Conference contribution :

Robinson, S., Pearson, J., Ahire, S., Ahirwar, R., Bhikne, B., Maravi, N. & Jones, M. (in press). *Revisiting "Hole in the Wall Computing": Private Smart Speakers and Public Slum Settings.* CHI 2018: CHI Conference on Human Factors in Computing Systems Proceedings,
http://dx.doi.org/10.1145/3173574.3174072

Revisiting “Hole-in-the-Wall” Computing: Private Smart Speakers and Public Slum Settings

Simon Robinson,[1] **Jennifer Pearson,**[1] **Shashank Ahire,**[2] **Rini Ahirwar,**[2]
Bhakti Bhikne,[2] **Nimish Maravi,**[2] **Matt Jones**[1]

[1] FIT Lab, Swansea University, UK
{ s.n.w.robinson, j.pearson,
matt.jones } @swansea.ac.uk

[2] Industrial Design Centre,
IIT Bombay, Mumbai, India
ahire.shashank@iitb.ac.in

ABSTRACT

Millions of homes worldwide enjoy access to digital content and services through smart speakers such as Amazon’s Echo and Google’s Home. Promotional materials and users’ own videos typically show homes that have many well-resourced rooms, with good power and data infrastructures. Over the last several years, we have been working with slum communities in India, whose dwellings are usually very compact (one or two rooms), personal home WiFi is almost unheard of, power infrastructures are far less robust, and financial resources put such smart speakers out of individual household reach. Inspired by the “hole in the wall” internet-kiosk programme, we carried out workshops with slum inhabitants to uncover issues and opportunities for providing a smart-speaker-type device in public areas and passageways. We designed and deployed a simple probe that allowed passers-by to ask and receive answers to questions. In this paper, we present the findings of this work, and a design space for such devices in these settings.

Author Keywords

Conversational speech, public spaces, emergent users.

ACM Classification Keywords

H.5.2 User Interfaces: Interaction Styles

INTRODUCTION

For many years, speech has been upheld as the ultimate communication channel between people and computers. Seductive future visions of human-like conversations with machines and robots have been depicted in films and television for decades, illustrating the dream of seamless spoken interactions with inanimate systems. In reality, of course, speech recognition systems are not as conversationally fluent as the creative industries would lead us to believe. Recent advances in the development of more sophisticated speech recognition technologies, however, are beginning to see people becoming comfortable speaking aloud to their mobile devices; and, more recently, to their surroundings, after the introduction of smart speaker systems such as the Amazon Echo or Google Home.

There are several reasons for this new-found success. The abundant availability of high-bandwidth data connections, with greatly matured back-end services, and the ability to mine vast amounts of data, have significantly improved both the accuracy and general user experience with the latest, now cloud-based, voice interaction systems. Moreover, with the increase in connected homes and lifestyles, and the dominance of mobile devices, there are now a great many compelling use-cases for spoken interaction with digital assistants. For example, actions such as asking Siri to send a message eyes-free while driving, or calling out to Alexa to turn on the lights when hands are full, allow people to multitask with ease rather than directing their attention to potentially fiddly touchscreens.

Clearly, then, conversational speech systems are emerging as a mainstream interaction approach that now allows even the most inexperienced and technology-shy user the ability to talk with and control devices. Furthermore, in addition to providing a more natural approach for all, voice is also often seen as a potential benefit for users with particular requirements. For example, blind or partially sighted individuals [2, 6], people with dyslexia [1], and those with mobility impairments [31] have all benefited from speech-driven developments.

There has also been a large body of research dedicated to spoken language systems for “emergent” users. Emergent users, as described by Devanuj and Joshi [11], are the hundreds of millions of people, typically those living in developing regions of the world, often resource-constrained and economically disadvantaged, who are just beginning to gain access to the latest mobile devices. Speech, as both an input and output modality, has long been seen to be important in these contexts (e.g., [18, 38, 41]) due to the high rates of illiteracy that are common in emergent user communities. However, there are several challenges, specific to these areas, that have limited the advance of speech recognition to-date. While recognition of diverse languages has previously been a significant issue,[1] recent advances by companies such as Google have lowered this barrier somewhat [12, 33]. Instead, the primary obstacles now include resource constraints: constant power and always-on internet connections are often not available or affordable; and, a lack of money to purchase devices in the first instance. To address these issues, rather than focusing technology developments around individuals’ homes or devices, previous collaborations with emergent users have experimented with siting technology in public places, allowing communities to

[1] E.g., as of 2001, India had 22 official languages, with at least a further 122 spoken by more than 10,000 people [23].

learn and engage on their own terms. For example, Mitra et al. [8, 9, 24, 25] deployed a PC in a "hole-in-the-wall," allowing exploratory computer and internet use by a multitude of people who would otherwise be unable to access due to the many resource constraints mentioned above.

In this research, inspired by these previous approaches, we explored a public deployment of speech recognition technology in two emergent user communities in Mumbai. As with the earlier interventions, we are motivated to see such systems being made available to these communities for two reasons:

- To provide information access to an under-served population of potential users, providing them and their wider community with the ability to access services regardless of literacy, education or social standing.
- To help raise awareness of the use of speech recognition, so that emergent users are aware of and confident with its use when recognition support for their languages is achieved, and the hardware and resources required are more in line with the contexts in which they live.

While we are of course aware that many emergent users are now beginning to get their hands on more sophisticated devices and services that could potentially give access to mobile voice recognition [11, 16], we argue that the current research is valuable as a probe for more smart-speaker-like interactions. Given the challenges faced by users in these environments, the likelihood of such a technology being integrated into single emergent-user homes in the not-so-distant future is very low. The high cost of the devices themselves, plus the power and internet access consumables required to run them are contributing factors to this lack of likely adoption, in addition to the small, cramped nature of many of the homes in such areas.

Common in informal settlements or slums (such as the ones we describe here), however, are larger public outdoor areas where community members congregate to interact, shop and work, and which are a potentially attractive setting for charitable, governmental, NGO or other funded models of public speech installation. The probe we describe, therefore, allows us to not only raise awareness of this type of interaction system, but also gather feedback on how such a public installation might be used, in order to refine and adapt it for future use.

