



MAKING VOICES HEARD
POLICY BRIEF



Making Voices Heard: Policy Brief

Research and Writing **SHWETA MOHANDAS**

Research Assistance **DIVYA PINHERO**

Review and Editing **PUTHIYA PURAYIL SNEHA, TORSHA SARKAR**

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



Shared under

Creative Commons Attribution 4.0 International license

Contents

1. Introduction	1
2. Voice interfaces in India	2
2.1. Mapping of actors in India	2
3. Key concerns/questions	4
3.1. Questions around connectivity and infrastructure	4
3.2. The need for Indian language voice data	5
3.3. Accessibility of government apps and websites	5
3.4. Emerging uses of voice and questions about privacy and data protection	7
4. Policy recommendations	7
4.1. The impetus for public-funded research	7
4.2. More funding for accessibility research	8
4.3. More clarity from PDP Bill about regulation of voice data	9
4.4. Need for more diverse voice datasets	9
4.5. Need for more funding towards community-led voice dataset collection	10
5. Conclusion	11
Appendix - Timeline of key voice interface events	13
Government initiatives	13
State initiatives	14

1. Introduction

Voice interfaces do not just provide an alternative way of interacting with a device; for people with low or no vision, they are the only way they can access the device. They allow people who are limited by text-only interfaces to navigate various aspects of their lives, by being able to access various services through voice. The development of voice technology has come a long way since the prototypes of the early 90s – they are now much cheaper, they can understand multiple languages and perform various tasks and can be integrated into different services. One of the earliest voice interfaces, the interactive voice response (IVR) system, emerged in the 1970s and is widely used even today. The technology has advanced by leaps and bounds since then, with the emergence of internet and smartphone-based voice interfaces that can be used to perform tasks of varying complexity, from setting alarms to ordering food.¹

In India, given that IVR systems have been widely deployed for service delivery in both the public and private domains, there is a growing interest in internet-based voice interfaces that can understand multiple Indian languages. These interfaces have the potential to enable people to access services that were earlier restricted by language (English) and interface (text-based systems). Although there is vast potential, some of which has been harnessed by voice interface start-ups like Niki,² there is a need to ensure that these applications are available to people with varying accessibility needs. Given the current push towards more digital-first public services (e.g., the Cowin³ platform), it is necessary to look at how accessible existing systems (such as websites) are and how voice interfaces can be integrated into them. Further, it is important to consider not just their potential but also the realities of a country where the infrastructural limitations can restrict access to services.

With respect to voice interfaces the advantages it can bring are curtailed by the unavailability of Indian language data. On the side of the individual there is also the need for better internet access to ensure that the people who will most benefit from voice interfaces can get to use them. Since voice interfaces are still in their beginning stages of uptake this is the right time to look at the challenges and possibilities towards their deployment. Additionally since the use of voice interfaces is still emerging in India, this is the right time to investigate the privacy concerns that may arise with the use of these interfaces and create policies in tandem with developments in data protection legislation.

This policy brief aims to bring into focus voice interfaces as an important policy question that needs more discussion and consideration, especially in India's quest for being a digital first nation. The policy brief also aims to shed light on the privacy concerns with respect to voice data, which seem to not get as much attention as facial data.

1 Kozuch, K., "The 30 best Alexa skills in 2021", *Tom's Guide*, 4 August 2020, accessed 3 November 2021, <https://www.tomsguide.com/round-up/best-alexa-skills>.

2 "Niki." *Niki*. accessed 9 September 2021, <http://niki.ai/>

3 "CoWIN." *CoWIN*. accessed 9 September 2021, <https://www.cowin.gov.in/>

In light of these questions, this policy brief will look at the existing companies working on voice interfaces in India, the key concerns that limit their uptake, and the policy challenges in realising their potential.

