Information Systems Frontiers (2020) 22:585–605
https://doi.org/10.1007/s10796-020-10015-6

Understanding Technology as Situated Practice: Everyday use of Voice User Interfaces Among Diverse Groups of Users in Urban India

Linus Kendall[1] · **Bidisha Chaudhuri**[2] · **Apoorva Bhalla**[2]

Published online: 7 June 2020

Abstract

As smartphones have become ubiquitous across urban India, voice user interfaces (VUIs) are increasingly becoming part of diverse groups of users' daily experiences. These technologies are now generally accessible as a result of improvements in mobile Internet access, [-8.5pc]Biography is Required. Please provide. introduction of low-cost smartphones and the ongoing process of their localisation into Indian languages. However, when people engage with technologies in their everyday lives, they not only enact the material attributes of the artifact but also draw on their skills, social positions, prior experience and societal norms and expectations to make use of the artifact. Drawing on Orlikowski's analytical framework of "technologies-in-practice" we engage in an interview-based exploratory study among diverse groups of users in urban India to understand use of VUIs as situated practice. We identify three technologies-in-practice emerging through enactment of VUIs on users' smartphones: looking up, learning and leisure. We argue that – instead of asking why and how users appropriate VUIs – identifying different kinds of enactments of VUIs present researchers and practitioners with a more nuanced understanding of existing and potential use of VUIs across varied contexts.

Keywords Voice user interfaces · Voice-based search · Technologies-in-Practice

1 Introduction

India has seen a steady growth in the number of smartphone users. As per the estimates reported by the technology consultant Counterpoint Research, there are about 650 million mobile phone users in India, out of which just over 300 million have a smartphone (E.T. T 2017). A joint report by Internet and Mobile Association of India (IAMAI) and KANTAR-IMRB, titled 'Mobile Internet in India 2017' (PTI 2018) reported the total number of mobile Internet users to be 456 million in December 2017 which was 17 per cent higher than December 2016. The report predicted that by June 2018, there will be 291 million mobile Internet users in urban India. These statistics show how smartphones have become ubiquitous across urban India paving the way for voice user interfaces (VUIs) to increasingly be part of diverse groups of users' daily experiences with technology. Voice User Interfaces employ modalities such as voice recognition and speech output to allow the user to interact with various functionalities of a digital device. They are implemented in variety of ways - from simple user interface elements to full-fledged conversational agents - combining prompts, grammars, and dialogue logic (also referred to as call flow) (Cohen et al. 2004). The emerging importance of VUIs is underscored by the fact that they are moving beyond being mere interfaces for interacting with smartphones or computers to computing platforms in their own right – with specialised hardware devices, unique services and ways for people to interact on them (Dale 2016). Increasingly,

✉ Linus Kendall
me@linuskendall.com

Bidisha Chaudhuri
bidisha@iiitb.ac.in

Apoorva Bhalla
apoorva.bhalla@iiitb.org

[1] Sheffield Hallam University, Howard St, Sheffield S1 1WB, UK

[2] International Institute of Information Technology Bangalore, 26/C, Electronics City, Hosur Road, Bangalore 560100, Karnataka, India

developers are able to create customised applications[1] that exist only via spoken interfaces and provide unique interactions and services. For instance, Amazon's own vision is for their device to become deeply intertwined in the everyday lives of regular families (Amazon Inc 2015).

With the burgeoning influence of smart voice technologies in everyday life, there is a gradually growing body of work on their usage patterns. There have been studies on the potential for voice-based agents in supporting meetings and collaborative work (McGregor and Tang 2017). Luger and Sellen (2016) explored the interactional factors affecting everyday use of different conversational agents across different devices in the United Kingdom. They found a gap between users' expectations and how conversational agents operated in reality, which eventually prevented meaningful engagement with these agents. Cowan et al. (2017) also reported infrequent use of intelligent personal assistants (IPAs), which according to them, were shaped by functional limitations, cultural norms and social concerns such as privacy and data transparency. In the home setting, there is a recent study on the everyday use of voice assistant devices such as Amazon Echo in collaborative action and informal conversation (Porcheron et al. 2018a). A study by Porcheron et al. (2018b) on the use of voice assistants among friends emphasised VUIs' ability - in contrast to smartphones - to democratise and make shared interaction with technology in social settings. Their study illustrated the use of VUIs specifically for looking up information in response to ongoing group discussions. Lau et al. (2018) studied the choice of use and non-use of voice assistant devices finding that privacy and security concerns were important reasons for non-use especially due to lack of trust in the company behind the device, as well as limited privacy control features. This included lack of ability to control secondary or incidental users' access to the device and its logs. Lovato and Piper (2015) in their study of voice input system for young children found three reasons for their engagement, namely, exploration for mostly fun, information-seeking and as way of operating a specific device. Druga et al. (2017) and Biele et al. (2019) talk about the use of voice user interfaces by children highlighting how the ubiquity of these interfaces create both new possibilities as well as raising new ethical and moral issues with regard to the conversations children and teens may have with them. Zamora (2017) study on the use of chatbots in routine life for users in India and USA has also shed some light on how users engage with such voice-based technologies to leverage a range of functionalities embedded in their everyday lives. However, most of these studies have so far looked at contexts within developed countries or at early adopters in developing countries.

Google launched voice search in Hindi in 2014 and additional Indian languages in 2017 (Turovsky 2017; van Esch 2017) with a specific aim to make "the Internet more inclusive". By providing localised input modalities - through keyboards and speech recognition - as well as search and automated translation of content into local languages, Google sought to "bring down language barriers". However, as Karusala et al. (2018) and others (Medhi et al. 2011) have shown, localisation and multilingualism in smartphone use is embedded in a complex negotiation of usability issues related to language input as well as social norms related to perceptions of literacy and English as a global language of the educated. Furthermore, the potential for voice technologies to support inclusion is complicated by both digital literacy in general as well as issues of space and privacy (Easwara Moorthy and Vu 2015). Robinson et al. (2018) has deployed an early-stage conversational speech probe in public space in a Mumbai slum. Their study brings out some interesting findings. They found adults to be more reluctant to try out the new device. They attributed this inertia to their lack of exposure to regular Internet search. The use of such technologies in public spaces led to two further important insights. First, it created friction among multiple users while approaching the system at the same time. Second, the importance of privacy concerns while using interactive search in presence of others. Finally, there is little work on the way voice user interfaces operate as part of gendered spaces. The importance of considering gendered patterns of access and use of new information has been long established, especially as ICTs become "domesticated" and increasingly prevalent in private spaces such as the home (Richardson 2009; Brown 2008; Venkatesh 2008).

Frequently when voice interfaces have been studied, they have been looked at in specific domains (McGregor and Tang 2017; Patel and Agarwal 2008; Richardson 2009; Zue et al. 2000; Porcheron et al. 2018b). However, as voice is set to become an ubiquitous platform in everyday lives we need to understand the diversity of use and user agencies and practices around it. On the one hand, we see a steady growth in the number of smartphone and mobile Internet users along with rising interest in using voice interfaces on smartphones in developing countries such as India. On the other hand, we see a gap in literature to capture and analyse how different groups of users are leveraging these voice-based new modalities (VUIs) on their smartphones and how their social contexts are shaping such usage (Schlesinger et al. 2017). Most of the existing literature on VUIs is either heavily focused on developed countries (Lovato and Piper 2015; Luger and Sellen 2016; Cowan et al. 2017;

[1] "skills", "actions" and "shortcuts" on Amazon's, Google's and Apple's platforms respectively

Porcheron et al. 2018a), exploring affordances of specific technologies such as "Siri" or "Alexa" (Porcheron et al. 2018a; Luger and Sellen 2016; Lau et al. 2018; Zamora 2017; Lee et al. 2019) or are looking at more specific groups of users, such as children (Lovato and Piper 2015; Druga et al. 2017; Biele et al. 2019), friend groups (Porcheron et al. 2018b; Lee et al. 2019), home device owners (Porcheron et al. 2018a; Zamora 2017) and so on. While drawing on useful insights from this existing body of literature, we extend our focus to a set of diverse users (in terms of their occupation, levels of education, gender and socio-economic background) and diverse possibilities of engagement with VUIs in practice within a larger context of urban India. Our selection of respondents in large Indian cities also brought in an additional dimension of a distinct linguistic culture shaped by multilingualism, code mixing and code switching[2] into the study.

In a broader sense, we look at how material structures of an artifact/technology create possibilities of actions by users and how such possibilities are further shaped by social norms and practices (Pozzi et al. 2013; Hsieh 2012). For instance, how women's use of public transport is shaped by gendered norms of mobility in a specific context. There are earlier studies on technology use that emphasise this recursive relationship between human action, technological and social structures (Zuboff 1988; Orlikowski 1999; Humphreys 2005). In examining early use of cellphones, Humphreys (2005) argues that effects of new technologies are always negotiated in people's interpretation and use of it. Hence, while trying to understand how new technologies work, one needs to pay attention to the social and cultural context in which users engage with new technologies (Humphreys 2005).

Following this orientation to technology use, we set out to understand the ways in which voice user interfaces affords social practices that are embedded in the everyday lives of smartphone and mobile Internet users in urban India. In our understanding, we are not interested merely in the usability or appropriation of voice user interfaces through smartphones, but rather how users enact these modalities through recurrent use that embeds the technology within their everyday contexts (Kumar et al. 2017). Our aim is to understand how recursive interaction between users, technology and social structures render use of voice user interfaces a socially entangled experience. This nuanced and contextually complex understanding of voice user interfaces, we believe, will yield important insights through which we can show how human agency of the users occupy a crucial position in experiencing more inclusive and diverse interactions with technologies (Sambasivan et al. 2011).

[2]Code mixing refers to the practice of interleaving two or more languages whereas code switching is the practice of switching between two or more languages within a single conversation or even sentence (Microsoft Research India 2018)

2 Analytical Framework

There are two specific dimensions of affordances that have been examined by information system researchers particularly in the context of technology use (Pozzi et al. 2013; Persico et al. 2014; Zheng and Yu 2016). Firstly, when people use a technology, how they draw on the material structures inscribed by the designers, and those added on through their previous interactions with technologies, their skills, power, knowledge, assumptions, and expectations about the technology and so on (Orlikowski 2000). Together, we can categorise these as technological affordances - specific features or facilities of the artifact coupled with users' conditions and contexts (Pozzi et al. 2013; Hsieh 2012). Secondly, while technological affordances address possibilities of human action through material properties, technology use is also shaped by the social environment within which any engagement with technology takes place. These can be labelled as social affordances (Pozzi et al. 2013; Hsieh 2012; Haider 2016; Treem and Leonardi 2013). These two notions of affordances imply a set of social practices afforded by the technology (Pozzi et al. 2013; Hsieh 2012). In this sense, our focus is not just on what people do with technologies (as they were designed to be used), but also on how people "can and do circumvent inscribed ways of using technologies — either by ignoring certain properties of the technology, working around them, or inventing new ones that may go beyond or even contradict designers' expectations and inscriptions" (Orlikowski 1999). This way of framing technology use is significantly different from technology appropriation, where the starting point is technology and how people appropriate its embodied structures (Orlikowski 2000). Instead, Orlikowski (2000) suggests we start from human action and how it enacts emergent structures through recurrent social practices. Thus, enactment of technology emerges through recursive interactions between people, technological structure and social structures. Enactment, as a practice-based lens, not only brings focus back on human actions, it also changes our treatment of technologies as artifacts with fixed properties to technologies as emergent structures reconstituted through human actions (Orlikowski 2000).