In this, our first step, our objective was two-fold. Firstly, we wanted to observe the effect of public speech interaction with different groups of emergent users. Our second aim was to identify and map out a range of design considerations for potential public conversational systems in such contexts, as a stimulus for further work by the wider community. In the rest of this paper, then, after surveying the related work, we describe focus group work with emergent users and its outcomes: a design space; and, a conversational speech probe that we deployed to investigate public speech interaction. Our results provide evidence that building publicly accessible conversational speech systems could lead to people experimenting with speech in a way that will not just be beneficial for public voice displays, but also in promoting awareness and community learning about this new modality that individuals might then use on their own devices as such access becomes viable.

BACKGROUND

There has been a long-term HCI interest in speech [39, 45] which has risen and fallen through many waves of optimism and experimentation, along with clear barriers around processing power for recognition, and interaction in challenging environments. We now see speech recognition widely deployed and used, and still very much a key research interest for the HCI community [29, 30]. The focus of this paper is on public speech interactions specifically within emergent user contexts. With this in mind, the following background focuses on speech services and public information access for emergent users, as well as more general public speech-focused installations.

Speech recognition for emergent users

It has been widely documented that the challenges of speech interaction in emergent user communities are greater than those that already exist for more "traditional" users of such technologies [19, 28, 38]. For example, noisy environments [4], multilingualism or dialectal variation [38] and a lack of support for local languages are all factors that affect the penetration of speech interaction in these areas.

Speech systems designed for emergent users, then, often focus on specific domain areas, such as crop prices or other work-focused information retrieval tasks (e.g., [22, 36]). Others use speech for literacy or language learning [20], or for healthcare [41]. Interactive voice response (IVR) forums—voice driven information systems for lower-literate users—are popular in emergent user communities to provide text-free advice on anything from farming to health. The Spoken Web [18], for instance, although primarily a keypad-based input system, was able to recognise some simple spoken words. A similar system, Avaaj Otalo [35], allowed a choice between touch (DTMF) input or voice commands for navigation, and found that 100 % of its users preferred touch-tone input. This is in contrast to research by Sherwani et al. [42], who discovered that speech interfaces outperformed touch-tone equivalents by lower-literate users for more conversational speech input.

There has also been research into techniques for automatic speech recognition (ASR) in less-widely-spoken languages [3, 19, 44]. Kumar et al. [19] evaluated the advantages and disadvantages of three types of ASR systems for developing regions, designing these to mitigate challenges such as unreliable cellular connections and low device processing power that are often the only options available to users in such environments. There are also spoken-based outreach programmes that are aimed specifically at largely lower-literate communities who are typically unlikely to be able to access information. QuestionBox[2], for example, is an organisation that deploys public helpline installations within lower-literate communities. These hard-wearing phone boxes allow community members to call a central group of outreach workers to retrieve spoken information on any topic from antenatal care to agriculture.

Multilingualism and speech recognition

Many emergent users across Africa and Asia speak a variety of different languages. Community members will often speak a

[2] http://www.questionbox.org/technical/

tribal or native mother tongue, as well as a more widely-spoken national language, with some understanding of neighbouring community languages, and often some English [36, 38]. As a result, multilingual interaction is commonplace, with communications switching between languages mid-conversation, or listening in one and answering in another. This type of conversational exchange poses a significant challenge to the majority of speech recognition services, which are typically monolingual in character (i.e., they are constrained to a single or low number of preset languages, and cannot detect arbitrary combinations in the same conversation [15]).

Most modern speech recognition services or devices follow this trait, and do not support multilingual dictation. Amazon's Echo, for instance, is available in English, German and Japanese (and has recently been rolled out in an Indian English variant [40]), but has to be preset into one of these languages beforehand. Apple's Siri can detect up to 21 languages but, again, must be preset. Google's services now have support for 119 languages [33], with the detection of up to five at a time, but do not support switching of languages mid-query [12]. As we discuss later in this paper, it is our view that it would be unnatural to preface a conversation by stating the language that is going to be spoken, which motivated the choice of a Wizard-of-Oz approach for the probe we tested in this research.

Devices for emergent users in public spaces

There have been many previous works that aim to provide public information access for emergent users. One of the most notable—and the inspiration for this paper—is the hole in the wall experiment (see [8, 9, 24, 25] and many more[3]). The aim was to place an enclosed desktop computer in a public location within emergent user communities and observe how the machine was used whilst giving no assistance. Primarily, the computer was intended as a learning tool for the community, who, at the time of its writing would have had little-to-no access to other sources of computerised information. The project was a success, and was subsequently deployed in many other communities after its initial installation[4].

Turning to other systems, Tamil Market [36, 37] was an early speech-driven query system designed for lower-literate users in rural communities. An initial instantiation of the Tamil Market system used a multimodal kiosk made from locally-available recycled components, and allowed simple one-word commands to be given verbally as well as via touch. In a later study of the system it was found that, despite many challenges, there was large potential for speech-based UIs in developing regions, and that lower literacy and technology experience did not adversely affect how users were able to navigate the system. In contrast to this evaluation, our probe was of more conversational speech (as opposed to simple navigational commands), and was conducted in a multi-user public environment rather than a single-user controlled laboratory study.

A related example is the StoryBank project [13], which was a walk-up-and-use kiosk for lower-literate emergent user communities in rural India. StoryBank used a public community screen for browsing of digital stories. More recently, the Com-Me [5] project deployed a public tablet installation in rural South African community for sharing media and digital stories. Both of these approaches used screen-based interfaces, whereas ours focuses solely on conversational speech.

Speech interaction in public spaces

There are several examples of previous research into speech interaction in public spaces. An early such installation was the Intelligent Kiosk [7], the goal of which was to create a "human-like" interaction with a walk-up-and-use public display. As part of this aim, the design included rudimentary speech recognition over a limited vocabulary (and requiring clear audio). The conclusion from this aspect of the design was that the recognition rate of the installation was too low for a commercial kiosk, and users found its lag frustrating. Consequently, pushing a button was seen as more favourable.

Another early example, which has since been refined and evaluated in a later investigation, is the MASK Kiosk [14, 21] – an advanced public service terminal with multimodal input and output capabilities (including speech input). The results showed that multimodality was more efficient than monomodality, indicating that a combination of speech and touch input was an advantageous solution for walk-up kiosks of this nature. Despite a prediction that users would be hesitant to speak to a kiosk in public, the results showed this was not the case, with 87 % of participants stating that they would be likely to use the device if it was positioned in a public place.

DESIGN INSPIRATION FOCUS GROUPS

Our starting point for this work was to revisit the hole-in-the-wall approaches first explored by Mitra [25], replacing the visual output and keyboard or touchpad input methods with a speech-based conversational system. Before creating a deployment probe, though, we first carried out a set of design inspiration and idea generation activities in three focus groups.