2. Voice interfaces in India

2.1. Mapping of actors in India

The voice interfaces ecosystem in India is slowly growing – a number of players provide voice services to businesses and consumers. However, when it comes to hardware-based voice interfaces, the key market players are Google⁴ and Amazon,⁵ which now support Indian languages spoken in different accents and integrate Indian apps (through Alexa skills) such as Ola.⁶ To understand the state of voice interfaces in India, we mapped 27 voice interface developers in India, in terms of type of voice interface, client, sector, languages, and data collection. This revealed a few trends, based on the type of individuals they cater to, the sectors that use voice technologies extensively, and the most preferred languages, that could provide insights on the uptake of voice interfaces in the country.

More business-facing interfaces than consumer-facing

Although only Google and Amazon offer device-centric voice assistants,⁷ a variety of mobile apps and smart devices incorporate voice interfaces. In our study of voice interfaces in India (including voice assistants), we were able to find only two apps – Niki⁸ and Vokal⁹ – that provided services to individuals directly. The remaining provided these services to businesses, which in turn offered them to the individual. Therefore, there are only a few general voice interfaces in India, as most are voice bots and chats developed for specific business purposes.

Sectors that use voice interfaces

The banking and finance sector features the highest number of chatbots and voice bots. These voice interfaces help individuals access information about their accounts as well as the services offered by the bank. HDFC Bank,¹⁰ Andhra

4 Akolawala, T. "Amazon Echo Dot Tops Smart Speaker Sales in India in 2020, Google Home Mini, Mi Smart Speaker Follow: techARC." *Gadget360*, 18 February 2021, <https://gadgets.ndtv.com/smart-home/news/amazon-echo-dot-most-sold-smart-speaker-india-2020-google-home-mini-mi-smart-speaker-techarc-report-2373059>

5 Akolawala, "Amazon Echo", *Gadget360*, 18 February 2021

6 "Ola", *Amazon*, <https://www.amazon.in/ANI-Technologies-Pvt-Ltd-Ola/dp/B075NGT52M,18> February 2021

7 A program on a device that can listen and reply to voice commands.

8 "Niki." *Niki*.

9 "India's Largest Vernacular Question & Answers Platform in Indian Languages", *Vokal*, accessed 20 October 2021, <https://www.vokal.in/>

10 Ani, "HDFC's Banking CHATBOT 'Eva' Now Compatible with Google Assistant", *Business Standard*, 20 December 2017, accessed 20 October 2021 https://www.business-standard.com/article/news-ani/hdfc-s-banking-chatbot-eva-now-compatible-with-google-assistant-117122000272_1.html.

Bank,¹¹ and Kotak Bank¹² all use voice interfaces to interact with customers. The second-most popular sector for voice interfaces is e-commerce, as apps such as Big Basket,¹³ Grofers,¹⁴ and Flipkart¹⁵ use or have proposed to use voice interfaces. Some local governments also use voice interfaces services (offered through their websites or apps), such as the Rajkot Municipal Corporation and Pimpri Chinchwad smart city.

Languages

Hindi was the first and is still at times the only Indian language other than English available on virtual assistants and voice bots. Out of the 27 companies we mapped, all of them provided voice features in English and Hindi. Both Google Assistant¹⁶ and Alexa¹⁷ can understand and speak Hindi now. However Google and Amazon are yet to launch the voice assistant in other Indian languages. The other languages that follow Hindi in popularity are Tamil, Bengali, and Kannada.

Accessibility

Voice interfaces provide accessibility support for individuals who are unable to see the screen or understand the text. However, no applications other than Google and Amazon claim to provide accessibility features. Amazon Echo's website lists the various features that customers with vision, hearing, mobility, and speech accessibility needs could use.¹⁸ Google Home provides accessibility features that allow the individual to control appliances and entertainment, make phone calls, broadcast messages, and manage tasks in addition to its voice assistant.¹⁹

11 Hans News Service, "Andhra Bank Unveils AL Chatbot Abhi", *The Hans India*, 15 July 2019, accessed 20 October 2021, <https://www.thehansindia.com/business/andhra-bank-unveils-al-chatbot-abhi-546877>.

