MAKING VOICES HEARD
DESIGN BRIEF

Making Voices Heard: Design Brief

Research and Writing **SAUMYAA NAIDU**
Research Assistance **SWETA BISHT, DEEPIKA NANDAGUDI SRINIVASA**
Review and Editing **SHWETA MOHANDAS, PUTHIYA PURAYIL SNEHA, DIVYANK KATIRA**
Research Inputs **SUMANDRO CHATTAPADHYAY**
Copyediting **THE CLEAN COPY**
Illustration **KRUTHIKA N.S.**
Report Layout and Design **SAUMYAA NAIDU**

CENTRE FOR INTERNET AND SOCIETY
Supported by Mozilla Corporation

Contents

1. Background

Given the increasing number of voice interface (VI) products in India, it is important to understand and analyse their design as part of the research and development of these technologies. In order to understand the challenges and opportunities in VI design, as well as to identify some best practices, we interviewed designers working with VIs.

The existing VI landscape comprises various actors, such as start-ups, global organisations, developers, policy-makers, and individuals using the VIs. Our mapping of actors in the VI industry suggests that a large number of the upcoming VI products in India are being developed by private companies.[1] In these companies, VI products are mostly conceptualised by developers, and designed by in-house teams. These teams include designers specialising in conversational design, and user interface (UI) and user experience (UX) design. Conversational experience design is still an emerging discipline in the country. It draws from human conversation patterns to make digital systems easy and intuitive to use.[2] The principles of conversational design can, therefore, be applied beyond voice assistants and chatbots to include all UI and web design.[3] However, in the case of most start-ups, the designers' role and scope are based on the UI and UX design process.

Our primary methodology involved interviews with designers working independently or with developers, start-ups, and global organisations, as well as developers who make broad design decisions and work with designers. These designers and developers include Preeti Sheokand, a user experience designer who specialises in conversational design at Symphony AI;[4] Kumar Rangarajan and Vinayak Jhunjhunwala, co-founder and marketing associate, respectively, at Slang Labs;[5] Jai Nanavati, co-founder at Navana Tech;[6] Megan Branson, senior product designer at NVIDIA, formerly at Common Voice;[7] Akshay Kore, senior product designer at Observe.ai;[8] and Keshav Prawasi, co-founder at Niki.[9]

In this study, we look at the design of VIs in India based on three key criteria: multi-

1 "Making Voices Heard: Mapping Actors," *Making Voices Heard*, accessed 02 February 2022, **http://voice.cis-india.org/mapping-actors.html**

2 Hampton, M., "Principles of Conversational Design," *Marvel Blog*, 30 October 2020, accessed 4 August 2021, **https://marvelapp.com/blog/principles-of-conversational-design/#:~:text=The%20concept%20of%20conversational%20design,more%20natural%20dialogue%20with%20systems**.

3 Hampton, "Principles of Conversational Design."

4 "Symphonyai: Transforming Businesses with Enterprise AI," *Symphony AI*, accessed 4 August 4 2021, from **https://www.symphonyai.com/**.

5 "Slang Labs: Add Accurate Multilingual Voice Assistants to Your App," *Slang Labs*, accessed 4 August 2021, **https://slanglabs.in/**.

6 "Navana Tech: Turn on the Conversation," *Navana Tech*, accessed 4 August 2021, **https://navanatech.in/**.

7 "Common Voice by Mozilla," *Common Voice*, accessed 4 August 2021, **https://commonvoice.mozilla.org/en**.

8 "Contact Center AI," *Contact Center AI | Observe.AI*, accessed 4 August 2021, **https://www.observe.ai/**.

9 "Niki: Aapke Ghar Ki Manager," *Niki*, accessed 4 August 4 2021, **http://niki.ai/**.

language support, accessibility, and privacy. As VIs gain popularity in the country, developers have realised that multi-language support is the key to success with Indian audiences. The focus on multi-language support has been a business enabler for the VI landscape. It has opened up a new design space, with a shift in focus from metros to start-ups aiming at the 'next billion users'.[10,11,12] However, developers have also identified challenges in supporting the numerous languages and dialects in India such as lack of language training data and technical expertise. Further, though VIs are globally recognised as accessibility tools for people with disabilities, at present, accessibility is not considered a primary objective for VI products in India. Instead, VIs are seen as tools to reach low-literacy individuals. In terms of privacy, the concerns around VIs usually focus on the device always listening for the 'wake word'.[13] There are no comprehensive guidelines on privacy standards for VIs. The design of VIs hence, also do not follow any specific privacy-preserving principles.

2. VI design and development processes

Over the course of our conversations with designers and companies, we observed that there is a largely standard process for the design and development of VIs. Based on the design thinking process, it comprises the broad steps of primary research, conversation-design modelling, testing, and refining.