Method

We recruited 12 emergent user residents of Dharavi—a large slum in Mumbai, India—in groups of 4 people (9F, 3M overall). Broadly representative of the Dharavi population, participants had generally low educational attainment and literacy, lived with daily resource constraints (e.g., limited access to electricity, sanitation and consumables), and resided in tightly-packed one- or two-room dwellings. Nine participants had smartphones and three had featurephones, but apart from this and, in some cases, access to a TV, none of the participants owned or had access to any other form of digital technology. None had previously used voice interaction.

Meeting in a community hall in Dharavi, we began the workshops with an IRB-approved informed consent process. We then demonstrated Google Now, and showed the group an Amazon Echo Dot, explaining how this device was typically used in homes. Participants were then asked to reflect on how such interfaces would or would not be useful in the lives of themselves, friends, family and their wider communities.

After this discussion (which lasted for around 30 min), for the first two groups, we went on a walking technology tour [43]

[3] See: http://www.hole-in-the-wall.com/publications.html

[4] See: http://www.gg.rhul.ac.uk/ict4d/workingpapers/mitra1.pdf

through a number of nearby streets and public spaces. The locations visited included a community square next to a church, an outdoor workplace, and streets outside homes. Over a period of around 45 min, during the tour we explored with the group: i) where they would place an Echo Dot-type device (asking them to place it or point to where in the location it should be attached); ii) the sorts of questions or requests they would ask of it in each location; and, iii) to discuss potential issues of its use. As well as asking direct questions, we employed a role-play technique where one of the participants played the part of "Alexa," and the other was the user. The user was instructed to ask this "device" something they saw as useful in that context. During these role-play activities, bystanders watched and added to the discussion, leading to approximately an additional 20 people who contributed to the insights gathered. For the third group, this role-play had to be done indoors due to the torrential rain flooding the streets outside.

Finally, we returned to the community hall and asked the group to reflect on the overall set of discussions we had had, identifying the most and least useful potential services, locations to install the device, and any other thoughts they had. Participants were given ₹500 each as a token of our appreciation for their taking part in the workshops.

Insights and observations

During the initial discussion, the groups saw value in using speech input, with three predominant reasons given: i) the speed of interaction compared with typing; ii) the access the approach might provide to their friends and family who were not literate or used to using technology; and, more intriguingly, iii) the companionship the agent could provide for them when bored or alone. From the technology walk and role-play, we elicited useful design pointers with regard to the practicalities of deploying the conversational service in public; the social and economic considerations; the training and education requirements; and, potential applications.

Practicalities

There was concern that the proposed device would not be robust enough for the physical context. Our visit was during the monsoon season, when a great deal of rainfall was experienced daily, often flooding the narrow passageways. Participants pointed out the need to make the device waterproof, but also, for the dry times, resistant to the heat and dust in Dharavi. The physical security of the device was also raised, with the suggestion that the valuable parts of the device should be locked within a building, with a wire running to cheap public-facing components (a speaker and microphone in the street). On the technology walks, our groups saw many people who had their hands full, carrying food, goods or building materials, or were busy cooking, cleaning or making. Our participants noted the value of talking to a service given the inability in these circumstances to access a mobile phone.

Social and economic

Participants were mindful of their own privacy, and also the impact on those around them. One suggestion was for the device to modify the volume of the speech output depending on whether the interaction was with one, several or a large group of people. Another participant noted the need for sensitivity in terms of the services or content served depending on the location, noting that it would be bad, for instance, for inappropriate loud music to be accessed through the device near the church. Participants asked us how the service would be paid for. We discussed a series of options (e.g., government/NGO funding or advertising). Participants felt that advertising (e.g., where the speech system might proactively speak out an advert as people passed; or, where a commercial sponsorship message was played before giving a reply) would be less desirable as it would add to the noise of the already busy streets and slow down interactions. They noted that there were "cybers" (internet cafes) in Dharavi where people paid per use, and wondered how the street service might be set up to work in a similar fashion.

Training and education

Participants noted that people they knew would not understand what an Echo Dot-style device was or could do if they saw it in the street. They also felt that people would be worried about being embarrassed initially to try out such a service. Suggestions to overcome these problems included a simple graphical "how-to" poster near the device; or, for the device to engage with and prompt people as they came close to it.

Applications

In reviewing the sorts of questions and requests acted out during the walks and role-play, we see several potential focused applications of speech in the street. Firstly, many of the questions participants asked were about wayfinding and getting from one place to the next. Dharavi is a complex maze of streets and districts, with poor to no signposting. Our participants and bystanders, then, said they wanted to be able to get en-route reassurance and directions, asking the speech system in the same ways they might ask people in the street. To assist them to make sense of the responses from such requests, our participants also suggested adding simple visual outputs, such as an arrow that could point out the direction while the instructions were spoken. Participants also asked about train and bus schedules, explaining to us that as they walked towards stations or bus stops they regularly wanted to find the best connections for their often convoluted travel needs.

Dharavi residents face a range of everyday challenges including thefts and resource failures (e.g., of power, water etc.). Much discussion during the walk was around how the device could help in these situations: participants acted out requests to log problems with the local council; to call for help from the police; to advise them if there were good or bad things happening in that area; or, to directly ask which direction a thief had gone to help them chase after him or her. There were, though, many lighter suggestions of uses for fun and entertainment.

Returning to the community hall after the walks, participants were enthusiastic in discussion about the possibilities of the public speech service, mentioning specifically the ways it could be particularly useful for those who needed hands-free interaction because of what they were doing, and for people not used to technology (such as the older community members). They saw most value in placing speech points in places that larger numbers of people currently congregate (such as the square we walked to; or, near stations and bus stops).

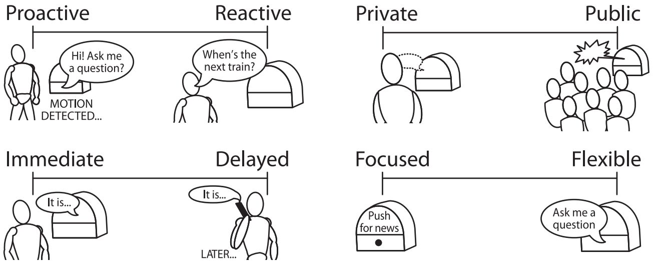

Figure 1. Design space for public speech interaction. Each of the four factors can be characterised on a sliding scale across the given dimension.