12 "Kota accessed 20 October 2021k Mahindra Bank Launches Keya – The First Voicebot in Indian Banking", *Kotak Mahindra*, 2 April 2018, <https://www.kotak.com/content/dam/Kotak/about-us/media-press-releases/2018/kotak-mahindra-bank-launches-keya-the-first-voicebot-in-indian-banking-02042018.pdf>.

13 Rangarajan, K., "Voice to Cart: Powering your Ecommerce App with Voice", *Slang Labs*, 6 October 2020, accessed 20 October 2021, <https://www.slanglabs.in/blog/voice-to-cart-powering-your-ecommerce-app-with-voice>.

14 Limited, J. H. T., "How Haptik Automated Grofers' Customer Support in Less than 48 Hours", *Haptik*, accessed 20 October 2021, <https://www.haptik.ai/resources/case-study/grofers-case-study>.

15 Schwartz, E. H., "Indian E-commerce Giant Flipkart Expands English and HINDI Voice Search Platform-Wide", *Voicebot.ai*, 4 March 2021, <https://voicebot.ai/2021/03/04/indian-e-commerce-giant-flipkart-expands-english-and-hindi-voice-search-platform-wide/>.

16 Tech Desk, "Google Assistant Now in Hindi: Here's How to Activate and Use", *The Indian Express*, 15 March. (2018, March 15). <https://indianexpress.com/article/technology/social/google-assistant-now-available-in-hindi-heres-how-to-activate-and-use-5098595/>.

17 Singh, M., (2019, September 18). "Amazon's Alexa Now Speaks Hindi", *TechCrunch*, 18 September 2019, <https://techcrunch.com/2019/09/18/amazon-alexa-hindi-india/>.

18 "Accessibility Features for Alexa", *Amazon*, accessed 20 October 2021, <https://www.amazon.in/gp/help/customer/display.html?nodeId=202158280>.

19 "Accessibility features on Google nest or home devices", *Google Nest Help*, <https://support.google.com/googlenest/answer/9286728?hl=en>.

Privacy

Voice interfaces have presented significant privacy concerns. The ‘always on’ feature of Google Home and Amazon Echo have attracted media attention for recording conversations even when the voice assistant was not summoned.²⁰ With respect to the voice interface companies that we analysed, it was difficult to assess privacy commitments as most developed voice interfaces for businesses, which then provided this service to customers. Hence, how these business-facing companies collect and store voice data is neither public nor addressed in their privacy policies. However, most companies developing voice interfaces have a publicly accessible privacy policy and terms and conditions. Some user-facing companies specified that they use, process, and store/retain voice data, whereas others failed to specify how they handle voice data. Although related laws, such as the Information Technology Act, 2001,²¹ Sensitive Personal Data/Information Rules, 2011,²² and Personal Data Protection Bill, 2019,²³ do not require companies to disclose if voice data is being processed, privacy policies that provide this information could help people make an informed choice of what they talk about or record on these applications.

3. Key concerns/questions

3.1. Questions around connectivity and infrastructure

The *Indian Telecom Services Performance Indicators* report published by the Telecom Regulatory Authority of India (TRAI) in 2020 revealed that as of 31 December 2019, there were 29.83 percent of rural internet subscribers in the country.²⁴ According to the license service area data that was provided, the states that had the lowest number of internet subscribers per 100 persons were Jammu and Kashmir (16 persons per 100) and Bihar and Uttar Pradesh (21 per persons per 100). The highest was Delhi (98.97 persons per 100).²⁵ The *Digital India* report of 2019 stated that India had 504 million active Internet users who were five years and above as of November 2019. In terms of usage frequency, nearly 70% of the internet-enabled population in India are daily users.²⁶ This data shows that although the number of internet users is large, the number of internet subscribers is still very

20 Guardian News and Media, “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>.

21 The Information Technology Act, 2000.

22 Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011.

