with a VUI may occasion users to do additional work to make their actions accountable. This fundamental feature of social interaction around others intimates the need for VUI designers to consider the request as a matter of *collocated* action around others, and not in isolation (i.e. the request must be accountable to the situation at hand). There is a variety of possible reasons and situations in which the request may be considered in need of explicit accounting, for example when the request does not 'fit' with a way of talking, or a social situation (e.g. a family meal), or perhaps that it is embarrassing in some way, or any number of other reasons.

This is not necessarily a problem for VUI users, who readily seem to account for requests where relevant. Rather, it is a sensitivity that designers may benefit from. Designing for accountability of requests is *not* about bestowing some intrinsic features upon requests such that they are accountable, but rather, considering how requests might play out within locally occurring conversations into which a designed request may become lodged or embedded. In other words, it is important to realise that the kind of accountability we talk of here is not a property of action but rather is an *interactional achievement*. Accordingly, designers could ask themselves: "might the requests we design for users to say be awkward to utter in certain circumstances?", "when might they be inappropriate or perhaps unusual?", and "will users need to do lots of accounting work around others?". Additionally, there is a link between considering the accountability of request design and the reflections on "observable-reportable abstractions" by Dourish and Button [6], which offer "a means for users to rationalise the activity of the system and therefore to organise their behaviour around it, as interaction proceeds, for their own practical purposes".

Responses as Resources for Further Interaction

We showed how family members in our fragments treat the response (or lack of a response) as indicators for the occurrence of some kind of trouble. Responses from the VUI themselves are analysed by members for the 'account' of sorts they provide on the state of the VUI device. Our data shows the inadequacy of the responses as resources to furnish this analysis to allow users to proceed with the interaction. To provide more resourceful responses, designers may find it useful to consider Dourish and Button's advice on "observable-reportable abstractions", which provide "cues as to not only what the system was doing, but why it was being done, and what was likely to be done next, uniquely for the immediate circumstances" [6]. For instance, we identified that, as participants repaired interactions with the Echo, they also attempted to identify the source of trouble, be it a system problem or a transcription problem. Insofar as can be achieved with a VUI, the response from the device is the primary 'account' of the system state and indicator of trouble: no-response (silence) is treated as an indicator of trouble as well, but it provides no mechanism for further interaction, and does not make available the state of the system, allying the VUI with notions of a 'black box'.

Conversely, a response from the VUI that provides reference to the activity of the device, or the transcription it processed and what provisionally might or might not happen next, provides its users with *resources* that can support and occasion further interaction with the VUI device. In designing responses, we might suggest designers consider questions such as: "is this response an interactional dead end?", "what resources does this response provide for a possible next request production?", "what might possibly be 'done' with this response?", "at what points might a user interrupt and take the next turn?", and "how does the response design employ moments of silence?". Thus, we suggest a conceptual shift towards considering response design *as the design of interactional resources for users*, rather than as phrases that follow an imagined 'script' of interaction.

CONCLUSION

The analysis presented here explicates how VUI use is routinely accounted for and embedded in talk-in-interaction. By drawing on fragments from our corpus of recordings of Amazon Echo use collected from multiple homes, our findings reveal how the use of the Echo is made 'at home', as situated actions, and becomes *embedded in the life of the home* rather than that of a discrete singular isolatable event. Our data reveals that the incipience of interaction with a VUI is achieved through its ready-availability, yet users may still methodically account for a request given the social context within which the use is done. We also unpacked the use of a VUI as sequentially organised in and through talk in the home. Ultimately, we identified two collaborative activities in using a VUI: addressing the device in turns-at-talk, and dealing with responses from the device. Finally, we turned to transferring our findings from that of matters of interaction to conceptual discussion points, to inform and shape future research and design on the use of VUIs.

ACKNOWLEDGEMENTS

This work is supported by the Engineering and Physical Sciences Research Council [grant numbers EP/G037574/1, EP/G065802/1, EP/N014243/1, EP/M02315X/1, EP/K025848/1]. Extended fragments of the data used in this paper are available at https://doi.org/10.17639/nott.342. The source code for the Conditional Voice Recorder is available from https://github.com/mixedrealitylab/conditional-voice-recorder.