DESIGN SPACE

After the workshops, our research team discussed the observations, insights and suggestions of the focus group participants. In doing so we identified a range of interaction design choices for a deployment prototype:

Comprehensiveness: Here we discussed the extent to which a deployment in Dharavi should allow for broad questions or requests, or be focused on specific topics (e.g., navigation or the most frequently asked topics).

Interaction triggers and user agency: While voice input was seen as an obvious trigger to begin the interaction, we also considered much simpler triggers. For example, a device could have a set of categories such as news, transport and navigation, accessed by pressing a button next to the topic, after which the service speaks out some key information as a starting point on that topic. Other potential triggers included more complex ones such as proximity (e.g., the presence and number of people passing the device); and, person identification (e.g., by recognising phone IDs via Bluetooth or WiFi scanning). In these latter two cases, the system would be proactive rather than being reactive.

Degree of interactivity: A user might be able to engage in a back-and-forth conversation with the system once triggered; alternatively, the device might simply speak out content or information when triggered, with that output generated based on context (e.g., the location of the device; the time of day; the people passing the device).

Figure 1 illustrates these choices in the form of a design space, systematised as four key properties that we envisage will shape and characterise future public conversational speech systems.

PROTOTYPE

In order to further explore the potential for public space speech interaction, we created a physical prototype of a deployable device. After considering and evaluating the design outputs from the Dharavi workshops, we chose to focus on a comprehensive, user-triggered interactive system. That is, passers-by are able to walk up, speak a question and receive a response.

We used a Wizard-of-Oz method [17] to create the prototype. The device was a wooden box with *"Ask Google a question"* stencilled on the front of its casing (see Fig. 3). Inside the box was a battery-powered Bluetooth hands-free speaker and microphone kit (see Fig. 2, left). During the deployment, this kit was connected to the mobile phone of a local researcher standing nearby. To activate the system, the researcher initiated a phone call. All audio output from this phone call was audible via the speaker hidden inside the box. Spoken input from users was detected by the hands-free kit's microphone and thus relayed to those on the other end of the phone call.

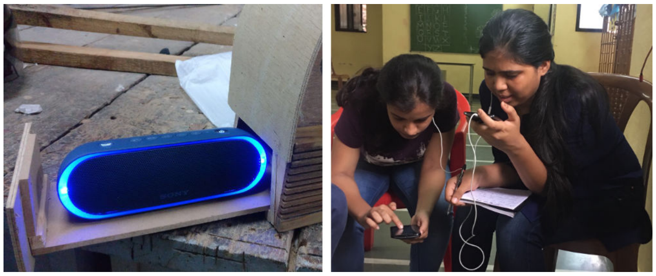

Figure 2. The "back-end" of the deployed system. Left: the inside of the prototype, showing the Bluetooth speaker concealed inside. Right: the "wizards" on the other end of the phone line, listening for questions, looking up answers and responding in a conversational system-like manner.

On the receiving end of the phone call (out of sight and hearing of those interacting with the system) were two multi-lingual local researchers. These "wizards" listened for questions asked to the system, searched for answers where necessary, and relayed these back to the questioner (see Fig. 2, right). The wizards strictly attempted to respond in a manner as close as possible to that used by existing conversational systems, rather than speaking in a human-like intelligent and conversational tone. For example, while one researched answers to more challenging queries, the wizard who was speaking responded with a set phrase asking participants to *"please wait a moment,"* and, if no answer was possible, relayed this to participants using a standard response. Our aim was to give the impression, as far as was possible, that it was a conversational speech system providing the responses, rather than a human.

The decision to use the Wizard-of-Oz method for the prototype deployment was to ensure a "best case scenario" version of a public speech interaction system. Had we based the deployment on an existing publicly-available speech system (e.g., Amazon Echo or Google Home) the interaction would have needed to be performed in English, as Hindi and Marathi versions of these are not yet available. Using a phone-based system that does support these and other Indic languages would have required a screen for certain output types, and either button presses or a trigger phrase (i.e., *"OK Google"*). Developing a multilingual speech recognition system is a serious engineering undertaking that is beyond the scope of this research. Although not without its own challenges, then, choosing to have a human answer the queries posed to the prototype meant that we were able to explore the potential future benefits of the technology before a more complete deployment; allowed for a rapidly-implemented and flexible user trial with no need for training or trigger words; and, maximised the possibility of the system being able to interpret multiple languages, nuanced phrasing, slang, or heavily accented speech.

USING THE PROBE

A week after the design inspiration focus groups, we invited the same groups back (with all participants now attending at the same time) to see and interact with the smart speaker prototype in order to gather feedback on its interaction and potential use in their everyday environments. We began by

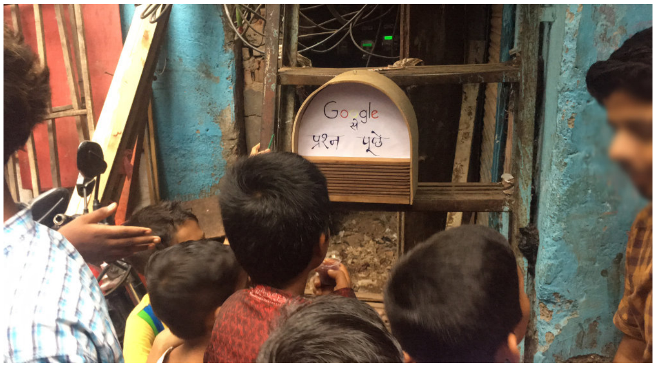

Figure 3. The prototype during the Dharavi street deployment. passers-by approach and ask a question, and receive a spoken response.

explaining how their feedback during the initial workshops had helped lead to the design. This was followed with a short demonstration of the probe's usage by a local researcher.

Each participant then asked the prototype a question (one-by-one, in whatever language they chose, in front of the whole group). Despite some initial hesitation, participants quickly became familiar with and confident talking to the probe. The most common queries at this stage were factual questions such as: *"who discovered electricity?"*; *"who is the Prime Minister of India?"*; or, *"what is the meaning of the word Google?"*. These questions were interspersed with more contextual queries such as: *"how long before it rains?"*; *"how far is Matunga station?"*; and, *"where is the nearest medical store?"*.