23 The Personal Data Protection Bill, 2019. http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

24 Ministry of Communications, “Internet Connectivity in Rural India. Unstarred Question No. 594 To Be Answered On 16th September, 2020, 2020”, 16 September, <http://164.100.24.220/loksabhaquestions/annex/174/AU594.pdf>

25 Ministry of Communications, “Internet Connectivity in Rural India”.

26 Nandita Mathur, “India now has over 500 million active Internet users: IAMAI”, *Mint*, 05 May 2020 <https://www.livemint.com/news/india/india-now-has-over-500-million-active-internet-users-iamai-11588679804774.html>

low. This is due to the fact that in most households one smartphone is used by multiple people in the house.²⁷

Thus, although several voice interfaces are being developed to cater to India's multilingual nature, they are limited in their reach until they can also be accessed by those without an internet connection or with intermittent access to the internet. A study on the use of IVR systems to support job searches by low-income domestic workers in India concluded that "for computer-based systems to solve developing-world problems often require significant work above and beyond an implementation of the technology."²⁸ Hence, although voice interfaces may benefit those limited by language and digital literacy, the proposed benefactors of the technology may be hindered by a lack of access to other key infrastructures.

3.2. The need for Indian language voice data

The developers and researchers interviewed for this study obtained voice training data from multiple sources such as open-source databases, at competitions set up by Google or Microsoft²⁹, user-generated anonymised data, databases like Mozilla's Common Voice, and hours of speech data recorded by professionals such as news readers or voice artists.

A common issue that the developers we interviewed highlighted was the scarcity of voice data in Indian languages. They noted that although there is now some data in Hindi and Indian English, there are several low-resource languages³⁰. If data on them was available, voice interfaces could be developed to help people access services in these languages via their phones. The Indian scenario is particularly challenging due to the scarce availability of open-source voice data. Initiatives such as Indic TTS,³¹ a consortium created and funded by the Government of India, have been making an effort to record data in various regional languages. However, finding the datasets and applying them to products is still a challenge. Another barrier that was highlighted was that technology giants such as Google and Amazon, with their abundant data and other resources, create an imbalance between start-ups that have to collect data from scratch and multinationals that already have data and systems in place.

3.3. Accessibility of government apps and websites

A 2012 study of 7,800 Indian government websites, which assessed their design

27 Dr Rajesh Tandon, "One Device Households", *The Times of India*, 17 July 2020. <https://timesofindia.indiatimes.com/blogs/voices/one-device-households/>

28 Smyth, T. N. (2010). Where There's a Will There's a Way: Mobile Media Sharing in Urban India. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://www.researchgate.net/publication/221514114_Where_there's_a_will_there's_a_way_Mobile_media_sharing_in_urban_india

29 Through our interviews we understood that developers and researchers alike were able to get voice data in different languages through participating in competitions organised by Google and Microsoft.

30 A resource language means a language that does not have or has only few data resources. This makes it even more difficult to develop machine-learning based systems for these languages.

31 "Indic TTS", *Indic TTS*, <https://www.iitm.ac.in/donlab/tts/>. accessed 3 November 2021

against the Web Content Accessibility Guidelines (WCAG) 2.0, revealed that 1,985 websites failed to open and the remaining 5,815 had some form of accessibility barrier, including a lack of non-text alternatives to text making them inaccessible.³² A more recent study, published in 2021, revealed that many government websites ranked low in usability, many did not follow WCAG 2.0 accessibility guidelines, and none of the 164 websites tested was fully accessible on mobile.³³ The study also stated that even in 2019, 62% of the websites they tested did not pass any MobileOK checks³⁴.