Following our interest in the recursive interaction between users, technology and social structures, we draw on the concept of "technologies-in practice" as developed by Orlikowski (1999). She argues that technologies can be analysed in two ways: as material embodiment of technical properties and as situated practices, that is, what people do with technological artifacts in their everyday lives

(Orlikowski 1999). Orlikowski uses the term "technologies-in-practice" to refer to "the particular structures of technology use that users enact when engaging recurrently with a technology" (Orlikowski 1999). This conceptual lens focuses specifically on how humans interact with technology and situatedness of technology use that depends on users' social positions and resources. In Orlkowski's framework of technology use, material structures or broadly technological affordances cannot be analytically separated from the social affordances. It is the interaction between the two that is realised through the notion of "technologies-in-practice". Orlikowski (1999) illustrates (see Fig. 1) this concept by looking at how structures (technological, social and institutional) interact with users' agency - how users employ interpretive schemes to apply and comply with socio-technical norms while enacting facilities offered by the technology. However, it is important to keep in mind that enactment within this framework can also result in ongoing, situated non-use of specific facilities or technologies. Thus, use and non-use is not a binary action, but rather a continuous trajectory of interaction with a specific technology.

3 Methodology

We deployed an interview-based exploratory study across several large cities covering people from different age, gender, educational background, and occupation. The data for our research was collected from May to July 2018 in three major metropolitan cities - Bangalore, New Delhi, and Kolkata. These three cities are listed in the top five urban centres in India, and each has a different official vernacular language (Kannada, Hindi and Bengali respectively). Moreover all the authors have professional and personal ties with these three cities which made it easy to identify and recruit respondents from diverse backgrounds. We conducted 29 semi-structured interviews with men and women in the age group 15-35, recruited primarily through snowball sampling where we used an initial set of respondents to get in touch with further respondents.

Our main aim for this study was exploratory - to observe a wide variety of everyday uses among diverse groups of people - and we sought to approach this through a qualitative, interview-based study. Accordingly, our sampling strategy was purposive based on a selection of diverse users when it came to gender, language, occupation, working and living conditions. Purposive sampling strategies are appropriate for exploratory qualitative studies where the goal is not to make broad generalised statements about a population group as a whole (Etikan et al. 2016). We used a form of critical case and variance sampling where we were seeking a variety of heterogeneous cases that could help illustrate a diverse set of concerns and uses (Etikan et al. 2016). The prerequisite for inclusion of respondents was that they were using Internet connected smartphones regularly and had been doing so for at least the past three months.

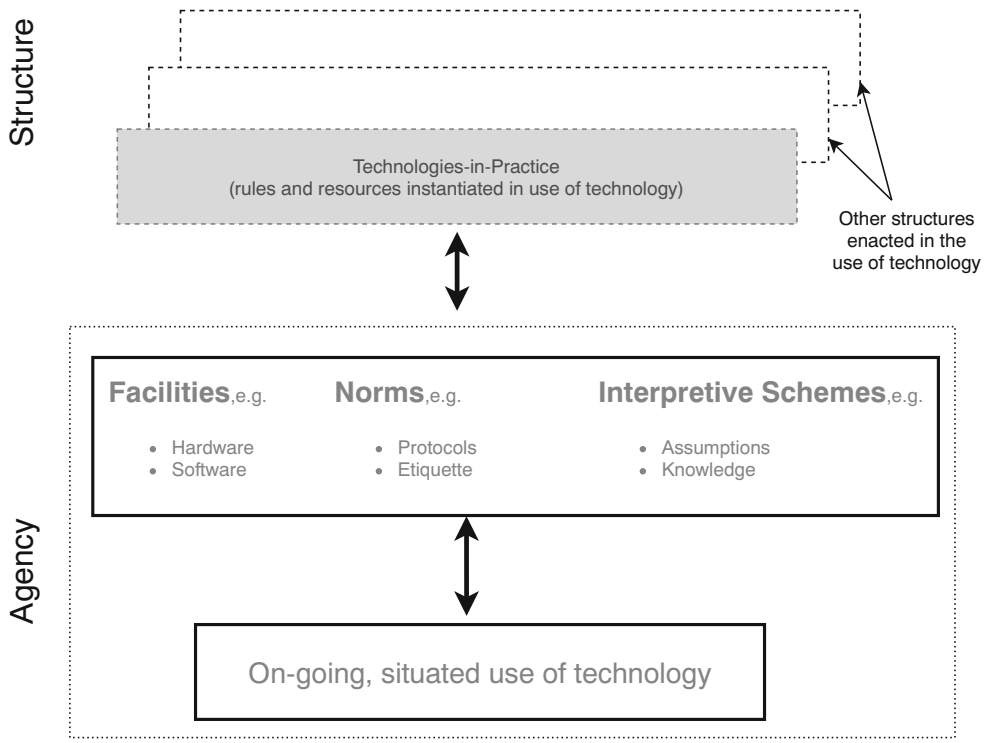

Fig. 1 Orlikowski's Technologies-in-Practice framework

Combining purposive and snowball sampling helped us get a group of respondents from similar backgrounds based on an initial purposive sample.

We selected participants from different educational[3] and occupational categories to represent socio-economic diversity. Occupational categories included, semi-skilled informal sector workers (Table 1), domestic workers (Table 1), self-employed merchants (Table 1; Table 2), and students (Table 2; Table 3). The students, however, were not a homogeneous category. While some of our student respondents were studying in premier institutes, others went to lower-end public schools or colleges. This posed a serious challenge for us when we tried to label respondents' socio-economic status through their educational qualification and occupation. Hence, instead of relying solely on these two categories we combined them with living arrangements, such as housing, neighbourhood, family size as a proxy for labelling respondents' different socio-economic backgrounds. As a result, other than gender, we use occupation, quality of education and living arrangements as a way to categorise our respondents into a higher (Table 2) and a lower socio-economic group (Table 1, Table 3). We employ these categories of gender[4] and socio-economic groups in our analyses of enactment of voice user interfaces.

Our emphasis on younger respondents was based on previous work where we found that this was the group that had thus far used voice user interfaces to a greater extent and our focus for this study was specifically on use rather than non-use. However, within the age group we covered there was considerable variation between the younger students, the more senior students and the group of working people in their mid 20s.

The diversity in age, occupation and gender also meant that our respondents had varying degrees of smartphone literacy. Relevant to our study, several of our respondents had (with the very recent introduction of 4G infrastructure, and the accompanied drastic price-reduction in data costs in India) only recently acquired smartphones and data connections capable of voice interface use. This meant that although many of our respondents had previously owned a mobile phone - even a smartphone - they did not necessarily have experience of services such as voice search or voice assistants that could be accessed only with high-bandwidth mobile Internet.

Table 1 Semi-skilled workers from the lower socio-economic group(names have been changed)

Participant	Gender	Age	Location	Education	Occupation
Praveen	M	21	Bangalore	High School	Vendor
Prerna	F	24	Bangalore	High School	Vendor
Radha	F	24	Bangalore	High School	Saleswoman
Rajat	M	22	Bangalore	College	Vendor
Rajesh	M	33	Bangalore	High school	Driver
Rakesh	M	22	Bangalore	High School	Vendor
Reema	F	33	Bangalore	High School	Security Guard
Rohit	F	22	Bangalore	Elementary	Vendor
Sagar	M	21	Bangalore	High School	Vendor
Shikha	F	24	Bangalore	High School	Saleswoman
Sudha	F	24	Bangalore	High School	Saleswoman
Sumit	M	22	Bangalore	High School	Vendor
Sunil	M	27	Bangalore	High school	Driver
Preeti	F	30	New Delhi	Elementary	Cook
Sahana	F	19	New Delhi	Elementary	Domestic Worker
Supriya	F	28	New Delhi	High School	Cook
Aparna	F	28	Kolkata	High school	Self-employed
Gopal	M	24	Kolkata	No education	Surgical mistri

Table 2 Students & professionals from the higher socio-economic group (names have been changed)

Participant	Gender	Age	Location	Education	Occupation
Anil	M	40	Bangalore	College	Vendor
Anupriya	F	19	Bangalore	College	Student
Ishita	F	21	Bangalore	College	Student
Nidhi	F	23	Bangalore	College	Student
Tanya	F	23	Bangalore	College	Student
Priya	F	22	Bangalore	College	Software Engineer

[3]We have used the terms "elementary" for completion of classes 1-9, "high school" for classes 10-12, "college" for any tertiary education.

[4]As per studies within the domain of ICTs for Development (ICTD), despite the access to the same technology, men and women show significant difference in the ways they make use of ICTs within a development context (Nguyen et al. 2017; Masika and Bailur 2015; Oreglia 2014; Balasubramanian et al. 2010; Best and Maier 2007; Volman et al. 2005). Given the crucial role of gender in both access to and use of ICTs, feminist science and technology studies (STS) and STS in general, have long established that technology and gender (and society in general) co-produce each other both by embodied structures of technology and in emergent use of technology in everyday practice (Faulkner 2001). Such studies explicate two kinds of relationship between gender and technology. First is gender in technology, where gender relations are embodied in the artifact (Cockburn 1983) and second is gender of technology, where in the gendering of the artifacts takes place through association rather than material inscription (Faulkner 2001; Lau et al. 2018; Cowan 1983). In case of association, sometimes it is hard to locate gender of technologies in its material properties. The gender of the technology only reveals in recurrent use as being reinterpreted as such (Berg and Lie 1995). In Human-Computer Interaction (HCI), in continuation with the "phenomenological turn", agency and identity of the users occupies a crucial position in designing interactions and here gender works as a major marker (Berg 1999; Bardzell 2010). As Bardzell (Bardzell 2010; Bardzell and Bardzell 2011) captures the vision of a feminist HCI, it is an attempt to "improve understanding of how gender identities and relations shape both the use of interactive technologies and their design" (Bardzell 2010).