2.1. Primary research

Preeti Sheokand explained that her design process begins with secondary research on the domain and the context within which the VI product will operate. She then identifies the limitations or challenges present in the context – for example, the presence of a noisy environment. Further, she conducts primary research to understand the needs and context of the people who are going to use the VI. Based on this, she works out the intents of the VI, including its 'hygiene intents'. An intent is the objective of the voice interaction or the individual's intention.[14] The VI understands these intents and responds to them. 'Hygiene intents', as Preeti calls them, are interactions needed to accomplish the intent. For example, if the objective is to buy groceries, the hygiene intent would include signing up or signing into the platform, creating a profile, and selecting items for

10 Majumdar, S., "Voice and Vernacular: The Future of E-retail in India," *Fortune India: Business News, Strategy, Finance and Corporate Insight*, 27 February 2021, https://www.fortuneindia.com/first-edit/voice-and-vernacular-the-future-of-e-retail-in-india/104630.

11 Choudhury, D., "Building Products for the Next Billion Users: Solving the Language Barrier, Monetisation Puzzle and More," *Inc42 Media*, 26 September 2020. https://inc42.com/features/decoding-the-psychology-of-the-next-billion-users-as-products-scale-up/.

12 Sachitanand, R., "Voice, Video and Vernacular: India's Internet Landscape is Changing to Tap New Users," *The Economic Times*, 7 October 2018, https://economictimes.indiatimes.com/tech/internet/voice-video-and-vernacular-indias-internet-landscape-is-changing-to-tap-next-wave-of-users/articleshow/66102478.cms?from=mdr.

13 Lynskey, D., "'Alexa, Are You Invading My Privacy?' – The Dark Side of Our Voice Assistants,," *The Guardian*, 9 October 2019, https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants.

14 "What Is a Voice User Interface (VUI)?" *Alan Blog*, accessed 20 May 2021, https://alan.app/blog/voiceuserinterface/.

purchase.

Preeti observed that while technologists create synthetic conversations for development purposes, there is a lack of understanding of natural conversations. She addresses this using the training data collected during the primary research. She then extrapolates starting points based on this research. She then works on the conversation flows following UX and conversation-experience design principles. Preeti explained that conversational design, when using artificial intelligence (AI), is a probabilistic system and not a deterministic one. In a probabilistic system, the occurrence of events cannot be predicted perfectly.[15] The behaviour of such a system can be understood in terms of probability. A deterministic system, on the other hand, is one in which the occurrence of all events is known with certainty.[16] The VI may have multiple responses based on what the individual says and how the VI understands it. The various possibilities of how a VI system understands a phrase are determined by its technological limitations and the individual's context. Hence, the conversation flows are decided for each of these possibilities or responses.

Our conversations with Navana Tech and Niki indicated that VIs are currently being envisioned for audiences with less experience with technology, and, hence, the initial design process revolves around capturing their interactions with existing platforms and assessing how they would potentially interact with VIs. Navana Tech has divided its audience into five literacy and technology cohorts by conducting tests to determine each segment.

The team at Niki too has significantly invested in primary research. Over the last three years, the team has spent about 30,000 hours talking to individuals using the Niki app. Keshav Prawasi informed us that Niki has a dedicated in-house customer insights and research team that consistently works towards understanding these individuals better. Their team of researchers, designers, and product managers has also travelled to Rajasthan and visited Tier 2, Tier 3, and other cities, like Chomu, Pushkar, Ajmer, and Udaipur, to conduct usability studies. They studied how people interact with platforms such as YouTube and WhatsApp. Further, they conducted hackathons, brainstormed for a few weeks, and recorded videos with people. They ideated multiple design concepts and finalised a few, based on which they built prototypes. They then tested specific use cases such as bill payment. They studied people's interaction with the prototypes and further refined the design. Then they ran tests again to observe for which functions people relied more on voice. Finally, they made upgrades and changes to reflect these observations.

Kumar Rangarajan at Slang Labs described the user research that he conducted for product design with Srishti Manipal Institute of Art, Design, and Technology, Bengaluru. The design researchers worked with some apps that used touch and others that used speech. They spoke to more than 50 people, including some who

15 Thakur, D., "Differentiate between Deterministic and Probabilistic Systems," *Computer Notes*, 30 January 2013, accessed 22 May 2021, **https://ecomputernotes.com/mis/information-and-system-concepts/differentiate-between-deterministic-and-probabilistic-systems**.

16 Thakur, D., "Differentiate between Deterministic and Probabilistic Systems."

are not well-versed with technology but who owned at least a smartphone. They asked shopkeepers and people on the streets and at bus stands how they would communicate to have a device perform a certain task. They also conducted more detailed, one to two-hour-long interviews with 10 people and observed them. They carried out primary research in English and Hindi. Then, they identified a use case and designed wireframes to test the flow of the interface. Like Preeti, Kumar also talked about considering the various intentions of the individuals using VIs, creating all possible conversation flows, and identifying different variables in these flows. He also spoke of expanding design details by adding various ways in which a primary use case, such as voice search, can be triggered while using the VI. Their focus is on 'productising' all the learnings from the research into their VI platform, so that businesses who integrate the VI do not need to start over.