More recently, one of India's COVID-19 measures, the Arogya Setu app, and its mandatory use by citizens, have been debated strongly as it requires a phone and a working internet connection to access, apart from several concerns related to privacy and data protection. The app was also flagged by persons with visual or hearing disabilities and disability rights activists, for failing to meet accessibility standards. The Union Social Justice and Empowerment Ministry informed the Ministry of Electronics and Information Technology (MeitY) and the National Informatics Centre (NIC), that the app lacked accessibility features.³⁵ A report by activist Anjee Agarwal stated that the visually impaired people who tested the app found it inaccessible, which amounted to a violation of the Rights of Persons with Disabilities Act, 2016. According to the report, the "the screen reader in the app did not announce the purpose of all controls or the type of control, whether a link or button". This means that the screen reader did not specify what tasks or options the app could provide, and it did not differentiate between whether there was a link or a button to enter the service. The app also did not mention the page numbers on the website, which would mean that the individual might miss out on the next page or the screen reader would keep on reading the pages on a loop. Additionally, on the "Your status", "COVID updates", and "E-Pass" tabs in the app, "the screen reader was not announcing the control type, so individuals did not know these were interactive tabs."³⁶ In May 2020, an IVR service was set up within Arogya Setu to aid people who had feature phones and landlines.³⁷ However, there were no known improvements with respect to the accessibility of the Arogya Setu app itself.³⁸ The Supreme Court, while examining issues relating to COVID-19 management, emphasised the need to conduct a disability audit for the CoWIN website and Arogya Setu to ensure that they were accessible.³⁹

32 "Accessibility of Government Websites in India: A Report", *The Centre for Internet and Society India*, 2012, <https://cis-india.org/accessibility/accessibility-of-government-websites-in-india>

33 Agrawal, G., Kumar, D., and Singh, M., "Assessing the Usability, Accessibility, and Mobile Readiness of E-government Websites: A Case Study in India", *Universal Access in the Information Society* (2021): 1–12.

34 The Mobile Ok checked by W3C performs various tests on a web page to determine the level of mobile-friendliness. The tests are defined in the mobileOK Basic Tests 1.0 specification. A web page is considered mobileOK only when it passes all the tests.

35 Nath, D., "Mandatory Arogya Setu App Not Accessible to Persons with Disabilities", *The Hindu*, 2 May 2020, <https://www.thehindu.com/news/national/coronavirus-mandatory-aarogya-setu-app-not-accessible-to-persons-with-disabilities/article31489933.ece>.

36 Nath, D., "Mandatory Arogya Setu", *The Hindu*.

37 "Arogya Setu IVRS", <https://www.mohfw.gov.in/pdf/AAROGYASETUIVRS1921.pdf>

38 Nath, D., "Mandatory Arogya Setu", *The Hindu*.

39 "In Re: Distribution of Essential Supplies and Services During Pandemic", In The Supreme Court Of India Civil Original Jurisdiction, 2021, https://main.sci.gov.in/supremecourt/2021/11001/11001_2021_35_301_28040_Judgement_31-May-2021.pdf

Hence, for India, there is a need not just for the implementation of voice interfaces, but also for other accessibility measures to be introduced to enable every person to benefit from the digital world.

3.4. Emerging uses of voice and questions about privacy and data protection

Despite their several benefits, particularly in terms of enabling individuals to access the internet and services in their own languages, voice interfaces present significant privacy concerns. Researchers and civil society have raised concerns regarding the potential for misuse and harm that might stem from storing and processing immense amounts of voice data. These recordings may have been made without the person's knowledge and may reveal extremely sensitive information – its most benign consequences range from targeted ads to being profiled based on what the device processes. One of the emerging concerns is how this voice data could be shared with law enforcement agencies and the consequences of such sharing.

Additionally, there seems to be a growing interest in using voice as a biometric identifier, especially in the banking sector. A report by Kaizen Secure Voiz detailed the benefits of voice biometrics such as fraud detection, rural banking, and remote verification.⁴⁰ However, the report also recognised the challenges that would come with switching to voice biometrics, such as user confidence (making the person confident in using their voice, and confidence in the safety of using voice), training of staff and capacity of the organisation implementing it. Some banks that have looked at implementing voice recognition are Citi Bank, HSBC, and Standard Chartered Bank, which seem to have implemented this in India as well. However, implementation of voice biometrics should also come with adequately addressing the privacy and data protection responsibilities of collecting and processing biometric data (in this case, voice data).