REFERENCES

1. J. Maxwell Atkinson and John Heritage. 1984. Transcript Notation. In *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, ix–xvi. https://doi.org/10.1017/CBO9780511665868
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. Graham Button, Jeff Coulter, John R. E. Lee, and Wes Sharrock. 1995. *Computers, Minds and Conduct*. Polity Press, Cambridge, UK.
4. 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
5. David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. *International Conference on Autonomous Agents and Multi-Agent Systems*, 1: 1061–1068.
6. Paul Dourish and Graham Button. 1998. On "Technomethodology": Foundational Relationships Between Ethnomethodology and System Design. *Human-Computer Interaction* 13, 4: 395–432. https://doi.org/10.1207/s15327051hci1304_2
7. Hasan Shahid Ferdous, Frank Vetere, Hilary Davis, Bernd Ploderer, and Kenton OHara. 2016. Technologies At Mealtime: Collocated Interactions In The Family Home. In *CHI '16 Workshop on Proxemic Mobile Collocated Interactions*.
8. Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall.
9. Nigel Gilbert, Robin Wooffitt, and Norman Fraser. 1990. Organising Computer Talk. In *Computers and Conversation* (1st edition), Paul Luff, David Frohlich and Nigel Gilbert (eds.). Academic Press, 235 – 257.
10. 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
11. Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. SAGE.
12. Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '13), 143–152. https://doi.org/10.1145/2484028.2484092
13. 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
14. 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
15. Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press.
16. 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
17. 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
18. Moira McGregor and John Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17). https://doi.org/10.1145/2998181.2998335
19. Michael McTear. 2002. Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computing Surveys* 34, 1: 90–169. https://doi.org/10.1145/505282.505285
20. Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3
21. Dong Nguyen, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2016. Computational Sociolinguistics: A Survey. *Computational Linguistics* 42, 3: 537–593. https://doi.org/10.1162/COLI
22. Kenton O'Hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor. 2013. On the Naturalness of Touchless : Putting the "Interaction" Back into NUI. *ACM Transactions on Computer-Human Interaction* 20, 1: 1–25. https://doi.org/10.1145/2442106.2442111

23. 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
24. 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
25. 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
26. Martin Porcheron, Joel E. Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with Conversational Agents in Collaborative Action. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17 Companion), 431–436. https://doi.org/10.1145/3022198.3022666
27. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17). https://doi.org/10.1145/2998181.2998298
28. Stuart Reeves and Barry Brown. 2016. Embeddedness and Sequentiality in Social Media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1050–1062. https://doi.org/10.1145/2818048.2820008
29. Jacob M. Rigby, Duncan P. Brumby, Sandy J. J. Gould, and Anna L Cox. 2017. Media Multitasking at Home. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (TVX '17), 3–10. https://doi.org/10.1145/3077548.3077560
30. Sean Rintel, Richard Harper, and Kenton O'Hara. 2016. The Tyranny of the Everyday in Mobile Video Messaging. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 4781-4792. https://doi.org/10.1145/2858036.2858042
31. 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
32. Harvey Sacks. 1992. *Harvey Sacks: Lectures on Conversation*. Basil Publishing, Oxford.
33. 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
34. Emanuel A. Schegloff. 1987. Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis. *Social Psychology Quarterly* 50, 2: 101–114.
35. Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511791208
36. Tanya Stivers, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter de Ruiter, Kyung-Eun Yoon, and Stephen C Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America* 106, 26: 10587–92. https://doi.org/10.1073/pnas.0903616106
37. Peter Tolmie and Andy Crabtree. 2008. Deploying Research Technology in the Home. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (CSCW '08), 639–648. https://doi.org/10.1145/1460563.1460662
38. Peter Tolmie, Andy Crabtree, Tom Rodden, and Steve Benford. 2008. "Are You Watching This Film or What?" Interruption and the Juggling of Cohorts. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work* (CSCW '08), 257. https://doi.org/10.1145/1460563.1460605
39. Sherry Turkle. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
40. 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
41. Robin Wooffitt. 1994. Applying Sociology: Conversation Analysis in the Study of Human-(Simulated) Computer Interaction. *Bulletin de Méthodologie Sociologique* 43, 1: 7–33. https://doi.org/10.1177/075910639404300103
42. Victor Zue, Stephanie Seneff, J. R. Glass, Joseph Polifroni, Christine Pao, T. J. Hazen, and Lee Hetherington. 2000. JUPITER: 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