2.2. Understanding the context

Going deep into the design process, Akshay Kore, who has previously been part of the team working on the Microsoft Cortana interface, shared several insights. To begin, he shared some questions to consider while designing a VI. The first is whether a voice interface is appropriate to that particular context. To answer this, Akshay shared some advantages of VIs and contexts where it is suitable:

- The VI substitutes for a complex action or task that requires multiple clicks or steps. For example, setting an alarm requires multiple steps and is time-consuming, but through VI, it can be set using a single command.
- When people are engaged in an activity where they cannot use their hands to access technology, voice becomes an important medium to interact with the device (for example, while driving or cooking).
- Using VI does not require any added learning. People can ask the VI questions, and the VI can either answer the question or respond that it cannot understand or address the query.
- Talking is more intuitive than other ways of interacting with technology. One can convey more through a VI than through a text-only interface, as other factors, such as tone, the difference between a question and an exclamation, and several emotions are conveyed more effectively through voice.

Akshay also mentioned contexts that are inappropriate for VIs:

- In public spaces, it is difficult to use VIs due to the presence of noise.
- If an application requires a lot of editing, using a VI is not advisable, as one cannot undo actions on VIs.
- Individuals may not be comfortable sharing health-related information or other private details with a machine, especially if the VI is being used in a public space.

Akshay reiterated Preeti's idea of context – he suggested that the designer be aware of the context for which they are designing the VI. They should consider the surroundings of the individual, whether they are a beginner or an expert in using technology, and the type of device they are using (a device with a screen or a speaker or both). For example, when designing a healthcare-based VI, research

may not be easily available as it comprises sensitive information, so the design process would need to involve several rounds of testing and feedback.

2.3. Testing and refining

Megan Branson, former senior product designer at Common Voice (CV), mentioned that they applied design at a conceptual level. The project used an iterative process which involved repeated testing with people and refining the platform. The project began with identifying the need for large quantities of publicly available voice data that could be used to train speech-to-text engines. Design thinking exercises with Mozilla community members were conducted to ideate on creating an open-source voice dataset.[17] Megan created paper prototypes of design concepts and gathered feedback on them. The initial assumption was that people would need an ulterior motive to share voice data. However, their research revealed that most people were willing to donate voice data. The team also realised that people wanted to learn more about the need for voice data collection. Hence, they designed a platform whose predominant objective is collecting voice data.[18]

In this initial iteration, CV developed an interactive model where people could 'teach' a robot to understand human speech by reading sentences to it.[19] This version intended "to tell the story of voice data and how it relates to the need for diversity and inclusivity in speech technology".[20] The team then gathered community feedback and developed further iterations. Megan explained that they did a UX audit of the working prototype at this stage and made further refinements. Since 2017, they have focused on improving the platform – primarily improving the experience of contributing voice data. They also took UX heuristics, competitor evaluations, and community feedback into consideration.

Our interviews indicate that there is a strong emphasis on primary research to understand the needs of people from varying backgrounds. Most interviewees placed a lot of focus on understanding how people with less experience of technology, in both rural and urban settings, use VI. Many VI companies aim to provide reliable banking and fintech services for rural audiences. As there is little precedent for designing VIs in India, designers follow the established UI/UX path. Most designers working on VI products are UI/UX or product designers by training or experience and have only recently familiarised themselves with the nuances of conversational experience design.

2.4. Conversation experience design

Conversation design is only just emerging as a discipline and specialised practice in India. Designers and services have put together guidelines and principles of

17 Branson, M., "We're Intentionally Designing Open Experiences, Here's Why," *Medium*, accessed 13 May 2021, https://medium.com/mozilla-open-innovation/were-intentionally-designing-open-experiences-here-s-why-c6ae9730de54.

18 Branson, M., "We're Intentionally Designing Open Experiences."

19 Branson, M., "We're Intentionally Designing Open Experiences."

20 Branson, M., "We're Intentionally Designing Open Experiences."

conversational design.[21] The core principles for conversation design in India were developed by Cathy Pearl in her book *Designing Voice User Interfaces*.[22] During his presentation on designing the best VI experiences, Vinayak Jhunjhunwala from Slang Labs talked about the best practices for conversation experience design based on Pearl's book and other resources.