We then asked participants to imagine that the prototype was positioned in the street in a public area near where they lived; and, if this were the case, what might it say when they were walking past? The responses to this question tended to lean more towards proactive as opposed to reactive interactions. Some participants, for instance, wanted the system to be used as a warning: *"there is some construction work down here; try another way"*; *"fire or danger!"*; or, *"dead end this way"*. Others simply wanted the system to verbalise locality information, such as naming the road they were on or the weather in the area. Finally, there were suggestions that the speaker could state facts, which could be beneficial for learners. Overall, participants were seemingly very engaged with the technology throughout the session, interacting not only with the box itself, but also amongst themselves about what questions to ask.

In the next sections we describe two deployments in-situ in public settings to further explore the system's potential utility.

Street deployment: Dharavi

We deployed the prototype for around one hour in the same busy public square in Dharavi that we had visited for the focus groups (see Fig. 3 for an anonymised photo of the device in-situ). Our goal was to observe passers-by and analyse the types of queries and behaviours that the prototype stimulated.

Method

Our initial approach was to place the speaker in a visible location and simply observe its use. However, it was clear that despite the attention driven by our presence, passers-by were reluctant to (or unaware of how to) interact with the system. To give context, and to stimulate initial use, a member of the local research team approached the prototype and asked a sample question, explaining and demonstrating to the group who had gathered that this was what the box was designed to do.

Interactions from this point were recorded both remotely (by capturing the work of the wizard team) and locally (one researcher stood near enough to the probe to overhear speech; others stood further back, 5–15 m, to capture global impressions). All researchers took notes, and several videos and pictures of the probe in action were taken to record the groupings of participants and their dynamics. Following their use of the system, we also approached six people to ask (if they were willing) for their first thoughts and opinions about the system.

After the session, each researcher independently wrote up their observations and participant interactions. Following this, one researcher categorised these into key themes and findings, with the others validating, critiquing and refining.

Results

While many adults in the vicinity were curious, at the start of the session most were unwilling to try the device themselves. Children, however, were more forthcoming, and eagerly approached the box, talking as if to a person, which drew in further observers (both adults and children) to take part.

The queries at the start of the session were primarily fact-based – for example: *"do hens lay eggs?"*; *"who is India's cricket captain?"*; and, *"who is the Chief Minister of Maharashtra?"*. Participants commonly chained questions together (e.g., *". . . and what is his age?"*). Contextually relevant queries were also popular, such as: *"which is the fastest train to Patna from Mumbai?"*, *"[. . .] train timetable to reach Ratnagiri station?"*; *"when is the next India [cricket team] match?"*; and, *"what is the weather like today?"*. Domain knowledge or quick factual searches about the most up-to-date information meant that answers to these queries were quickly found by those in the research team who were acting as wizards. There were also queries with more opinion-based answers, however, such as: *"will Lalu Prasad resign?"*; *"why did Pakistan win?"*; and, *"which came first, chicken or egg?"*, to which the wizard team responded with statements such as: *"everyone is searching for that answer!"*. Requests for playing music were also common, as were questions about the questioner (e.g., *"what is my name?"*), or about the device (e.g., *"where is your home?"*, *"when were you born?"*).

Due to the noisy environment, for several questions the wizard team had to ask the speaker to repeat their enquiry, which resulted in many users starting with *"Google, listen carefully . . . "*. Late in the study one person asked a question, and was subsequently heard saying to another *"that's a person talking, not the internet"*, after which we decided to conclude the deployment in this location. For the most part, however, it was clear that passers-by interacting with the system were not aware that responses were generated by a human rather than a computer.

Street deployments: Chaitanya Nagar

We carried out two further deployments in a different, more publicly accessible area of Mumbai (but still an informal, slum-like setting), placing the device near to a cafe and then near to a snack centre, for around one hour in each place. We followed the same procedure as described for the deployment in Dharavi. That is, in both locations we gave a demonstration of prototype to the first few curious passers-by to alleviate initial apprehension. Once these people had begun interacting, observational effects were enough to sustain constant use, and the research team stood back to observe use from a distance as in Dharavi. Here, a total of five participants were asked for more detailed responses to the system.

Once again, younger individuals were initially more inclined to ask questions. Several older passers-by mistook the device for a fortune telling robot[5] and so did not engage with the prototype. As with the previous deployment, there were many general single-answer factual queries, such as those about capitals of countries or states, or queries about celebrities or politicians. In these deployments, users were more inclined to use the system for location based queries, asking for directions to nearby places such as hospitals, picnic places, government offices and even film stars' homes. After using the system, for instance, one participant stated *"generally we don't rely on people giving us directions as they can fool you, and some of them are also reluctant to help – in such cases, devices can help us with more accurate locations"*. Several autorickshaw drivers asked the device to check the distance between two locations. These drivers frequently needed assistance with directions to an unfamiliar place, but were cautious about asking others for help: *"we need some reliable source to get the direction and we don't know how to use maps, so such devices can help us to find the location faster"*.

During the rainy season in Mumbai there are often abrupt weather changes, so many participants used the device to check for weather updates. One participant, for instance, said: *"before leaving for some place I would like to know whether it will be a sunny or rainy day, or if there are going to be heavy showers sometime soon – this will help me to plan my day"*. Questions about the timings of trains and buses were also popular, as were more in-depth questions about specific topics that required more significant research. For example, users asked about the procedure for opening a bank account, general information about a recently introduced country-wide tax system (GST), as well as suggestions for easy-to-cook recipes, and queries about the correctness of spoken vocabulary.

DISCUSSION

A recent study conducted by Edison Research [32] found that 7 % of the US population now own a smart speaker. The reasons for wanting such a device, from the perspective of those taking part in the study, varied from convenience (*"to listen to music"* (90 %); *"to ask questions without needing to type"* (87 %)) to exploration (*"because it's a fun new gadget"* (86 %)), to more accessibility-based reasons such as: *"to help with a disability"* (16 %); or, *"to help the elderly"* (12 %). Perhaps unsurprisingly, given the context in which this study took place, none of the respondents cited literacy or technological accessibility as a reason for wanting to own a smart speaker.

Question category	**Dharavi**	**Chaitanya Nagar**
Basic facts	16	7
Context-specific information	15	13
Domain-specific queries	12	4
Philosophical questions	9	7
Total	52	31

Table 1. Categories of questions asked by passers-by during the street deployments in Dharavi and Chaitanya Nagar.