Table 3 Students from the lower socio-ecomomic group (names have been changed)

Participant	Gender	Age	Location	Education	Occupation
Deep	M	15	Kolkata	High school	Student
Jyoti	F	15	Kolkata	High School	Student
Prithi	F	16	Kolkata	High School	Student
Priyanka	F	17	Kolkata	High School	Student
Sudesna	F	16	Kolkata	High School	Student

Another source of diversity was linguistic groups. Even though the cities we covered were dominated by a single vernacular language, many of our respondents, regardless of their occupational category and socio-economic background, were migrants and hence multilingual. For the lower socio-economic group, this meant that they at least spoke or understood Hindi as well as a regional language. For the higher socio-economic group, they were primarily English speaking and often understood at least one other vernacular language. The semi-structured interviews we conducted were based on an interview guide which was divided into sections which covered participants' use of smartphones, use of voice interfaces along with information related to location, time and kind of queries they used it for. The interview guide evolved throughout our study allowing us to check some of our early interpretations in later interviews (Krefting 1991). Employing a snowball approach, we stopped reaching out to new respondents as we reached a point of saturation when it came to both responses as well as demographic diversity (Etikan et al. 2016; Mason 2010). Data from our respondents was gathered in the form of interviewer notes and voice recordings. These were transcribed and when necessary translated into English. In order to analyse these findings we adopted the coding strategy of open and axial coding.[5] The transcripts, and in some cases the voice clips directly, were first coded by the researchers separately in an open coding process where we paid special attention to why, what for, where and when our respondents used the VUIs. From these open codes, we used axial coding to identify groupings of codes that yielded a hierarchy of categories or themes (Khandkar 2009; Price 2010; Wicks 2010). This process was done iteratively, and we returned to the data to re-code based on the axial codes identified. The final set of categories were based on codes that were shared by both researchers, as well as those identified by a single researcher. We could triangulate our themes between different geographical locations, different researchers as well as different groups of respondents (Krefting 1991). Through our multiple phases of open and axial coding, specific patterns of enactment of VUIs emerged along the categories identified. We employed our analytical framework to organise these categories into a theoretical frame. In this way, we could not only evaluate the fit of the theoretical framework to our data but also employ it to better understand the patterns of codes that emerged from the data. These methods of data collection and interpretation were aligned with our interpretive epistemology and approach to research. We draw on similar interpretive studies (Luger and Sellen 2016; Lovato and Piper 2015; Cowan et al. 2017) that have done exploratory work studying everyday practices in relation to voice user interfaces and conversational agents.

[5] Coding was conducted using regular word processing and spreadsheet software.

4 Enactment of Voice User Interfaces as Technologies-in-Practice

We report on three technologies-in-practice that have emerged through recurrent use of voice user interfaces in participants' everyday activities. We call these looking up, learning and leisure. Through these technologies-in-practice, we illustrate how different social positions as represented by gender and socio-economic status interact with the material structures to enact voice technologies on their smartphones.

4.1 The Looking up Technology-in-Practice

All participants in the study used voice for what we term "looking up" - different forms of informational queries. These were not only general searches, but sometimes map/location look ups or searches for a specific name or content for which they expected the VUIs to give a precise answer. Even though there were similarities between all participants in our study in the content they looked up, the way in which they approached looking up varied by gender and socio-economic position.

Common among all participants was looking up entertainment content in the form of videos on YouTube. Voice search was commonly used both directly in the YouTube app as well as in the Google search bar:

> Deep: "I use the voice search every day at least 5-6 times. I use it for songs, the songs which I don't know how to write. I speak the name [of the song] and it comes written."

> Priyanka: "I can search for songs, [especially in English], as I may know its name but not how it is spelled."

In addition to these types of searches, participants from the higher socio-economic group would use searches that provided direct answers to their queries. For example:

Nidhi: “Sometimes for weather, mostly for that I have used it. Otherwise if for finding short like short note on particular topics. If you want to know something like something trending is happening and you want to get a gist of it, so you just say that and it gives you this entire information. ”

These queries would be formulated so that they could get a direct, specific answer as opposed to a list of search results. None of the participants from the lower socio-economic group reported using this sort of searches, and would rather use general keyword searches for finding content, such as “Bengali remixes”. This, however, did not mean that they were unaware of how to phrase themselves to find the results they were interested in:

Interviewer: “One time you show me?”

Gopal: [clicks microphone, speaks:] “ ‘Bengali remix’. [pause] When searching for bangla, you have to say ‘Bengali’. If you search for ‘bangla’, then it won’t give you anything, it won’t happen. But if you search Bengali then it can come.”

Another common pattern was to use voice searches to look up map locations and directions, especially helpful for navigating in the large metropolitan cities where this work was conducted:

Reema: “I commute by a two-wheeler. When one day on my way back home I got stuck because of traffic, my husband gave me directions on phone. I reached home crying as he was angry because I could not figure out a way on my own. Then, my children told me about Google Maps and that I just need to speak the location I want to go to.”

When it came to these type of lookups, however, we found that among women there were concerns about safety while searching for location and direction in public spaces:

Radha: “ . . . we search for place. In map, we don’t know about a place. So we use voice to search for it. We type but there will be a mistake. Sometimes it doesn’t give right results. But in voice, it will show all the results like where to go, how to go, where is traffic, all that it will tell. But like while traveling with a friend, we used it. But I won’t use it when alone. People might hear what I am searching and might start following me.”

Unlike, for example, when asking a friend for directions over the phone, the VUIs required the user to provide a detailed description of the address they were searching for. Furthermore, for a person listening in, the entire conversation would be audible - again unlike normal phone conversations.

These kind of concerns were not reported by any of the men we interviewed. Here, the material property of voice searches - being publicly audible - combined with gendered notions of security in public spaces to shape the way in which women enacted the looking up practice. Furthermore, socio-economic position - which is also tied to smartphone literacy and tech-savviness - show differences in enactment of this practice. For instance, even though all our respondents had access to the same features on their phones, only our respondents from premier colleges and from professional backgrounds used the built-in assistant functionality to make requests or demands as opposed to generic searches.

4.2 The Learning Technology-in-Practice

Another practice of voice search use is what we would term as learning. We found three kinds of voice interface use which we classify as learning - searching for spelling of words, learning meanings of words, and finally finding encyclopedic content. These are distinct from looking up, as our respondents clearly identified these uses as motivated by a purpose of learning. Spelling was a common use of voice search for all groups but especially among the lower socio-economic group:

Reema: “Spelling mistakes like venter (winter), I will pronounce it correctly but spelling is wrong. So when I speak, it understands and gives me the correct spelling. Like terrace, I don’t know the spelling.”

Sahana: “I search for videos, movies, English words for Hindi words [meaning transliterations of Hindi words]. I use it at home as well as my workplace. Sometimes I even ask my mallik’s[6] children if I am stuck and need help with something. For example, I asked bhaiya [the mallik’s son] about how to download an app I didn’t know the spelling of but knew how to call its name. He told me about the audio feature by which I can simply speak.”

This was not limited to English, and several participants reported being able to use voice search to identify spellings in Hindi, despite not having made the specific setting changes to enable such support in the VUIs:

Nidhi: “For us, for me, it has been a good thing as we as educated people we are able to use typing and voice equally well ... Like we don’t have the habit of typing in Hindi and if we type something in Hindi, Google doesn’t understand it so well. But if we use voice to

[6]Mallik denotes “boss”, “owner” (of car in case of drivers) or “manager”.

speak in Hindi, maybe it will be able to understand nicely."

Interviewer: "So all the searches that you have done using voice, it has been in English but was it anything related to Hindi or any other language? Like searching for a Hindi song."

Nidhi: "Yeah, that I have done, like I have searched for song, the lyrics in Hindi and it has been able to identify that correctly."

Contrary to what we might assume, people with inadequate training in English were more keen to find spellings in English as it served their expectation to improve their reading or writing abilities in the language; those from a higher socio-economic position (who had "English medium" education) were to a greater degree interested in vernacular spelling.

Another specific pattern of learning that we found amongst the higher socio-economic group was using the voice user interface to identify and learn meanings of words that they may have heard but did not feel comfortable asking the meanings of:

Interviewer: "OK, oh so you don't use Google Assistant to search for...?"

Ishita: "... I use it but very rare. Like some swear words, I don't know the meaning but all my friends know. I don't want to do it before them. So I go to the room, then I do it. I am, like, lets see what this word means. So I didn't know the meaning of "slut", like sounds fancy let's see what it means. [laughter]"

At a young age it can be socially awkward to reveal your ignorance of commonly used expressions in front of your peers, but using voice search she could - in private - search for the meaning without even knowing how it was spelled.

Respondents would also make specific searches for dictionary, news or encyclopedia entries:

Nidhi: "So in my room whenever I am feeling like knowing something new and I don't want to waste my time doing time pass and I want to utilise it nicely, I do search or whenever I am getting bored and I want to read about new things, that time I generally use it."

In case of these three types of queries, again we saw clear differences between socio-economic groups. As mentioned earlier, the lower socio-economic group would mainly use the voice user interface for finding spellings of words. This pattern of use was also present among the higher socio-economic group but in their case it was almost exclusively used for spelling in vernacular languages. Searches for dictionary definitions, encyclopedic information and news items were also exclusively done by the higher socio-economic group.

4.2.1 The Leisure Technology-in-Practice

A very common use of smartphones was for leisure and accordingly this was also an important use for voice user interfaces. The leisure content that our respondents would often use voice interface to find were forms of music numbers or clips from regional or Bollywood movies. Irrespective of their mother tongue all our respondents were interested in Bollywood movies. However, these would have Hindi titles, and all groups of users - especially if they were not comfortable in Hindi - would use voice search to find Hindi content. In general, our participants would know how to pronounce - but not spell - the title of the movie and thus voice search would allow them to find the content they were interested in:

Interviewer: "[Since he said he can't write in Hindi] . . . so when you want to find Hindi songs, on YouTube, how do you do it?"

Gopal: "I speak with voice, I speak with voice, then it comes."

Interviewer: "Show me?"

Gopal: [Clicks the microphone icon] "Dil tu hi bata, piece 3" [name of a popular song]

While Hindi was the most commonly used vernacular language for this purpose, some respondents also used voice search to find entertainment content in other regional languages:

Rajesh: " . . . I don't need to speak in Kannada to watch a Kannada movie. I just speak [the title of the movie] . . . and it gives me the results. I choose anyone which I like."

For the lower socio-economic group, leisure was one of their main uses for their smartphones overall, and in particular YouTube was the primary, often only, way in which they consumed entertainment content on their phones. For the higher socio-economic group, however, the voice functionality itself became a source of leisure:

Anupriya: "Oh and we just use it for some fun thing then, we just ask Siri "who's your girlfriend", "who's your boss", "where do you work?", "are you a girl? Or boy? Or transgender?"... We do everything, if we are bored, we keep this Siri or Google Assistant, Android vs iPhone. I will be with my friend, and we'll ask the same question [of Siri and Google Assistant] and see whose answer is more funny or interesting."

This further illustrates the general differences between the two groups in the way they approach the voice user

interface. Even if they both accessed it through the same interface on their phone,[7] the higher socio-economic group would ask questions and give commands, while the lower socio-economic group would use more general search queries for content.