- Defining expectations using convention: In VI design, it is important to break away from existing conventions and unlearn previous digital behaviour.
- Setting the right expectations: It is important to eliminate open-ended greetings and rhetorical questions from the design, as they are difficult to answer for the VI due to cognitive overload.
- Discoverability: Elements should be easily accessible in the VI. For example, the individual should be able to discover the voice button and quickly understand how to use it.
- Affordance: Vinayak emphasised that voice needs novel affordance strategies, such as audio prompting and adding visual depth to the interface.
- Fail-safes: Fail-safes should be built into the interface to counter instances of when a phrase is not heard, or when it is heard incorrectly.
- Use cases: He recommended picking common use cases, creating sample dialogues for each case, and testing them with different people. Sketching a voice user interface (VUI) flow diagram is another recommended technique.
- Confirmations: There should be explicit audio confirmation of commands to assure the individual that the VI has understood the task.
- Error Handling: VI needs to be designed to handle errors and latency in responses.

Other principles in Pearl's book include using conversational markers that let the individual know where they are in the conversation; adapting to the experience and expertise of novice and expert individuals keeping track of the context of the input; including a set of universals such as 'repeat', 'main menu', and 'help', at every stage; using audible or visual cues to communicate unavoidable system delays; designing experiences for accessibility; and prioritising personalisation over personality.[23]

3. Challenges in designing VUI

Some of the key challenges that designers faced were the poor memory of the VI; the need for multiple potential conversation flows; the need to handle errors; technological barriers; and a lack of language compatibility.

3.1. Poor memory

Akshay warned us that maintaining a record of previous interactions is a concern

21 "Voice Principles: Clearleft," *Voice Principles | Clearleft*, accessed 7 June 7 2021, **https://www.voiceprinciples.com/**.

22 Pearl, C., *Designing Voice User Interfaces: Principles of Conversational Experiences*, O'Reilly Media, (2017).

23 "Voice Principles," Voice Principles.

for VIs. While designing, it is important to ascertain what sort of memory the machine should have. This can be difficult to judge, as in some cases, multiple individuals may access the device. Products such as Amazon Echo are designed for the home setting, where there are multiple family members and activities.[24] But most voice assistants are designed to talk to one person at a time. The design challenge here is creating voice assistants that can address a group, know how many people are present, and be able to distinguish the situations and profiles of these individuals.[25] Akshay suggested that profiling the individuals talking to the device can enable it to provide contextual responses based on who is interacting with it. While it is not clear what this profiling would entail, it could mean differentiating individuals based on their speech patterns and/or voice biometrics.[26] This can then be used to build a history of their commands and identify and list their intent. However, voice profiling can have grave privacy implications, such as voice-based surveillance, targeted advertising, and the leakage of sensitive voice biometrics.[27]

3.2. Designing potential dialogues

Akshay also stated that the information provided by the VI must be related to the conversational context. While humans understand this intuitively, automated responses may not always be appropriate. Niki also refers to this challenge as "bringing the individual back into the conversation". Thus, it is necessary to write rigorously and design potential dialogues between the interface and the individual. As we mentioned earlier, VI uses probabilistic technology, which means there can be multiple responses in different contexts. For instance, how a VI interprets homophones such as 'pair' or 'pear' would depend on the context. In the case of a food delivery app, 'pear' takes precedence over 'pair'. These potential dialogues must be designed to understand the questions accurately, respond appropriately, set the right expectation, and provide confirmation to the individual.

3.3. Handling errors

Many variables affect VIs – such as background noise and accents – and which make them less than perfect. Handling errors becomes significant when designing VIs. The accuracy of the device, or the confidence of output, as Akshay phrases it, is impacted by design and product thinking.

3.4. Focus on technical approaches

Preeti pointed out that the VI industry has only recently realised the value of UX. Most developers are not designing for experiences but for the completion of tasks. The interfaces are mostly created by people with technical expertise. Preeti

24 Santos, M. E., "Designing Better Voice interfaces for Everyday Life," *Medium*, accessed 23 June 2021, **https://uxdesign.cc/designing-better-voice-interfaces-for-everyday-life-2cb344913fae**.

25 Santos, M. E., "Designing Better Voice Interfaces."

26 Turow, J., "Shhhh, They're Listening – Inside the Coming Voice-profiling Revolution," *The Conversation*, 28 April 2021. **https://theconversation.com/shhhh-theyre-listening-inside-the-coming-voice-profiling-revolution-158921**.

27 Turow, J., "Shhhh, They're Listening."

suggested that it is critical, even from a business perspective, that one moves beyond this purely technical approach and starts looking at how individuals use VI. The focus should shift from data sets to studying use cases; the design process requires greater sensitivity towards the purpose and the audience and their comfort.

3.5. Controlling or restricting content

Preeti also emphasised the need to focus on accessibility, transparency, and ethical practice when designing VIs, as the applications are a lot more open-ended. To illustrate her point, she used the example of Alexa, where children may ask the device for information that their parents may not want them to know yet. While parental controls can be applied to visual content, controlling or restricting content on VIs is more challenging as identifying and differentiating between the individuals using the device is a complex operation.