The value of such technology for those with little to no literacy, however, is potentially high. As Dell and Kumar [10] report, it is generally accepted amongst the HCI4D community that speech interaction is growing in popularity, and *"could be so huge for poor, low-literate people"* (cf. [10]) The potential barriers to appliance form-factor conversational speech systems being introduced into the individual homes of emergent users are great, however. A general lack of infrastructure to provide consistent power and internet connectivity, coupled with the lower economic status of many users in these areas, means that smart speakers are currently prohibitively expensive to buy, and technically demanding to run.

In our probe studies, taking inspiration from the hole-in-the-wall set of deployments (cf. [25]), we observed a wide variety of reactions and potential use-cases for deploying spoken systems in public slum settings. Of course, it is important to nuance the benefits of smart speakers in such settings. In our focus groups, participants suggested visual outputs for some queries; and, for inputs, discussed the benefits of buttons against those of speech just like traditional users might do. Participants, then, saw audio, and smart speakers, as just one of a range of modalities, rather than a panacea.

Overall, we can categorise many of the general queries into the following types (see Table 1 for further detail):

- Basic facts (e.g., *"who is the president of America?"*);
- Context-specific information (e.g., travel times, prices of local commodities, weather, news);
- Domain-specific queries (e.g., recipes, playing music);
- Philosophical questions, or those directed at the device itself (e.g., *"where are you [the device] from?"*)

When we consider mainstream smart speaker use, we see that the most common queries resonate somewhat with our public deployment experiment. Two of the four most common tasks conducted by traditional smart speaker users—asking about the weather and general fact-based questions—were also amongst the top queries in our public probe. The most used function of smart speakers for traditional users, however—playing music—was something our probe saw far less of.

One of the more popular queries observed during our public trials was to get directions from one place to another, which is a use-case not explicitly mentioned by traditional users (cf. [32]). Another common query during our probe involved asking for local bus or train times, which relates only slightly to the 10 % of traditional users who use their smart speaker to request flight information. There are also very clear differences in use between what could be considered as personal and public tasks. For instance, the fifth most used function of traditional smart

[5] A highly-decorated electronic attraction often seen being operated by enterprising sellers on Mumbai's beaches: https://goo.gl/4gTphT

speakers is for timers or alarms (43 %), a query we did not observe during any of our public trials. Similarly, controlling smart devices (33 %), editing to-do lists (26 %), adding to shopping lists (26 %), placing events in a calendar (23 %), and a range of other more home- or personal-focused interactions were not asked in any of our deployments.

Turning now to a comparison between our deployments and that of the hole-in-the-wall experiments (cf. [25]). One clear similarity was the uptake and excitement shown by younger individuals. In both installations, children were initially far more keen to make use of the system than adults. There were also parallels in terms of the searches chosen. For instance, participants using both our probe and the hole-in-the-wall installations often wanted to play specific songs, request horoscope information, or hear local news. Another similarity between the studies is the potential for external factors to diminish the user experience – in the hole-in-the-wall case, sunlight [25], vandalism [26] and wear-and-tear affected the interaction; in our case there is potential for external noise to interfere in a similar way to that reported by previous work in public spaces (e.g., [4]). We did observe this in some cases (e.g., participants asking the device to *"listen carefully"*), but it did not stop people from interacting with the device.

One obvious difference between the context of the hole-in-the-wall deployment and that of our installation is the change in availability of external information sources. It is generally accepted that many of the emergent users who live in and around the areas in which we have been working are beginning to get access to sophisticated devices and services [11, 16], which could now include mobile voice recognition services such as Google's assistant. While this is a potentially valuable source of information for such users, issues still remain around data costs and availability, as well as language and dialect challenges. A public installation of such a service, however, not only potentially overcomes issues relating to resource constraints, but can also contribute toward scaffolding communal education around the device, with users learning from one another to perform tasks, as seen in the hole-in-the-wall studies.

TAKING THINGS FURTHER

In this work we have explored an early-stage conversational speech probe in public. Taking such a deployment to a higher level of technology readiness will bring additional design challenges, however. Several of these can be identified and designed for in advance based on our own and others' results.

Turning first to user interaction with the probe. We saw in each of our deployments that adults were initially reluctant to interact with the system, while children used it uninhibited. We also saw minor instances of friction between users (mainly with people talking over one another). Looking back to the hole-in-the-wall deployments, there are resonances with both of these interaction aspects. In that system, children were first to interact with the PC, as is often the case in other areas of technology, with digital natives regularly leading the way [34]. The friction that sometimes arose amongst hole-in-the-wall users was primarily caused by increased demand for time on the machine, which saw some get frustrated when they could not have a turn. In our deployment, after overcoming their initial hesitation, it was easy for participants to join in at any moment by simply shouting queries from behind the person currently speaking. Future research on how multiple users can amicably share and interact with a single speech device could build on previous investigations of multiple inputs to a single screen-based system (e.g., [27]) to help inform design decisions.

Privacy issues, as we have already touched upon in the insights from focus groups, are highly likely to be a concern when interacting with speech systems in public. There will certainly be questions that people did not or would not ask because of privacy or other concerns. Here, we point to our design space (Fig. 1), and its dimensions that could be combined to preserve privacy. For example, sensitive content could be presented in a delayed fashion (i.e., when fewer potential listeners are nearby), or in a reactive manner upon speaking a passphrase.

It is also important to consider levels of query elicitation (both by the researchers and other participants), and participants' awareness of the possibilities afforded by the system. Our participants did not have the sort of search capability conditioning that regular users of internet search often have. We saw it as beneficial to appropriate the Google logo for brand recognition on our prototype (Google services are heavily advertised in India), but participants will not necessarily have used an internet search engine themselves. In our view, this is a benefit of the work – just as in the hole-in-the wall studies, "naive" participants probed and explored the system without the potential conceptual barriers that existing ideas of suitable uses might bring. The novelty of the system is likely to have affected participants' behaviour in this regard. While we did not control for novelty effects, we did see how participants shaped their queries, iterating and refining just as those familiar with internet search might do. We have not yet explored long-term interaction with such a system, so the ways in which interaction might change in the longer term, once any novelty effect has worn off, are a subject for future research.

Finally, it is important to consider the long-term requirements for maintaining a public conversational speech system (e.g., covering ongoing costs such as power and network access; repairing damage). Our participants felt that advertising was less desirable than charitable or NGO-supported services. A combined model, with advertisements clearly delineated from other content could, as previously discussed, contribute toward scaffolding communal education around the device

CONCLUSIONS AND FUTURE WORK

Smart speaker appliances are fast becoming commonplace in homes worldwide. Among the many advantages of speech, its text-free nature potentially makes it especially useful for the lower-literate, such as the many so-called emergent users living in resource-constrained areas of the world. The often challenging nature of these environments, however, means that standard, in-home smart speakers—devices that require constant power and a reliable, high-bandwidth internet connection—are typically beyond reach for emergent users.