This kind of use shows a very distinct attitude towards technology - enacted through users' social position and living environment. One of the groups of our respondents is steeped in an environment where digital technologies are abundant and have been for quite some time. Hence this set of users are more at ease while interacting with technology. For them the smartphone is not just an instrumental tool for entertainment or communication but has become an integral part of their everyday lives to fool around with. This stands in contrast to the other set of participants for whom the smartphone and fast mobile Internet connectivity is only a relatively recent entrant in their social milieu. For them, they are still in the process of getting to know a new way of interacting with their phones which was hitherto unknown to them. The generic nature of their search queries and their non-use of assistant commands are testimony to their lack of long-drawn experience of technology which in turns reflects in the way they enact voice user interfaces.

5 A Framework of Enactment of VUIs

In this section, we draw on the enacted technologies-in-practice and place them in a framework organised around the material facilities of the technology that the users draw on, the norms that become involved in this enactment and the interpretive schemes that users apply to their enactment.

5.1 Facilities

The use of voice user interfaces among our participants was almost exclusively through smartphones. On our respondents' smartphones, there are three distinct interfaces providing voice functionality - "voice search" (Fig. 2),"voice assistant" (Fig. 3), and "voice typing" (Fig. 4). The voice assistant would be accessed either via a specialised app, via pushing and holding a specific button or via voice activation ("Hey, Siri", "OK, Google"). Voice search was available through the assistant interface or - primarily in the case of Android - through either the search bar in the default main screen interface or in apps such as YouTube. Voice typing or dictation would be present across the interface through a button next to the on-screen keyboard. We found that most of our respondents' used voice search and voice assistant on smartphones. While among the higher socio-economic

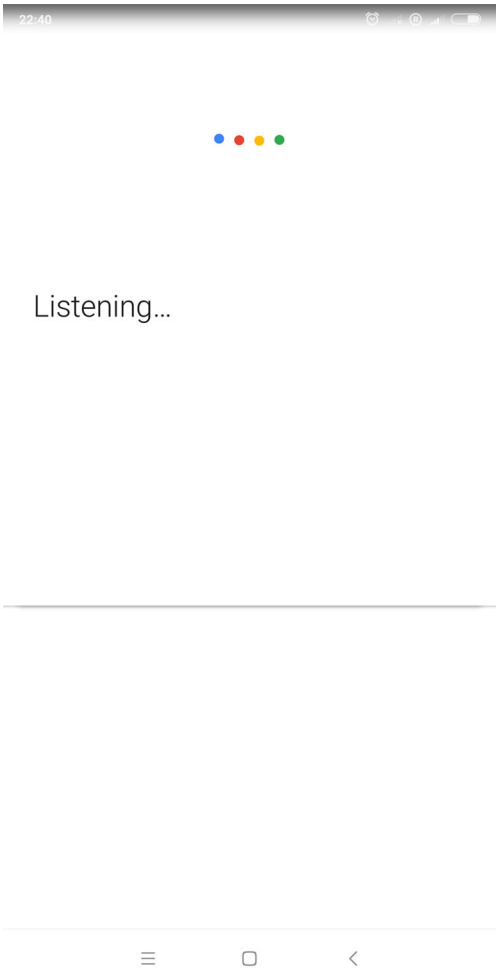

Fig. 2 The "voice search" interface which was the primary voice use case by our respondents

group dedicated hardware devices such as Google Home and Amazon Echo have started to be used, they were the exception. Thus, we have not considered such use in detail in this paper.

From our participants, mainly those who used Apple's iOS and its built in assistant Siri identified the above mentioned difference between using voice search and voice assistants. For Google Android users - who formed the majority of the participants in our study - voice functionality was primarily used through the voice search interface. Thus, they would access the voice functionality by pressing the microphone icon and speaking. Even if they used the built in "assistant commands" , they would do so through the "voice search" interface. However, such use of "assistant commands" - regardless of interface - were limited to respondents from higher socio-economic backgrounds. None of our participants had installed the separate Google Assistant app and only a small number would access the voice assistant interface directly. In the case of iOS, voice

[7]For instance the microphone button next to the search field on their home screens

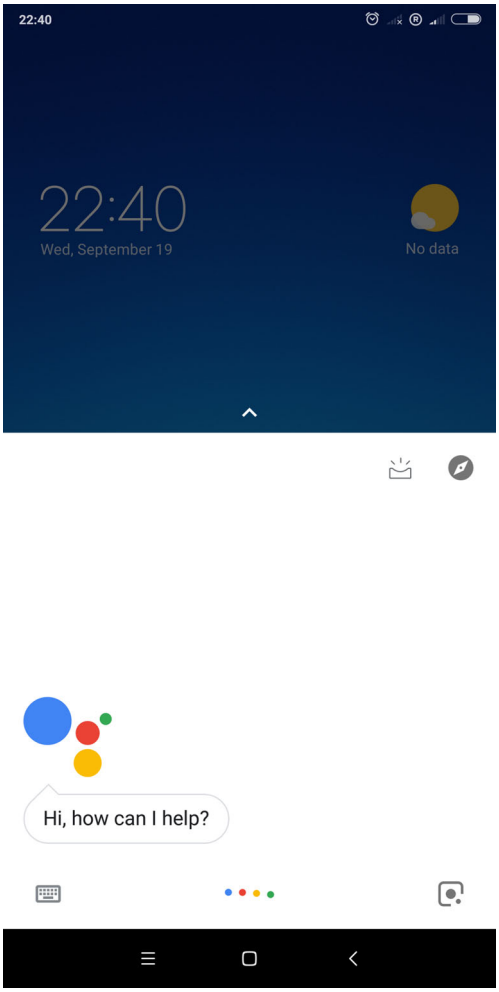

Fig. 3 The "Google Assistant" interface

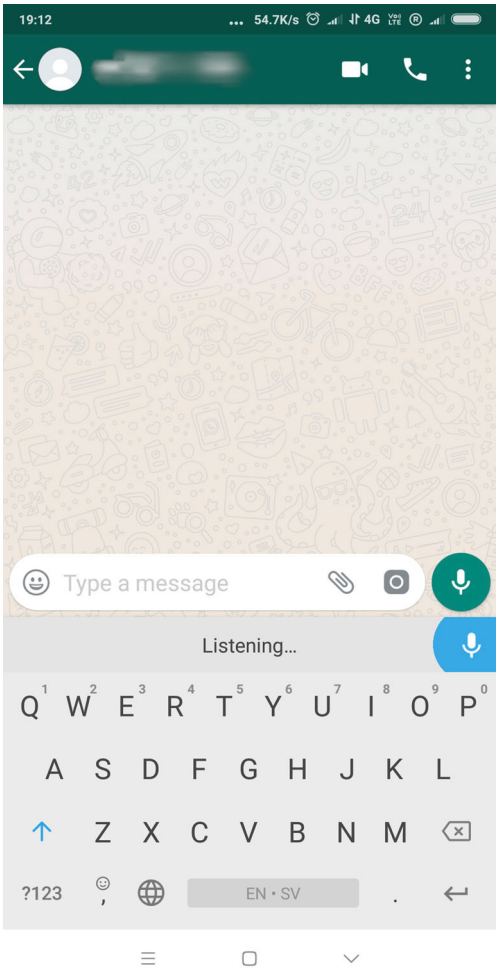

Fig. 4 "Voice typing" in the WhatsApp interface

features were only accessed through either a long press button press or - more commonly - by voice activation ("Hey Siri"). Voice activation was also used for Android voice use, in the form of "OK, Google", however again only among those of the higher socio-economic group. Almost all respondents from lower socio-economic backgrounds had never used the "assistant interface" or "assistant commands" via the "voice search" interface. These differences in the approach and understanding of the way these assistant operate highlight important differences in the technological affordances these platforms provide. Part of this difference in perception of the VUIs may be due to variations in digital literacy between the two groups. However, an important aspect is that the content and services currently available on the "assistant" are still targeted at relatively affluent users. This means that the technological affordances provided by the platform for the lower socio-economic group is limited to its use as a text-to-speech or voice recognition engine. While voice assistants on smartphones (and even independent hardware devices) are now broadly available at price points where lower socio-economic groups can afford to own and use them, in the way they are presented and imagined, the assistants are still primarily there to service the higher socio-economic group.

Apart from using voice user interfaces through dedicated access points as built in to the phone operating system, participants also used it in individual apps. By far the most common individual apps where voice interfaces were used was in YouTube and WhatsApp and this broadly correspond to the apps our respondents used the most on their phones. Other places where voice interfaces were used was in the Maps app as well as on the App Store.

While Apple's voice assistant does not support any Indian languages, the voice interface provided by Google provides multilingual support for many of the regional languages spoken by our participants - including Hindi, Bengali, Tamil and Kannada. To access this support, however, the user would need to specifically enable it in their settings for voice search and voice assistant, something that none of our participants had done. What we see

here is that - in our respondents' perception - the voice user interface should be able to interpret the type of mixed language that they use in everyday speech without additional settings or customisation. This "English" - text written primarily in English with any vernacular in Latin characters as opposed to Indic alphabets - was the language of the smartphone:

> Rakesh: "I only type in English. I don't search using voice. Very rare. I directly type. Bengali is our language to talk but not when searching on Google."

An important material difference between VUIs and other voice-based modalities on the smartphone – such as regular voice calls – is that when using the VUI the entire conversation is audible to those around you. Furthermore, the VUI still lacks a complete understanding of the user's context, meaning that commands or queries need to be spelled out in full. Unlike - for example - asking for directions from a friend over the phone, queries such as "how can I get to your house" need to be translated into specific requests such as "how do I get to Koramangala 7th block".

Through our interviews it is clear that the facilities provided by smartphone VUIs - voice search, voice typing and messaging and voice assistants - interact with social structures - such as socio-economic position, digital and general literacy, gender - to shape the enactment of the practices of looking up, learning and leisure.

5.2 Norms

From our analysis of how the material structures of VUIs has become enacted through everyday practices, we can begin to see how the technology when viewed through a technologies-in-practice lens, becomes entangled with users' identities, social position and cultural norms. In this section, we further illustrate the way in which these entanglements take place and in what way they shape the relationship between voice user interfaces and its users.

5.2.1 A "mixed" Linguistic Culture

As has been noted by others (Karusala et al. 2018; Microsoft Research India 2018) code-switching and code-mixing is a common practice among Indian language speakers and forms a distinct linguistic culture. In fact speaking in "mixed" language was the "normal" way of speaking for all our respondents. The linguistic complexity of the Indian subcontinent meant both limitations of as well as unique affordances provided by voice user interfaces for our respondents. This becomes especially important in metropolitan cities where there are diverse regional and linguistic communities - originating from historic or more recent migration from other parts of the country. English as the language of aspiration and respect (Karusala et al. 2018) in combination with Hindi as the dominant vernacular provide for a complex - often code-mixed - lingua franca for urban India. In addition to this, our respondents would use - and consume content in - not just their own linguistic community's vernacular but also their metropolitan region's majority language.[8]

> Reema: "I speak in normal language - mix. More of Hindi words than English. It understands properly. But the results are mostly in English. If I say Hindi shayari, then only the results would be in Hindi."