3.6. Designing with optionality

Navana Tech's co-founder, Jai Nanavati, also talked about the challenge of dealing with optionality in voice menus. He explained that although banking facilities can offer 20–30 service options, it is difficult for the individual to remember the options and effectively provide input to the VI. He suggests that a chat-based interface is more helpful in such a scenario. Effective VUIs should provide brief information and ask individuals if they want to hear more before offering additional options.[28] It is also important to allow individuals to request that information is repeated whenever they need it.[29]

3.7. Clarifying scope

Kumar observed that while touch limits options to those on the screen, in the case of VI, the individual using it can say anything. Hence, when an individual says something that is beyond the scope of the VI, there needs to be some feedback to inform the individual that their request is out of the scope of the service.

Based on her experience designing the CV website, Megan talked about the difficulties in designing for responsiveness. She believes that accommodating a large amount of information in a small device or screen is even more challenging with the localisation of CV in various languages. She also saw this as a sign that CV is growing. The varied perspectives of multiple languages, the politics of language, and locale codes or language identifiers in computing have presented interesting pain points while working on the platform. Within linguistic communities themselves, the question of locale codes for specific languages on CV is a big discussion.

28 Sengupta, A., "A Sound Relationship: 4 Tips to Build an Engaging Voice User Interface," *Wipro Digital*, accessed 18 May 2021, https://wiprodigital.com/2019/05/22/a-sound-relationship-4-tips-to-build-an-engaging-voice-user-interface/.

29 Kamm, C., "User Interfaces for Voice Applications," in *Voice Communication between Humans and Machines*, National Academies Press: OpenBook, 2019, 426 https://www.nap.edu/read/2308/chapter/30#426.

4. Designing for multiple languages

While most companies and designers recognise the need for multi-language support in VI products, most VIs lack regional language compatibility. According to Akshay, the predominant languages for VI in the country are English (US), English (UK), French, and English (India). He stated that as there is insufficient data to train regional language models, the accuracy of VIs in these Indian languages will be very low. Preeti recommended that language experts must understand the technology well, and technologists must be appreciative of more diverse language models. This will enable them to collaborate better and develop higher language adaptability. The Niki team also indicated the difficulty they faced in finding the right technical expertise to build language compatibility in VI.

4.1. Using colloquial translations

Companies such as Niki and Navana Tech talked about conducting primary research on multiple-language use in languages including Hindi, Tamil, Kannada, Telugu, Oriya, Maithili, and Gujarati. The team at Niki claimed that like their technology, their design is scalable to multiple local languages. They do, however, recognise the challenges involved in adapting the app to local dialects. Before adding any new languages to Niki, the team familiarises themselves with the colloquial language used by the community in a specific region and for a specific use case. This helps them design responses according to the intent of the individuals using the app. Given the linguistic diversity of India, the biggest challenge that Niki faces is in hyper-localising conversations. The Niki team also realised from their focus group research that colloquial translations are more useful than literary ones. They aim to keep chat messages colloquial.

4.2. Support with iconography

Jai mentioned that it is challenging to incorporate speech to text as the individual may speak in a combination of English and Hindi. He shared that Navana Tech uses a combination of audio files and illustrative iconography, as this enables the individual to understand the flow of the app better. He believes that audio files allow for more realistic engagement and are a near-necessity until text to speech becomes more natural-sounding. This also helps in language accessibility.

4.3. Crowdsourcing voice data

CV has attempted to address the lack of language data by crowdsourcing it. In her interview, Megan explained that CV began with English as the primary input language but it aims to eventually have diverse language inputs. The initial prototypes of the platform were tested in Taipei. Feedback from individuals whose first language is not English, but who wanted to contribute to the platform, made it clear that CV must be made available in more languages. The team designed a process by which individuals could contribute in their preferred language, instead

of making the platform available in several arbitrary languages. The CV interface has a simple mechanism for choosing and adding languages. Further research by the team also revealed an audience for language preservation. Currently, CV is evolving to include lesser-known languages.

The CV team observed in its iterative design process that the quality of data collected needs to be more diverse in terms of gender, accent, dialect, and language. They organised an experience workshop to ideate on how to support multiple languages and improve the quality of voice data contributions.[30] Based on their learnings, they added dedicated language pages and community dashboards to the CV interface.

Our interviews also revealed that bigger consumer-based VI companies, like Amazon, Microsoft, and Google, are also considering including Indian languages. However, it is safe to assume that since the demographic of individuals using voice assistants and similar devices is likely to be English-speaking, it is easier for technology companies to continue catering to this consumer base by focusing solely on English.

5. Designing for accessibility

5.1. Accessibility for Persons with Disabilities

While voice is considered useful for people with visual impairments and certain cognitive disabilities such as dyslexia,[31] there are several other disabilities that VI design processes do not fully account for. These include cognitive disabilities, hearing impairments, physical disabilities, and non-normative speech patterns. Most of the designers we interviewed emphasised this perspective. Akshay also added that the first VI ever launched was meant to address accessibility concerns.