The hole-in-the-wall computing concept has long been seen as a highly valuable resource for the demonstration, learning and awareness of screen-based technologies in emergent user communities. As a first step to identify whether this type of installation would be beneficial for speech-based technology, we conducted a Wizard-of-Oz investigation in two slum areas of Mumbai. We have described the promising results of this probe, and related our findings to the highly-regarded hole-in-the-wall literature. We have demonstrated the opportunities and values of conversational systems—imagined in the Western world for single family use in a domestic home—for public slum settings. Of course, we are not claiming the impact or longer-term insights of the hole-in-the-wall project. Rather, we argue that a key take-away from our work is that it is clearly worth investing the resources to create a deployable conversational speech system for the sorts of unsupervised learning and interaction benefits seen in that project. Our work provides evidence, then, that building publicly accessible conversational systems could lead to people experimenting with speech in a way that will not just be beneficial for any public "voice displays," but also in promoting awareness and community learning about this new modality that individuals might then use on their own device as this becomes viable.

ACKNOWLEDGEMENTS

This work was funded by EPSRC grant EP/M00421X/1. We would like to thank Sam Jones for his assistance during the fieldwork studies, and IIT Bombay technicians for their support in constructing the casing for the deployed probe.

REFERENCES

1. Theologos Athanaselis, Stelios Bakamidis, Ioannis Dologlou, Evmorfia N. Argyriou and Antonis Symvonis (2014). Making assistive reading tools user friendly: a new platform for Greek dyslexic students empowered by automatic speech recognition. *Multimedia Tools and Applications* 68, 3, 681–699. DOI: 10.1007/s11042-012-1073-5.

2. Shiri Azenkot and Nicole B. Lee (2013). Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '13). ACM, New York, NY, USA, 11:1–11:8. DOI: 10.1145/2513383.2513440.

3. Laurent Besacier, Etienne Barnard, Alexey Karpov and Tanja Schultz (2014). Automatic speech recognition for under-resourced languages: a survey. *Speech Communication* 56, 85–100. DOI: 10.1016/j.specom.2013.07.008.

4. Nicola J. Bidwell and Masbulele Jay Siya (2013). Situating asynchronous voice in rural Africa. In *Human-Computer Interaction – INTERACT 2013: 14th IFIP TC 13 International Conference, Proceedings, Part III*. Springer Berlin Heidelberg, Berlin, Heidelberg, 36–53. DOI: 10.1007/978-3-642-40477-1_3.

5. Nicola J. Bidwell, Simon Robinson, Elina Vartiainen, Matt Jones, Masbulele Jay Siya, Thomas Reitmaier, Gary Marsden and Mounia Lalmas (2014). Designing social media for community information sharing in rural South Africa. In *SAICSIT '14: Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014*. ACM, New York, NY, USA, 104–114. DOI: 10.1145/2664591.2664615.

6. Stephen Brewster, Matt Jones, Roderick Murray-Smith, Amit A. Nanavati, Nitendra Rajput, Albrecht Schmidt and M. Turunen (2011). We need to talk: rediscovering audio for universal access. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '11). ACM, New York, NY, USA, 715–716. DOI: 10.1145/2037373.2037494.

7. Andrew D. Christian and Brian L. Avery (2000). Speak out and annoy someone: experience with intelligent kiosks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '00). ACM, New York, NY, USA, 313–320. DOI: 10.1145/332040.332449.

8. Ritu Dangwal, Swati Jha, Shiffon Chatterjee and Sugata Mitra (2005). A model of how children acquire computing skills from hole-in-the-wall computers in public places. *Information Technologies & International Development* 2, 4, 41–60. DOI: 10.1162/154475205775249319.

9. Ritu Dangwal and Preeti Kapur (2008). Children's learning processes using unsupervised "hole in the wall" computers in shared public spaces. *Australasian Journal of Educational Technology* 24, 3, 339–354. DOI: 10.14742/ajet.1213.

10. Nicola Dell and Neha Kumar (2016). The ins and outs of HCI for development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 2220–2232. DOI: 10.1145/2858036.2858081.

11. Devanuj and Anirudha Joshi (2013). Technology adoption by 'emergent' users: the user-usage model. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction* (APCHI '13). ACM, New York, NY, USA, 28–38. DOI: 10.1145/2525194.2525209.

12. David Eustis (2014). Speak more than one language? So does Google. https://search.googleblog.com/2014/08/speak-more-than-one-language-so-does.html (visited on 25/08/2017).

13. David M. Frohlich, Dorothy Rachovides, Kiriaki Riga, Ramnath Bhat, Maxine Frank, Eran Edirisinghe, Dhammike Wickramanayaka, Matt Jones and Will Harwood (2009). Storybank: mobile digital storytelling in a development context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 1761–1770. DOI: 10.1145/1518701.1518972.

14. Jean-Luc Gauvain, Jean-Jacques Gangolf and Lori Lamel (1996). Speech recognition for an information kiosk. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP '96)*. Vol. 2. IEEE, 849–852. DOI: 10.1109/ICSLP.1996.607734.

15. Javier Gonzalez-Dominguez, David Eustis, Ignacio Lopez-Moreno, Andrew Senior, Françoise Beaufays and Pedro J. Moreno (2015). A real-time end-to-end multilingual speech recognition architecture. *IEEE Journal of Selected Topics in Signal Processing* 9, 4, 749–759. DOI: 10.1109/JSTSP.2014.2364559.

16. Matt Jones, Simon Robinson, Jennifer Pearson, Manjiri Joshi, Dani Raju, Charity Chao Mbogo, Sharon Wangari, Anirudha Joshi, Edward Cutrell and Richard Harper (2017). Beyond "yesterday's tomorrow": future-focused mobile interaction design by and for emergent users. *Personal and Ubiquitous Computing* 21, 1, 157–171. DOI: 10.1007/s00779-016-0982-0.

17. John F. Kelley (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems* 2, 1, 26–41. DOI: 10.1145/357417.357420.

18. Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty, Sheetal K. Agarwal and Amit A. Nanavati (2007). WWTW: the world wide telecom web. In *Proceedings of the 2007 Workshop on Networked Systems for Developing Regions* (NSDR '07). ACM, New York, NY, USA, 7:1–7:6. DOI: 10.1145/1326571.1326582.