5.2.2 Cultural Norms Around Education and Literacy

One of the expected benefits of voice user interfaces in developing countries - where multilingualism, illiteracy and low levels of literacy are pervasive - is that it can provide opportunities for digital inclusion (Turovsky 2017). But as others (Karusala et al. 2018) have already noted literacy is not just a matter of the ability to read and write. It is imbued with complex social and cultural meanings (Spring 2007). In the Indian context, English is often associated with class and educational status. The choice of whether to use English or vernacular is a negotiation between different factors such as where you are speaking, about what you are speaking and with and amongst whom you are speaking. These concerns carry over into our respondents' use of voice user interfaces on their smartphones. For example, for several of our respondents among the lower socio-economic group, there was a clear link between the use of voice search and low literacy or low education.

5.2.3 Privacy, Safety & Moral Codes

As previous studies (Jones et al. 2017) indicated, privacy becomes a major concern while using voice-based interactive technologies in public spaces. When asked about the environments in which they prefer to use voice search, both men and women said they use it mostly in their personal space.

Our male respondents expressed less qualms than women about using voice search irrespective of space, viewing it more as a problem of whether it was quiet enough for the voice interface to pick up what they were saying. Our female respondents, on the other hand, expressed an acute awareness of their surroundings at all times. They were constantly making choices about what can be searched for where, including whether to use voice search at all.

[8] In our case Kannada, Bengali and Hindi.

This constant act of self-censoring while using voice search was for women also driven by broader moral codes. For example, consumption of pornographic material by women is a social taboo in India. Access to Internet via a personal device may have eased women's access to such material. While several of the women we interviewed did in fact report that they also used voice search for this purpose, the moral codes relating to this kind of content shaped the spaces where women felt comfortable using voice interface. Again, the public nature of voice interfaces and the potential for being overheard while using them, set limitations on when and where the women interviewed felt comfortable using it for this purpose:

> Interviewer: "Is there any difference for what you use it at home, for what all searches kind of searches that you do at home?"
>
> Anupriya: "In home, we don't do all the searches what we do [laugh] here right? Only some . . . I mean . . . One category [implying sex] of searches and all we won't do at home, right?"
>
> Interviewer: "What all are you comfortable searching using voice? When you are in your room, when you are with your friends, vs. when you are at home?"
>
> Nidhi: "When I am at home, I don't want to search anything related to, eh, sex. Porn. When I am with my friends also 99% I don't use."
>
> SSSInterviewer: "You don't use it for this purpose or you don't use it at all?"
>
> Nidhi: "I don't use it for_that_ purpose. And some terrorist, killing, violent anger, I don't use it [for these kinds of searches] at home, my parents [would] kind of freak out."
>
> Interviewer: "Really?" [Laugh]
>
> Nidhi: "I don't like to use it before others also, because I don't want to them to think I am a psychopath.''

While these kind of restrictions may also hold true for men, they did not report such a specific awareness of the type of content that could not be searched for among friends, family or in public spaces. In this instance, this could be because the social taboo for consumption of pornographic content is significantly less for men.

5.3 Interpretive Scheme

5.3.1 Speaking in "mixed" Language

While our respondents' use of their smartphones and voice search was primarily oriented towards English, they would freely mix in words or sentences in vernacular and would expect the VUI to accurately interpret this:

> Interviewer: "So in which language have you searched?"
>
> Tripti: "So I have searched in English and in Hindi.. In Hindi it becomes - maybe it is because of the accent or something - it sometimes become difficult for Google to identify correct word. Initially when we used to use [voice search] 2-3 years back, it wasn't that good, but these days [Google voice search] is able to catch a lot of words [in Hindi]. Maybe one in fifty words it is not able [to] catch properly. Not one in fifty, but one in ten."

Code-switching was especially common when searching for entertainment content as much of it would have titles in Hindi. Here, voice user interfaces were not just a convenience, but necessary for locating the content in the first place as they may know the name of the movie or the actor but not how to spell it. Even though some of our respondents were aware that the phone had support for languages other than English, none of them had enabled it specifically in the settings. Rather they relied on the voice user interface to understand their mix of vernacular and English while technically having the interface set to "English".

However, when it came to accessing vernacular content - apart from Hindi - the voice user interface would not be able to pick up their code-mixing as accurately. This may, in part, be why our respondents' preferred generic search queries such as "Kannada songs" to find vernacular content:

> Sunil: [speaks to his device:] "Kannada songs" [gets a list of results]
>
> Interviewer: "Why don't you use Kannada to search?"
>
> Sunil: "When I can search for anything in English and Hindi, why should I speak in Kannada?"
>
> Interviewer: "And what about Bangla?"
>
> Swarup: "If I speak in Bangla, then over there you can't see any Bangla [referring to the search box]. Over there if you speak in English, then I can find Bangla [material]."
>
> Praveen: "I am from Assamese [speaking background], so we have difficulty in speaking [to the voice user interface]... because of our accent, it was not able to understand."

Regardless of these limitations, multilingualism and code mixing was not just present throughout our respondents' use of voice user interfaces but was also an important reason - and interpretive scheme - for their use of this functionality

in the first place. While most of them followed the norm of mixing languages, their decision to leverage code-switching on VUIs depended on their specific purpose of use, and their individual ability to make use of this affordance.

5.3.2 Feelings of Apprehension and Anxiety in VUI Use

For some of our participants - primarily men - use of voice search was associated with feelings of shame or shyness due to their culturally contingent connection between education, literacy and the VUI use. This gendered difference is illustrated by this conversation between a male and female respondent during a group interview:

> Deep: "I feel very ashamed, when I [have] learn[t] these things [reading and writing], but I can't actually type so I have to say it out loud, so I feel very bad about myself. I feel shy of course."
>
> Jyoti: "But typing takes a lot of time also. With voice it goes fast. Therefore it is good. It is not possible to remember all the spellings."
>
> Deep: "If I am studying, why won't I be able to write? This is very shameful thing."
>
> Jyoti: [shrugs] "I actually say it quickly and it is done."

We observed this pattern in multiple contexts - where men in the lower socio-economic group expressed a need to hide or not disclose the fact that they were using voice search. This need to hide their use of voice search would influence the context where our respondents could make use of the technology. As Gopal related, he would feel comfortable using voice search in his own neighbourhood as those who lived there knew him and knew about the fact that he had not completed school. When he left the neighbourhood, however, he would become self-conscious about using voice search for fear of judgement:

> Interviewer: "... any time when you use [voice search] do you feel shy?"
>
> Gopal: [embarrassed laugh] "I can't use [voice search] in every place."
>
> Interviewer: "You can't use [voice search] in every place, why not?"
>
> Gopal: "What you say, everybody can hear. They may know I haven't studied, that I don't know [how to read and write]. So people may make fun."
>
> Interviewer: [I see], "...for this reason. When you are outside?"
>
> Gopal: "Yes, well when you are outside. Meaning, when you are [in the local area - his zone of everyday familiarity], like at Haldar more, up to there it is okay to use [voice search]. People know [I haven't studied]. But when I go further away, then I won't use it."
>
> Interviewer: "So it is okay up to Haldar more?"
>
> Gopal: [laughing] "Yes, up to Haldar more it is okay. But when I go further away, then I won't use it."

5.3.3 VUIs Making up for my Limited Capabilities

For several of our female respondents, voice search could be a convenient way to make up for their limited ability to write or spell. Reema related how voice search helped her to not have to disclose the fact that she could not spell very well:

> Reema: "Sometimes people have big names, and I do spelling mistakes, so only for that mostly I speak. For example, Vikas sir checks whatever we write in the visitor book and if he sees a spelling mistake, I would feel so ashamed. So for that not to happen, I speak."

For the higher socio-economic group - most of whom were educated in English - voice search was not a trade off between speaking and typing. As they would not regularly write in vernacular, some would use voice search to find spellings of words in, for instance, Hindi. Confidence in their English education, and the status associated with it in the Indian context, meant this usage did not have any connotations of shame or sense of inadequacy. However, even for the higher socio-economic group the potential for embarrassment or need for secrecy about their voice interface use would shape where they felt comfortable using it:

> Nidhi: " ...I sometimes [voice search] lame things. Like very obvious meanings of few words."
>
> Interviewer: "Yeah?"
>
> Nidhi: "Like sometimes I Google the word "lame" also just to see what it means. It is not like I don't know the word...I just want to know how a dictionary would explain it. So I just don't [laugh] want others to know that. "

5.3.4 Where Can I Talk to my Phone?

As we have discussed norms of privacy were crucial in shaping use of VUIs. We further saw that there are clear differences in the way that users from different genders and socio-economic groups interpreted these norms when enacting VUIs. Men's desire for privacy was motivated by two factors. For the lower socio-economic group it was related to their concerns about notions around being

educated and associated social implications. For the other set of men, where education was not an issue, it was more about a social performance. They wanted to be correct in their manner of speaking and also they did not want to be perceived as "a strange guy talking to his phone":

> Interviewer: "So what environment do you use it in, like, generally? Like is it in your room or are you okay using it, like in public, or?"
>
> Anupriya: "Yeah, in even public it is fine. But [eh] you see, some of people, like some of my friends,they feel very shy to you, you see, not really [uh] using English and all .... So I generally use it in public [I use it] both [in public and in private]."
>
> Interviewer: "So you don't have any concern of speaking or you don't feel shy or awkward?"
>
> Anupriya: "No, not at all. But my friends, some of my friends feel shy."
>
> Interviewer: "So they don't use it in public?"
>
> Anupriya: "Not really ... you see, many people are worried you know, they are worried with their English grammar and all, so they feel shy talking in public [using voice interfaces]."

While women have also expressed this sense of embarrassment around the use of technology they were more concerned about other's reaction to or moral judgements of their searches. This tendency was visible among women across class, age, and educational status. They did not feel comfortable using voice search either in public spaces or in places where they could be easily heard by others, such as the workplace:

> Shikha: " ...I don't use it [voice search] in public spaces ...when in public spaces people can use their mind and do something wrong. [So], I use it secretly mostly."
>
> Interviewer: "What would be the main reason that you are not comfortable using voice in public, but you are comfortable using your phone by typing and all?"
>
> Tanya: "Just that the thing that other people would get to know what I am doing. Everything you cannot disclose to everyone. And there are like people so ...sometimes it becomes even to your close friends or your parents or ... maybe very, very close ones also. That you can't even say this stuff."