Preeti, who has worked on the design of a voice-based scribe for people with visual impairments, mentioned that though the overall design process was similar to that of other VI products, the design research for the scribe project was more detailed. Her team prepared the questionnaire for their research after speaking to people with visual impairments. She mentioned the need to let go of existing biases while working on the project. The team tested the product by conducting examinations with people with visual impairments and low vision; this helped them understand all the possible interactions between the individual and the scribe. Preeti believes that voice is fundamentally useful for people with low vision and those with low digital literacy. She highlighted that so far, she has not come across further work on leveraging VIs for accessibility.

CV takes certain accessibility concerns into account when designing its platform interface. They analysed the CV website using Lighthouse,[32] an open-source,

30 Branson, M., "Prototyping with Intention," *Medium*, accessed 10 May 2021, https://medium.com/mozilla-open-innovation/prototyping-with-intention-33d15fb147c2.

31 Nielsen, J., "Voice interfaces: Assessing the Potential," *Nielsen Norman Group*, 26 January 2003, https://www.nngroup.com/articles/voice-interfaces-assessing-the-potential/.

32 "Lighthouse | tools for web developers | google developers," *Google*, accessed 17 June 2021,

automated tool that audits for performance, accessibility, and search engine optimisation (SEO) on web pages. According to their Lighthouse score, their execution of colour contrast was not up to accessibility standards. They are now ensuring that their website matches these standards.

We noted that most developers and designers do not consider accessibility when conceptualising a VI product. Many VI apps use voice alongside visuals. These could be in the form of illustrations that suggest context or iconography that communicates options. Navana Tech shared the example of a voice-based banking app that uses icons to indicate options. They observed through their primary research that people needed further direction to navigate the audio-based app. The VI, therefore, uses icons as well as audio prompts for each button. The VI product created by Slang Labs is an added voice layer on an existing touch-based app. These products work well with the visual interface, and would not be as beneficial for the visually impaired as they do not operate on voice alone. Slang Labs also mentioned that by reducing the amount of screen-based interactions, they can help people who have motor disabilities like tremors.

5.2. Access and inclusivity

Preeti emphasised the need for access, along with accessibility, for the elderly, and for people with low digital literacy. Access here is the overall inclusivity for varying groups of people while accessibility here is being referred to specifically for persons with disabilities. There are also infrastructural concerns. The CV team also stressed the importance of the quality of the dataset. Large datasets are difficult to download, and they are working towards improving access. They are also working on creating a web app version of the website, which can be accessed on phones with lower connectivity, so that contributors can use it online and offline. It is evident that designers are identifying access concerns for people with low connectivity and low experience with technology; the elderly; and rural communities. However, many of these concerns are yet to be addressed in existing VIs. They still seem to be more popular with the English-speaking, mostly urban population – which is already well-versed with technology. While this focus on access is a welcome change among designers, there are clear gaps in their understanding of accessibility needs.

5.3. Learnings from grassroot initiatives

There are also key lessons to be learnt in the area of access from initiatives such as Avaaj Otalo and Gram Vaani, which have applied VI in rural India. These services have successfully used voice to increase the penetration of mobile phones even in the context of low literacy and internet access. We must note, however, that these initiatives are simpler, deterministic applications, and the probabilistic VI products currently in use face more complex training challenges.

Avaaj Otalo is a service developed in 2008 for farmers to access relevant and timely agricultural information over the phone.[33] Their team put together a set of guidelines for researchers designing VIs for developing regions. They

https://developers.google.com/web/tools/lighthouse.

33 Stanford, "Voice-based social media," *Awaaz.De*, accessed 23 June 2021, https://hci.stanford.edu/research/voice4all/.

recommend leveraging existing systems, ideating with people using the platform, and evaluating design choices empirically.[34] Putting their guidelines into action, Avaaj Otalo integrated with existing radio programmes and switched to explicit, directive-style prompting to avoid confusion.[35]

Gram Vaani is a social tech company founded in 2008 at the Indian Institute of Technology (IIT), Delhi.[36] One of their services, Mobile Vaani, is a social media platform for social development in rural areas.[37] Mobile Vaani uses an interactive voice response (IVR) system that allows people to call a number and leave a message about their community or listen to messages left by others.[38] The platform allows communities to discuss wide-ranging issues on culture, local updates and announcements, and government schemes, and to share other information.