19. Anuj Kumar, Anuj Tewari, Seth Horrigan, Matthew Kam, Florian Metze and John Canny (2011). Rethinking speech recognition on mobile devices. In *Proceedings of International Workshop on Intelligent User Interfaces for Developing Regions (IUI4DR)*. http://repository.cmu.edu/lti/117/.

20. Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal and Matthew Kam (2012). Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 1149–1158. DOI: 10.1145/2207676.2208564.

21. Lori Lamel, Samir Bennacef, Jean-Luc Gauvain, Hervé Dartigues and Jean-Noël Temem (2002). User evaluation of the mask kiosk. *Speech Communication* 38, 1, 131–139. DOI: 10.1016/S0167-6393(01)00048-6.

22. Gautam Varma Mantena, S. Rajendran, B. Rambabu, Suryakanth V. Gangashetty, B. Yegnanarayana and Kishore Prahallad (2011). A speech-based conversation system for accessing agriculture commodity prices in Indian languages. In *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, 153–154. DOI: 10.1109/HSCMA.2011.5942384.

23. Ministry of Home Affairs, Government of India (2001). Census of India 2001: General Note. http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/gen_note.html (visited on 01/09/2017).

24. Sugata Mitra, Ritu Dangwal, Shiffon Chatterjee, Swati Jha, Ravinder S Bisht and Preeti Kapur (2005). Acquisition of computing literacy on shared public computers: children and the "hole in the wall". *Australasian Journal of Educational Technology* 21, 3, 407–426. DOI: 10.14742/ajet.1328.

25. Sugata Mitra (2000). Minimally invasive education for mass computer literacy. In *Conference on Research in Distance and Adult Learning in Asia*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.112.9984.

26. Sugata Mitra (2005). Self organising systems for mass computer literacy: findings from the 'hole in the wall' experiments. *International Journal of Development Issues* 4, 1, 71–81. DOI: 10.1108/eb045849.

27. Neema Moraveji, Kori Inkpen, Ed Cutrell and Ravin Balakrishnan (2009). A mischief of mice: examining children's performance in single display groupware systems with 1 to 32 mice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 2157–2166. DOI: 10.1145/1518701.1519030.

28. Preeti Mudliar, Jonathan Donner and William Thies (2012). Emergent practices around cgnet swara, voice forum for citizen journalism in rural India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development* (ICTD '12). ACM, New York, NY, USA, 159–168. DOI: 10.1145/2160673.2160695.

29. Cosmin Munteanu, Matt Jones, Steve Whittaker, Sharon Oviatt, Matthew Aylett, Gerald Penn, Stephen Brewster and Nicolas d'Alessandro (2014). Designing speech and language interactions. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 75–78. DOI: 10.1145/2559206.2559228.

30. Cosmin Munteanu and Gerald Penn (2015). Speech-based interaction: myths, challenges, and opportunities. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15). ACM, New York, NY, USA, 2483–2484. DOI: 10.1145/2702613.2706679.

31. Maia Naftali and Leah Findlater (2014). Accessibility in context: understanding the truly mobile experience of smartphone users with motor impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '14). ACM, New York, NY, USA, 209–216. DOI: 10.1145/2661334.2661372.

32. NPR and Edison Research (28th June 2017). The Smart Audio Report. http://nationalpublicmedia.com/smart-audio-report/.

33. Thuy Ong (2017). Google now recognizes 119 languages for voice-to-text dictation. https://www.theverge.com/2017/8/14/16142786/google-recognises-119-languages-dictation-voice-typing (visited on 25/08/2017).

34. John Gorham Palfrey and Urs Gasser (2010). Born Digital: Understanding the First Generation of Digital Natives. Basic Books.

35. Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave and Tapan S. Parikh (2010). Avaaj otalo: a field study of an interactive voice forum for small farmers in rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 733–742. DOI: 10.1145/1753326.1753434.

36. Madelaine Plauché, Udhyakumar Nallasamy, Joyojeet Pal, Chuck Wooters and Divya Ramachandran (2006). Speech recognition for illiterate access to information and technology. In *2006 International Conference on Information and Communication Technologies and Development*. IEEE, 83–92. DOI: 10.1109/ICTD.2006.301842.

37. Madelaine Plauché and Madhu Prabaker (2006). Tamil market: a spoken dialog system for rural India. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06). ACM, New York, NY, USA, 1619–1624. DOI: 10.1145/1125451.1125746.

38. Madeline Plauché and Udhyakumar Nallasamy (2007). Speech interfaces for equitable access to information technology. *Information Technologies & International Development* 4, 1, 69–86. DOI: 10.1162/itid.2007.4.1.69.

39. Nitin Sawhney and Chris Schmandt (2000). Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction* 7, 3, 353–383. DOI: 10.1145/355324.355327.

40. Sunny Sen (2017). The backstory of Alexa's Indian makeover. https://factordaily.com/amazon-alexa-india-makeover-review/ (visited on 22/11/2017).

41. Jahanzeb Sherwani, Nosheen Ali, Sarwat Mirza, Anjum Fatma, Yousuf Memon, Mehtab Karim, Rahul Tongia and Roni Rosenfeld (2007). Healthline: speech-based access to health information by low-literate users. In *2007 International Conference on Information and Communication Technologies and Development*, 1–9. DOI: 10.1109/ICTD.2007.4937399.

42. Jahanzeb Sherwani, Sooraj Palijo, Sarwat Mirza, Tanveer Ahmed, Nosheen Ali and Roni Rosenfeld (2009). Speech vs. touch-tone: telephony interfaces for information access by low literate users. In *2009 International Conference on Information and Communication Technologies and Development (ICTD)*, 447–457. DOI: 10.1109/ICTD.2009.5426682.

43. Roger Silverstone and Leslie Haddon (1996). Design and the domestication of information and communication technologies: technical change and everyday life. In *Communication by Design: The Politics of Information and Communication Technologies*. Oxford University Press, 44–74. http://eprints.lse.ac.uk/64821/.

44. Charl Van Heerden, Etienne Barnard and Marelie Davel (2009). Basic speech recognition for spoken dialogues. In *INTERSPEECH-2009*. International Speech Communication Association, 3003–3006. http://hdl.handle.net/10204/3649.

45. Ian H Witten (1983). Principles of computer speech. Academic Press, Inc.