As we have previously mentioned, this difference between men and women was clear also in the practice of using look ups for map locations. While men expressed no hesitation to using voice for this purpose, for women there were concerns about their safety when disclosing their destination via voice search. This self-awareness manifested itself in many different ways in our conversations with female respondents. For example:

> Sudha:". . . Now here if you see, I can't search for anything related to Hindu culture because then people would judge me based on what I am speaking and why." [It's a Muslim-dominated locality.]
>
> Tanya: "I could say that.. using it for private stuffs it is not good, because if you are surrounded by people you can't just say it, that way."
>
> Interviewer: "so how would you call private, what private stuff?"
>
> Tanya: "I ...[uh] what kind of private stuff, like okay [uhm], [silence] what should I say ...okay I say something that you are not able to disclose to everyone that you are actually searching for such things, so ... if you are like [clicks tongue] for the funny purpose also, it sometimes become weird to search in front of people, even for like if you are normally using it for normal searches also, so you just can't speak it in front of public, because obviously people do have ears, they would listen to you while you are doing this thing, so it becomes, that time you can't just do it that way."

There were differences between women from different socio-economic groups in the content that they used VUIs for. For instance, women from lower socio-economic backgrounds did not report any "adventurous" use of voice interfaces. One reason for this could be the concept of and access to private space. Some of the younger women in this group, however, did allude to such use but were not comfortable talking about it. Regardless, for these women it was not just about where to use voice interfaces but also what time of the day they had a comfortable space in which to do it:

> Sudesna: "Using it in the house is easier. Mummy and pappi are not there, they are at the shop [where they work]. So in the house, we can use it. It is more calm to use in the house [when] no one is there. Outside, there are others who could hear. For example, some autorickshaw driver is there ..."

In sum, we see that the time and space people choose to use voice interfaces in are interwoven with many levels of considerations ranging from linguistic practices, cultural notions of literacy and education to concerns for safety and moral judgements. The way in which these concerns are enacted in everyday practices around voice interfaces is inextricably driven by users' social position such as gender identity, class, living conditions and education.

6 Discussion

Instead of looking at why and how VUIs are appropriated by diverse users in different circumstances, enactment as a lens enables us to focus on people's engagement and the particular emergent structures of using the VUIs as technologies-in-practice. In our frame of analysis, VUIs are not merely technological artifacts but are considered to be enacted in three technologies-in practice, namely looking up, learning and leisure. These enactments of ongoing situated use of VUIs emerges from an interplay between social and technological structures and people's interpretations of them.

Orlikowski (2000) identifies three kinds of enactments - inertia, application and change - as a way to categorise technology use. In all three categories, people use technology in relation to their existing practice either to retain them (*inertia*), or to refine and enhance them (*application*) or to substantially change and alter them (*change*). Here we discuss the three technologies-in-practice that we identified in relation to this categorisation and what such categorisation implies for researchers and practitioners of VUIs .

6.1 Looking Up

When it comes to the Looking up Technology-In-Practice (Fig. 5) we see a clear distinction in the enactment of VUIs between the higher and lower socio-economic groups. For the higher socio-economic group, we classify this enactment as inertia, where they are applying the VUIs to retain an existing practice of using their smartphone for informational queries and directions. They would use speech and text interchangeably for this purpose, without specifying any strong preference. For users from the lower socio-economic group, however, their enactment of this technology-in-practice was - especially for people with limited literacy - one of change. The enactment of a Looking Up Technology-in-Practice provided a substantially new way of accessing information and navigating the city. We observed one exception within both the socio-economic groups, where gendered structures and norms of safety and privacy made women from both the groups interpret the technology-in-practice as a potential risk. The inertia and selective non-use of the Looking up Technology-in-Practice by women observed in our study expands on earlier studies that have

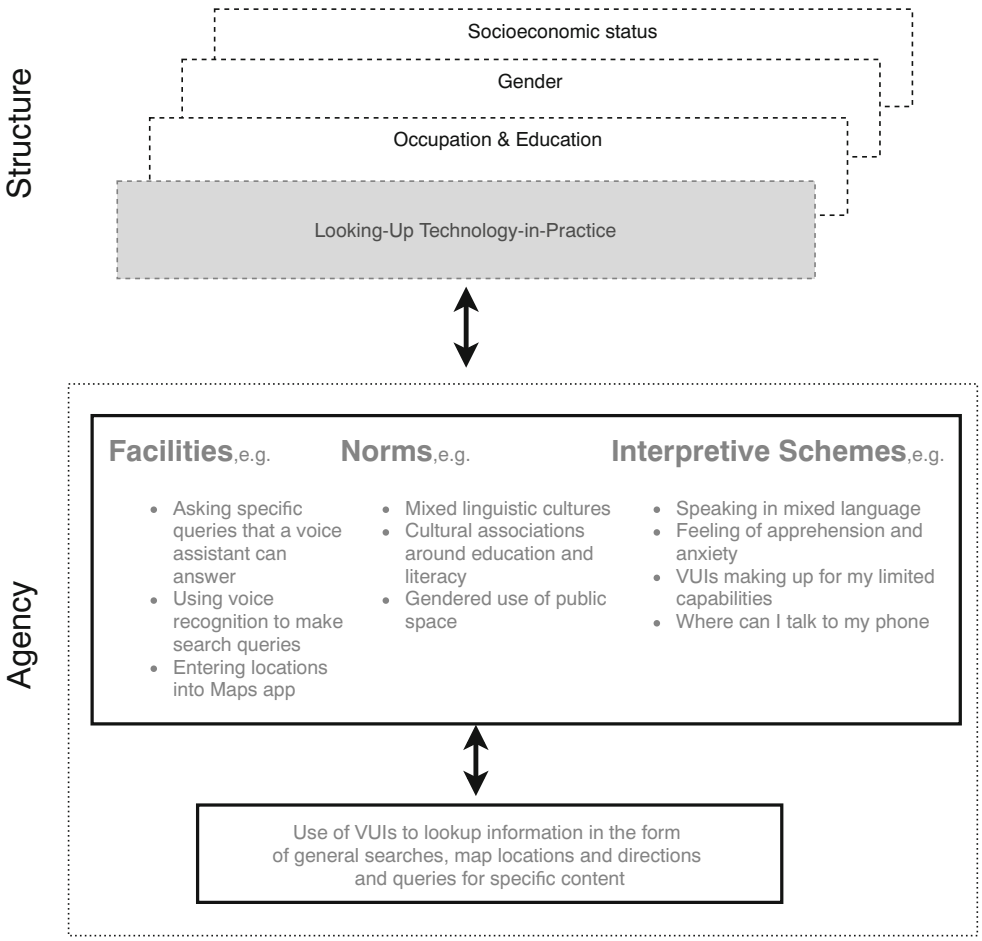

Fig. 5 The Looking Up Technology-in-Practice

highlighted the privacy implications of voice user interfaces in public spaces (Robinson et al. 2018; Easwara Moorthy and Vu 2015). What we find is the complex interaction between users' choices and preferences, social and cultural norms of behaviour as well as a layered understanding of space of use. Here, unlike earlier literature on technology appropriation, we argue that space needs to be viewed as a continuum rather than a binary of public and private. Who you are with or surrounded by, what information you are disclosing and your perception of risk in any given situation will shape how you look up using VUIs. This understanding of enactment of VUIs leads to the critical question of how to create a shared informational context between the user and the interface, that allows the VUIs to be sensitive to the users' situation and accordingly adapt the modalities of the interface.

6.2 Learning

The Learning Technology-in-Practice (Fig. 6) enacted by our respondents can be viewed as ranging from application to change. It enhances our respondents' ability to find meanings and spellings of words in more than one language. While they have other means of accessing this information, the facility of only having to pronounce (albeit sometimes incorrectly) the word, improved their ability to learn across different levels of literacy and language competency. For people from the higher socio-economic group, this was primarily used to overcome limitations related to their knowledge of vernacular language. This worked in two ways. Most users in this group had been educated in English medium and therefore did not use their knowledge of vernacular language regularly. With the Learning Technology-in-Practice they could find meanings and spellings they may have forgotten or were uncertain about. This also applied to those who had moved to other regions of India where they had limited familiarity with the local language. Hence, their enactment of VUIs can be seen as one of application, where users' engagement with the technological structures enhance their existing practices of learning.

For the lower socio-economic group, learning English was an aspiration which they could easily explore with their engagement in this technology-in-practice. For this group, it was also a way to overcome issues of limited or low literacy. This meant, there was significant change in their

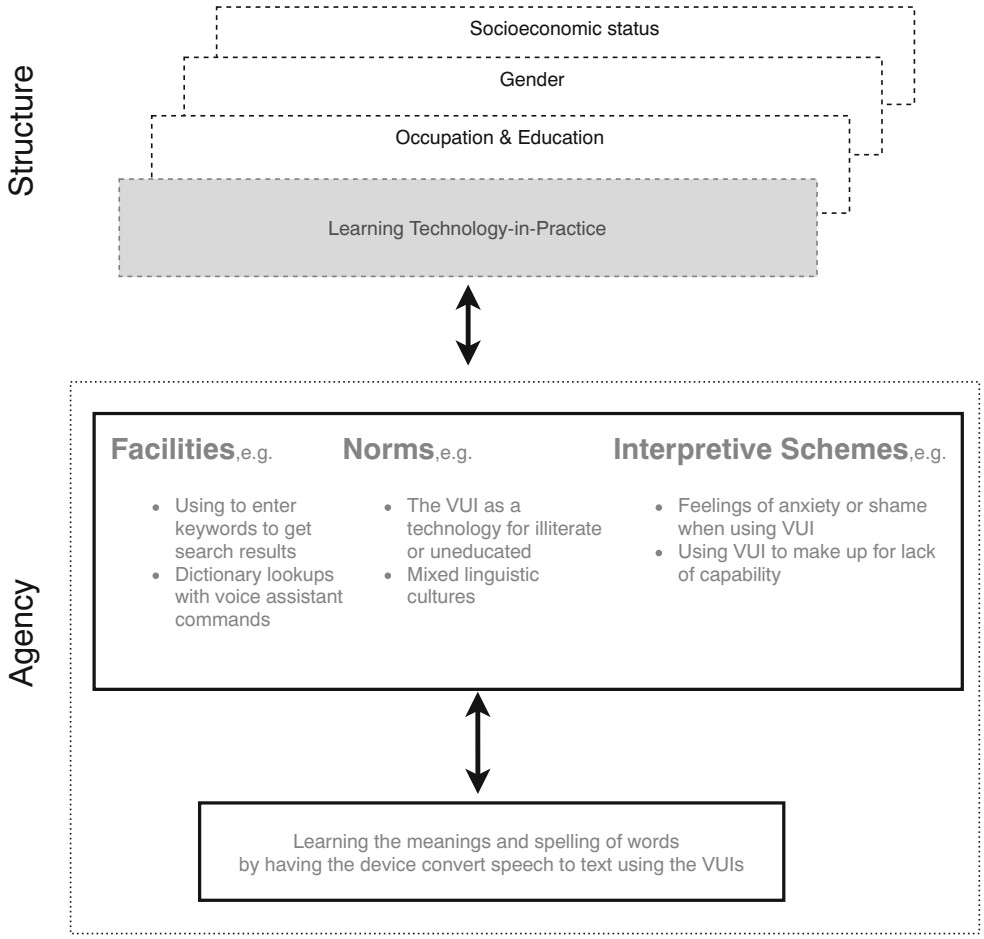

Fig. 6 The learning technology-in-practice

existing practices of learning through their enactment of the Learning Technology-in-Practice. However, this association to low literacy also lead to patterns of non-use due to social embarrassment. This pattern, as we have noted, was gendered. While women with low literacy saw it as a useful affordance of technology, men within the same category found its social affordance much more limited. They associated the facility with a sense of social stigma attached to illiteracy.