To design VIs that are inclusive, it is important to ensure that the training data used covers a diverse population, so that the quality of speech recognition is improved for everyone.[39] It is important to view accessibility as beneficial for everyone and not just one sub-group.[40] Taking into account the needs of individuals with visual impairment or low vision, voice interactions should be kept brief and must allow for interruptions. The application should let the individual control the speech rate. To address cognitive disabilities, a linear and time-efficient architecture is helpful. Important information should be placed at the beginning or end, and sufficient context should be provided. For individuals with hearing impairments, the VI should provide volume control and alternatives to speech-only interactions. Designers can also consider providing transcriptions of audio files or transmission to hearing devices. Physical disabilities can be addressed by enabling the VI to capture and understand broken or varying speech. Designers must include appropriate pauses in the VI's listening. For people with non-normative speech patterns, designers must provide text-to-speech alternatives.[41]

6. Designing for privacy

One of the most common privacy concerns with VIs is that the device might be always listening and collecting data. The designers we spoke with brought up

34 Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., Parikh, T., 2009. "Experiences Designing a Voice Interface for Rural India," paper presented at *Spoken Language Technology Workshop*, 2008. 21–24. 10.1109/SLT.2008.4777830, https://www.researchgate.net/publication/224382217_Experiences_designing_a_voice_interface_for_rural_India.

35 Patel, N. et al. "Experiences Designing a Voice Interface for Rural India."

36 "About Us," *Gramvaani*, accessed 13 July 2021, https://gramvaani.org/?p=495.

37 "Community-powered-technology," *Gramvaani*, accessed 13 July 2021, https://gramvaani.org/.

38 "How Mobile Vaani Works," *Gramvaani*, accessed 13 July 2021, https://gramvaani.org/?page_id=15.

39 Pearl, C., "Using Voice Interfaces to Make Products More Inclusive," *Harvard Business Review*, 16 May 2019, https://hbr.org/2019/05/using-voice-interfaces-to-make-products-more-inclusive.

40 Kulkarni, M., "Digital Accessibility: Challenges and Opportunities," *IIMB Management Review 31*, no. 1 (2019): 91–98, https://doi.org/https://doi.org/10.1016/j.iimb.2018.05.009. (https://www.sciencedirect.com/science/article/pii/S0970389617301131)

41 Pearl, C., "Using Voice Interfaces." .

this concern as well. Many devices, especially voice assistants, use 'wake words' (for example, 'Hey Alexa' or 'Okay Google') that invoke the VI. These devices are thus always on the lookout for this wake word. This is a concern if companies start recording and storing this data on the cloud. However, most companies assure users that they place the recording on the cloud and process it only when the 'wake word' is spoken. Before that, the data is stored locally on the device. This still raises questions regarding the retention of surplus data and safeguards against leaks and sharing. One of our interviewees pointed out that many privacy implications are dependent on the design of the system architecture. For example, for Google Pixel devices, all the processing happens within the device, and no data is uploaded to the cloud. However, Google's intention behind this was not to safeguard privacy but to optimise the processing of data as it becomes much faster compared to cloud-processing devices. In 2019, there were reports that Apple was sharing a portion of the recordings from Siri with its contractors for quality control.[42] Following this controversy, Apple updated its policy to allow people to opt into sharing audio samples of their requests to train Siri.[43]

While discussing privacy in CV, Megan spoke about the need to be transparent about how the platform is utilising the information collected. She mentioned that the dashboard helps contributors control who can see their profiles. They can hide their visibility from others on the platform. The website has been created to be as malleable as possible when it comes to contributors' interactions with it. Contributors do not need to necessarily have a profile to contribute voice data. The terms and conditions agreement states that the data is being collected for research and that personal information in the form of their voices is being collected by the website.

Preeti affirmed that while she has not seen training data that reveals an individual's identity so far people should still be aware that their voice data is being processed and utilised. This is especially essential for people who do not have a lot of experience with technology and lack trust in digital services. Preeti explained this through the example of an elderly person who finds it difficult to trust online banking and hence would be unable to use a VI to access it. She spoke about the importance of notice and consent, transparency, and opt-outs. Another interviewee also made a critical point regarding notices. While voice-only devices are always listening, they do not indicate that they are listening, in contrast to webcams, which clearly indicate – through a flash next to the lens – that they are turned on. Likewise, there needs to be a similar kind of indication in VI devices. Companies such as Amazon and Google reassure people that their devices are not violating privacy – but they also do not provide any indications of continued listening.

Another key insight that Preeti pointed out concerned designing privacy notices for voice-based products. She highlighted that communicating a privacy notice

42 Hern, A., "Apple Contractors 'Regularly Hear Confidential Details' on Siri Recordings," *The Guardian*. 26 July 2019 , **https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings**.

43 "Improving Siri's Privacy Protections," *Apple Newsroom (India)*, accessed 29 September, 2021, **https://www.apple.com/in/newsroom/2019/08/improving-siris-privacy-protections/**.

through voice could be difficult, and suggested that in such cases, there needs to be an option to view the notice as text. Designing text-based privacy notices is a challenge due to the complexity and length of these texts. Even a textual notice is difficult to access and understand.[44] Moreover, communicating a privacy notice is difficult if the device does not have a screen or if it is entirely voice-based. Taking consent verbally is another complicated design task, as the quality of consent will vary depending on the use of voice biometrics, the quality and volume of audio input, and environmental noise.[45] Preeti called for an increase in overall sensitivity about data collection and use among people. She proposed an opt-out for people who do not want their data to be used for training and development purposes, while recognising that this could lead to difficulties in sourcing data for research.