While introducing VUIs in a development context, a common pattern is to think of them as providing a means towards learning, literacy and digital inclusion (Turovsky 2017). However, this framing of the affordances of VUIs risks strengthening certain existing apprehensions and anxieties, e.g. literacy and class in the above case, leading to discomfort around the use of VUIs or choices to not use them at all.

6.3 Leisure

The enactment of the Leisure Technology-in-Practice can be categorised as one of application rather than change or inertia. Consumption of entertainment content - such as YouTube videos - was one of the core features our respondents used their phones for. This did not change through their use of VUIs. However, their practice of leisure was enhanced when they could easily access content across language divides.

Among the higher socio-economic group, a change in practice that we observed was the ability of using the VUI itself as a source of entertainment - by asking it lewd questions, attempting to confuse it, having small conversations with it or requesting jokes and other information.

While the enactment of leisure in itself is unsurprising, the complex linguistic milieu of the users and their negotiation of it with help of VUIs (in our case - when finding content) poses a critical challenge for further research (Microsoft Research India 2018). This is relevant both in terms of the development of the technological affordances of VUIs as well as in thinking about what they could be designed to do. For example, language choice cannot be encompassed in a single setting but must take into account code mixing, code switching and other complex linguistic practices (Fig. 7).

In all three technologies-in-practice, we see their kinds of enactment varied across users agency, identity and their

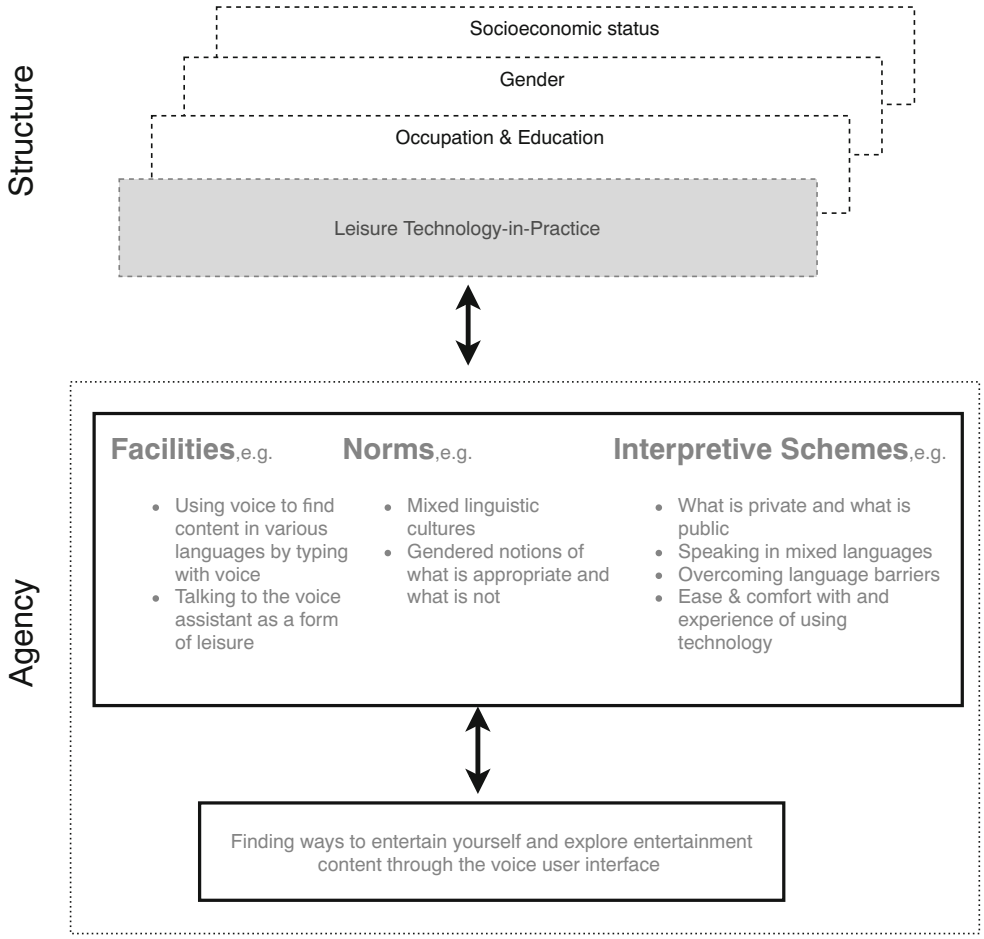

Fig. 7 The leisure technology-in-practice

interpretations of technological and social affordances of the VUIs. While earlier studies on VUIs among specific groups of users has identified similar usage patterns, our objective in this paper is to capture the processes of enactment through which such usage patterns emerged. We argue that understanding the recursive relationship between users, technology and social structures that makes VUIs a socially embedded experience is crucial for understanding the emerging usage patterns of VUIs in a more nuanced way.

6.4 Limitations

This is an exploratory study which follows a qualitative method and interpretive epistemology. Hence, our sample selection was purposive, where we aimed to enrich both diversity of users profile and their contexts of use. We expanded our sample size through a snowball method where each participant led to another participant in their respective network. As we started to get similar responses from every new interview we decided to stop the process. The motivation of our study was to identify specific patterns of use of VUIs and how we map these patterns of use across different profile and context of users. Hence, we were able to document a variety of uses even with a relatively small sample size.

However, we understand that with a larger number of participants enacting VUIs, we will be able to find more technologies-in-practice. Therefore, it is important to reiterate that our findings of three technologies-in-practice are only indicative and neither conclusive nor exhaustive. This implies, firstly, that our way of labelling technologies-in-practice is not the only way to capture what we observed and secondly, that there might have been practices that were not observed through our study. While our first limitation is in general present in all interpretive studies, the second limitation can easily be addressed by engaging in more long-term ethnographic study of VUI use among the same groups of users or by extending the study to other categories of users, such as high income working professionals, elderly people and children.

Moreover, our study was based on users in urban areas, which also limits our findings. Partially, this is by design as we wanted to capture VUI use in heterogeneous settings in terms of socio-economic status, linguistic cultures, occupational categories and degrees of exposure to technology. It is also within these urban areas that sufficiently high-speed mobile Internet access and affordable smartphone devices have been available for a longer period of time. While this provides us with a rich sample, it limits our understanding of VUIs in low-resource settings. For example, how people in locations with poor technical infrastructure make use of VUIs.

Finally, our criteria of selection of respondents based on their regular use of VUIs, meant that most of our respondents were young people. We do not consider this as a major limitation as our motivation was to understand situated use of a technology. However, there is scope for further research that also looks at reasons behind non-use of VUIs.

7 Conclusion and Future Work

India is projected to be the second largest country in terms of total numbers smartphone users by the end of 2018, right after China (E.T. T 2017). Thus, understanding how diverse groups of Indian users use different modalities on their smartphones will be critical to further develop VUIs. In this paper, drawing on Orlikowski (1999) technologies-in-practice framework, we have presented a nuanced, contextual understanding of everyday use of voice user interfaces on smartphones among mostly young, urban users from diverse socio-economic backgrounds in three large cities in India.

Through qualitative semi-structured interviews with 28 respondents, we identify three technologies-in-practice (looking up, learning and leisure) that emerged through enactment of voice user interface on participants' smartphones. We show how each of these practices draw on the material structures of the interface, user preferences and ability to choose from the range of possible engagements with the interface. Last but not the least, we highlight the interaction between these practices and intersecting social structures of gender, occupation, education, class that the users inhabit at large. From our findings we contribute insight into how contextual complexities of multilingual practices, culturally embedded notions of literacy and education, gendered norms of social behaviour, notions of space and privacy are some of the major factors that shape the way people engage with and enact different features of voice user interfaces on their smartphones.

Here, we want to address another significant stream of studies that look at social affordances of platforms. These studies also address the complex entanglements that play out in the way different platforms position themselves strategically for diverse set of users across varied contexts. As Gillespie (Gillespie 2010) argues, the term platform is used to mitigate tensions arising from conflicting constituencies of intended partners, including users, advertisers,media producers, policymakers. Such studies unravel the politics of platforms by examining its discursive use to cater to conflicting interests converging on to a technological platform (Gillespie 2010; Helmond 2015; Gerlitz and Helmond 2013). While we acknowledge the significance of these studies to understand social affordances of platforms, the discursive politics of platforms remain beyond the scope of our work. Instead of looking at social

affordances emerging from structural and symbolic politics within the platforms, we focus on specific social practices around voice user interfaces as a technological platform that cater to diverse use contexts.

Given VUIs growing popularity among urban users, their platform-like structure and the emergent nature of their use, we solicit further research of ethnographic orientation across more diverse users in India and elsewhere. For example, in this paper we focused on social positions around categories of gender, occupation and education. We infer future work of similar approach to focus on issues of age, disability and caste among other social factors shaping users' agency to engage with voice user interfaces on smartphones and other devices.

References

Amazon Inc (2015). Introduction of Amazon echo. https://www.youtube.com/watch?v=6V5I8HHFTNQ.

Balasubramanian, K., Thamizoli, P., Umar, A., Kanwar, A. (2010). Using mobile phones to promote lifelong learning among rural women in Southern India. *Distance Education*, *31*(2), 193–209.

Bardzell, S. (2010). Feminist HCI : Taking stock and outlining an agenda for design. Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp 1301–1310. https://doi.org/10.1145/1753326.1753521.

Bardzell, S., & Bardzell, J. (2011). Towards a feminist HCI methodology: Social science, feminism, and HCI. SIGCHI Conference on Human Factors in Computing Systems (CHI'11), pp 675–684. https://doi.org/dc47ft.

Berg, A.J. (1999). A gendered socio-technical construction: the smart house, in "The Social Shaping of Technology"

Berg, A.J., & Lie, M. (1995). Feminism and constructivism: Do artifacts have gender? *Science, Technology, &, Human Values*, *20*(3), 332–351.

Best, M.L., & Maier, S. (2007). Gender, culture and ICT use in rural South India. *Gender, Technology and Development*, *11*(2), 137–155.

Biele, C., Jaskulska, A., Kopec, W., Kowalski, J., Skorupska, K., Zdrodowska, A. (2019). How might voice assistants raise our children? In *Advances in intelligent systems and computing*, (Vol. 903 pp. 162–167). Cham: Springer. https://doi.org/10.1007/978-3-030-11051-2_25. http://link.springer.com/10.1007/978-3-030-11051-2_25.

Brown, S.A. (2008). Household technology adoption, use, and impacts: Past, present, and future. *Information Systems Frontiers*, *10*(4), 397.

Cockburn, C. (1983). *Brothers*. Boulder: Westview Press Inc.

Cohen, M.H., Cohen, M.H., Giangola, J.P., Balogh, J. (2004). Voice user interface design. Addison-Wesley Professional, google-Books-ID: PI_n2EcJfT0C.

Cowan, B.R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., Bandeira, N. (2017). What can i help you with?:nfrequent users' experiences of intelligent personal assistants. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, p 43.

Cowan, R.S. (1983). More work for mother. Basic Books.

Dale, R. (2016). The return of the chatbots, (Vol. 22. https://doi.org/10.1017/s1351324916000243, arXiv:0910.0834.

Druga, S., Williams, R., Breazeal, C., Resnick, M. (2017). "hey google is it ok if i eat you?". In *Proceedings of the 2017 Conference on Interaction Design and Children* (pp. 595–600). https://doi.org/10.1145/3078072.3084330.

Easwara Moorthy, A., & Vu, K.P.L. (2015). Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, *31*(4), 307–335.

van Esch, D. (2017). Type less, talk more. https://www.blog.google/products/search/type-less-talk-more/.

E.T. T (2017). India to have 530mn smartphone users in 2018: Study. https://telecom.economictimes.indiatimes.com/news/india-to-have-530mn-smartphone-users-in-2018-study/61097817.

Etikan, I., Musa, S.A., Alkassim, R.S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1–4.

Faulkner, W. (2001). The technology question in feminism: a view from feminist technology studies. In *Women's studies international forum, Elsevier*, (Vol. 24 pp. 79–95).

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, *15*(8), 1348–1365.

Gillespie, T. (2010). The politics of 'platforms' *New Media & Society*, *12*(3), 347–364.

Haider, J. (2016). The shaping of environmental information in social media: Affordances and technologies of self-control. *Environmental Communication*, *10*(4), 473–491.

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media+ Society*, *1*(2), 2056305115603080.

Hsieh, Y.P. (2012). Online social networking skills: The social affordances approach to digital inequality. *First Monday 17* (4).

Humphreys, L. (2005). Reframing social groups, closure, and stabilization in the social construction of technology. *Social Epistemology*, *19*(2–3), 231–253.

Jones, M., Robinson, S., Pearson, J., Joshi, M., Raju, D., Mbogo, C.C., Wangari, S., Joshi, A., Cutrell, E., Harper, R. (2017). Beyond "yesterday's tomorrow": future-focused mobile interaction design by and for emergent users. *Personal and Ubiquitous Computing*, *21*(1), 157–171. https://doi.org/10.1007/s00779-016-0982-0. http://link.springer.com/10.1007/s00779-016-0982-0.

Karusala, N., Vishwanath, A., Vashistha, A., Kumar, S., Kumar, N. (2018). Only if you use English you will get to more things: Using Smartphones to Navigate Multilingualism. Proc of CHI pp 1–14, https://doi.org/10.1145/3173574.3174147. http://dl.acm.org/citation.cfm?doid=3173574.3174147.

Khandkar, S.H. (2009). Open coding. *University of Calgary*, *23*, 2009.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American journal of occupational therapy*, *45*(3), 214–222.

Kumar, N., Karusala, N., Seth, A., Patra, B. (2017). Usability, tested? *interactions*, *24*(4), 74–77. https://doi.org/10.1145/3098571. http://dl.acm.org/citation.cfm?doid=3115390.3098571.

Lau, J., Zimmerman, B., Schaub, F. (2018). Alexa, Are You Listening? *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–31. https://doi.org/10.1145/3274371.

Lee, K., Lee, K.Y., Sheehan, L. (2019). Hey alexa! a magic spell of social glue?: Sharing a smart voice assistant speaker and its impact on users' perception of group harmony. *Information Systems Frontiers*, 1–21.

Lovato, S., & Piper, A.M. (2015). Siri, is this you?: Understanding young children's interactions with voice input systems. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 335–338): ACM.

Luger, E., & Sellen, A. (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* (pp. 5286–5297): ACM.

Masika, R., & Bailur, S. (2015). Negotiating women's agency through ICTs: A comparative study of Uganda and India. *Gender, Technology and Development*, *19*(1), 43–69.

Mason, M. (2010). Sample size and saturation in phd studies using qualitative interviews. In: Forum qualitative Sozialforschung/Forum: qualitative social research, vol 11.

McGregor, M., & Tang, J.C. (2017). More to meetings: challenges in using speech-based technology to support meetings. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 2208–2220). https://doi.org/10.1145/2998181.2998335.

Medhi, I., Patnaik, S., Brunskill, E., Gautama, S.N., Thies, W., Toyama, K. (2011). Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction*, *18*(1), 1–28. https://doi.org/10.1145/1959022.1959024. http://portal.acm.org/citation.cfm?doid=1959022.1959024.

Microsoft Research India (2018). Project Mélange: Understanding MixEd LANguaGE and Code-mixing. https://www.microsoft.com/en-us/research/project/melange/.

Nguyen, H., Chib, A., Mahalingam, R. (2017). Mobile phones and gender empowerment: Negotiating the essentialist-aspirational dialectic. *Information Technologies & International Development*, *13*, 181–185.

Oreglia, E. (2014). ICT and (personal) development in rural China. *Information Technologies & International Development*, *10*(3), pp–19.

Orlikowski, W.J. (1999). Technologies-in-practice: an enacted lens for studying technology in organizations.

Orlikowski, W.J. (2000). Using technology and constituting structures: a practice lens for studying technology in organizations. *Organization Science*, *11*(4), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600.

Patel, N., & Agarwal, S. (2008). Experiences designing a voice interface for rural India. In *Spoken Language Technology Workshop* (pp. 21–24). https://doi.org/10.1109/SLT.2008.4777830, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4777830.

Persico, D., Manca, S., Pozzi, F. (2014). Adapting the technology acceptance model to evaluate the innovative potential of e-learning systems. *Computers in Human Behavior*, *30*, 614–622.

Porcheron, M., Fischer, J.E., Reeves, S., Sharples, S. (2018a). Voice interfaces in everyday life. In *Proceedings of the 2018 CHI conference on human factors in computing systems - CHI '18* (pp. 1–12). New York: ACM Press. https://doi.org/10.1145/3173574.3174214. http://dl.acm.org/citation.cfm?doid=3173574.3174214.

Porcheron, M., Fischer, J.E., Sharples, S. (2018b). Do animals have accents?: Talking with agents in multi-party conversation. https://doi.org/10.1145/2998181.2998298.

Pozzi, G., Pigni, F., Vitari, C. (2013). Affordance Theory in the IS discipline: A review and synthesis of the literature. In: AMCIS, 2014. Proceedings.

Price, J. (2010). Coding: Open coding. Encyclopedia of case study research, pp 155–157.

PTI (2018). Mobile internet users in India seen at 478 million by June. https://www.thehindubusinessline.com/info-tech/mobile-internet-users-in-india-seen-at-478-million-by-june/article23383790.ece.

Richardson, H.J. (2009). A 'smart house' is not a home: The domestication of ICTs. *Information Systems Frontiers*, *11*(5), 599–608. https://doi.org/10.1007/s10796-008-9137-9, https://link-springer-com.lcproxy.shu.ac.uk/content/pdf/10.1007%2Fs10796-008-9137-9.pdf.

Robinson, S., Pearson, J., Ahire, S., Ahirwar, R., Bhikne, B., Maravi, N., Jones, M. (2018). Revisiting "Hole in the Wall" computing: Private smart speakers and public slum settings. In *Proceedings of the 2018 CHI conference on human factors in computing systems - CHI '18* (pp. 1–11). New York: ACM Press. https://doi.org/10.1145/3173574.3174072. http://dl.acm.org/citation.cfm?doid=3173574.3174072.

Sambasivan, N., Weber, J., Cutrell, E. (2011). Designing a phone broadcasting system for urban sex workers in India. In *Proceedings of the 2011 annual conference on human factors in computing systems - CHI '11* (p. 267). New York: ACM Press. https://doi.org/10.1145/1978942.1978980. http://dl.acm.org/citation.cfm?doid=1978942.1978980.

Schlesinger, A., Edwards, W.K., Grinter, R.E. (2017). Intersectional HCI. In *Proceedings of the 2017 CHI conference on human factors in computing systems - CHI '17* (pp. 5412–5427). New York: ACM Press. https://doi.org/10.1145/3025453.3025766. http://dl.acm.org/citation.cfm?doid=3025453.3025766.

Spring, J. (2007). The triumph of the industrial-consumer paradigm and english as the global language. *International Multilingual Research Journal*, *1*(2), 61–78. https://doi.org/10.1080/19313150701489655. http://www.tandfonline.com/doi/abs/10.1080/19313150701489655.

Treem, J.W., & Leonardi, P.M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, *36*(1), 143–189.

Turovsky, B. (2017). Bringing down the language barriers - making the internet more inclusive. https://india.googleblog.com/2017/04/bringing-down-language-barriers-making.html.

Venkatesh, A. (2008). Digital home technologies and transformation of households. *Information Systems Frontiers*, *10*(4), 391–395.

Volman, M., Van Eck, E., Heemskerk, I., Kuiper, E. (2005). New technologies, new differences. Gender and ethnic differences in pupils' use of ICT in primary and secondary education. *Computers & Education*, *45*(1), 35–55.

Wicks, D. (2010). Coding: axial coding. Encyclopedia of case study research, 154–156.

Zamora, J. (2017). Rise of the Chatbots. In *Proceedings of the 22nd international conference on intelligent user interfaces companion - IUI '17 Companion* (pp. 109–112). New York: ACM Press. https://doi.org/10.1145/3030024.3040201. http://dl.acm.org/citation.cfm?doid=3030024.3040201.

Zheng, Y., & Yu, A. (2016). Affordances of social media in collective action: The case of Free Lunch for Children in China. *Information Systems Journal*, *26*(3), 289–313.

Zuboff, S. (1988). *In the age ofthe smart machine*. New York: Basic Book.

Zue, V., Seneff, S., Glass, J.R., Polifroni, J., Pao, C., Hazen, T.J., Hetherington, L. (2000). JUPLTER: A telephone-based conversational interface for weather information. *IEEE Transactions on speech and audio processing*, *8*(1), 85–96.

Linus Kendall is a PhD student at Sheffield Hallam University working on the use of information and communication technologies in development (ICTD). His present research is focused on design of agricultural knowledge management systems for smallholder and subsistence farmers in West Bengal, India.

Bidisha Chaudhuri is an Assistant Professor at the International Institute of Information Technology Bangalore (IIITB). Sheteaches undergraduate and graduate courses on modern social theories, digital sociology and social studies of science and technology.

Apoorva Bhalla graduated from the Masters in Digital Society program at IIIT Bangalore in 2019. Post that, she started working for Google as a UX Researcher. Currently, she is utilising and building on her qualitative and quantitative research skills to help design and build products used by millions of users across the world.