Preeti remarked on larger design patterns in the voice industry: she believes that at present, the issue of privacy does not receive enough focus, and most VI companies still do not consider privacy at the conceptualisation stage. She feels the industry is still struggling to create a working system, and so privacy concerns are relegated to the end of the design process. She asserted that early conceptualisation on privacy is extremely relevant, and design is best placed to enforce and create awareness about both accessibility and ethical practices. She also pointed out the lack of guidelines for VI design. Many designers are also not familiar with data practices, privacy policies, and data protection laws. The design of privacy notices, and other elements in the interface, do not account for transparency of data collection, storage, and use. It is imperative to place VI design practice in an ethical framework and focus on privacy and transparency while designing.

Besides the policy interventions necessary to enhance privacy in VI products and services, privacy can be addressed through the interface design of VI products.[46] Designers need to provide explicit opt-ins along with enhanced notices at the time of setting up voice features. These features should not be pre-enabled. A hard 'off' switch should be available to eliminate the possibility of the device activating at inconvenient or unintended times. Creators can use prominent visual cues to practise informed consent by notifying people when a device is on and recording.

7. The future of VI design

According to Akshay, voice is just one more modality through which we interact with devices. According to him, enhanced adaptability in a VI depends on the context of the device. VIs can be more enabling for people who are not well-versed with technology. If companies begin to give importance to multi-language support, they could boost the reach of this technology. Preeti believes that VIs are going

44 Naidu, S., "Design Concerns in Creating Privacy Notices," *CIS India*, 29 May 2018, **https://cis-india.org/internet-governance/blog/design-concerns-in-creating-privacy-notices**.

45 Sigg, S., Nguyen, L. N., Zarazaga, P. P., and Backstrom, T., "Provable Consent for Voice User Interfaces," *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2020, 1–4. **https://doi.org/10.1109/percomworkshops48775.2020.9156182** (**https://ieeexplore.ieee.org/document/9156182**)

46 Gray, S., "Always On: Privacy Implications of Microphone-enabled Devices," *FPF*, accessed 8 June, 2021, **https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf**.

to become more and more relevant in the future. She foresees an initial phase of possible friction, but as familiarity with the technology increases, VIs will become more mainstream. She predicts that people who are not digitally literate – such as the elderly – will become the primary users of voice. VI is also likely to reduce the accessibility gap for people with disabilities.

Megan also speculated about the future of the CV website if it were to adopt voice-activated interaction as opposed to its current touch or point-and-click-based interface. The CV team feels that it is important to enable some sort of voice detection in the website, as this will allow for recordings to be more succinct and accurate. People will then be able to donate recordings of their own voices. The team could collect voice data to tune voice recognition on the platform as well. Speaking of the future of VI, Megan observed that soon there will be a homogenisation of VIs as has been the case with visual interfaces. She mentioned that this homogenisation is already underway with the use of wake words in all voice assistants. She wonders if the open-source data on CV can make this homogenisation look different. She believes that it can allow people to compete against the idea of what voice interfaces should look like. Megan also made a critical recommendation that designers integrate ethical practices for voice at an early stage. UI/UX has already become established, but VI is still new, and the ethical foundations can be laid early in collaboration with designers.

8. Insights and further questions

Voice is projected to be a time-saving alternative to touch-based interfaces. It is often pitched as a tool for people who cannot type or read. But present applications of VI do not demonstrably bridge the gap of access and inclusivity for marginalised and vulnerable communities. As the design processes of various VI products suggest, the homogenisation of voice-based products is already underway. It is important for design to break the 'templatisation' of interfaces and allow varying applications, formats, and structures to emerge. The absence of an inclusive and contextual design practice guideline for VI is evident in the existing VI design scenario in India.

This early stage of the technology presents an opportunity to establish an ethical design framework that focuses on inclusivity, accessibility, privacy, transparency, and openness. The focus on primary research and usability is pronounced among designers, but centring on digital rights – and not just usability – is a desperate need in design practice. Our study led us to the following critical questions on design:

- How can ethics and rights become central to design practice for VIs?
- What kind of ethical guidelines should be created for designers?
- How can design enable VIs to support multiple languages?
- How can designers be familiarised with privacy and data protection practices?
- In what ways can the design process of creating VIs focus on inclusivity and

accessibility?

These and other emerging questions must inform the growing landscape of work on VI in India. It is therefore imperative that the research, design, and development of these technologies are also shaped by a sustained and meaningful engagement with these thematics, and, most importantly, with the communities that would benefit the most from these advancements in technologies.