4. Policy recommendations

4.1. The impetus for public-funded research

A project at the scale of Indic TTS was possible because of the availability of government funding. There is a need for increased public funding of voice-based research in Indian languages to allow researchers and developers to create localised voice interfaces. However, one of the issues with publicly funded research is that open access research and databases require continuous funding to be sustainable. Unlike private for-profit companies, public-funded research or datasets are usually made available free of cost.

In the case of Indic TTS, the datasets are all open access and can be used by start-ups and researchers alike; the objective is to allow more projects and research questions to stem from the existing work and to foster an environment

40 Kulkarni, A., "Indian Banking – Adoption of Voice Biometrics", 2020, <https://kaizenvoiz.com/wp-content/uploads/2020/11/Kaizen-white-paper-for-Indian-banking-ver-6.1.pdf>

of collaborative, open-access research. Our conversation with Indian start-ups working on voice revealed that they mainly relied on datasets from large companies such as Google for their voice data which these startups either purchased or won as a part of challenges organised by the companies. While initiatives such as Indic TTS do exist, there seems to be a disconnect between researchers and start-ups working on voice in Indian languages. One way to foster innovation is to have public-private partnerships that would not only ensure that the research is relevant to the needs of the industry but also that the industry benefits from the research and the development. Another way to boost further research on voice interfaces specifically for Indian languages could be to set up a system of royalty-free licensing for start-ups, where once the start-up starts to seek commercial value from the datasets, the license can be changed to a revenue-sharing model.⁴¹ This system would ensure that the researchers receive feedback after deploying the research in the real world and the start-ups can test and verify the same. The above system could be beneficial for start-ups that do not have the capacity or the funding to set up public-private partnerships.

4.2. More funding for accessibility research

There has been a worrying decline in budgetary allocations towards schemes for persons with disabilities in India. The budget for the Scheme for Implementation of Persons with Disabilities Act (SIPDA) was cut from INR 315 crore in 2019–20 to INR 252 crore—a 20 percent reduction—in 2020–21. Similarly, the budgetary allocation for both research on disability-related technology and the National Institute of Mental Health and Rehabilitation in FY 2020–21 was missing, compared to INR 20 crore in the previous year.⁴² The assistance for Disabled Persons for Purchase (ADIP)/Fitting of Aids and Appliances has also not seen any increase in allocation of funds and stands at INR 230 crore for the entire population of persons with disabilities.⁴³ The national pre-budget consultation held by the National Centre for Promotion of Employment for Disabled People (NCPEDP) emphasised the need to incentivise companies that make accessibility products (both hardware and ICT) by providing rebates and concessions.⁴⁴ As recently as August 2021, the Standing Committee on Social Justice and Empowerment (Department of Empowerment of Persons with Disabilities) expressed that the progress of the Accessible India Campaign, launched in 2015, has been “rather slow”.⁴⁵ The campaign aims to make accessing services such as transport, public spaces, tourist places, international airports, railway stations, and information and communication technology in India easily accessible for persons with disabilities.

41 Ali, F. and Mohandas, S., “The Compulsive Patent Hoarding Disorder”, *The Hindu*, 24 March 2017, <https://www.thehindu.com/opinion/op-ed/the-compulsive-patent-hoarding-disorder/article17617888.ece>.

42 Ali, A, “Scheme for Implementation of Persons with Disabilities Act (SIPDA) Has Been Reduced from Rs 315 Crore”, *Indian Express*, 30 January 2021, 20, <https://indianexpress.com/article/lifestyle/life-style/pandemic-has-hit-persons-with-disabilities-hardest-union-budget-should-address-their-concerns-7167840/>.

43 Ali, “Scheme for Implementation”, *Indian Express*.

44 Ali, “Scheme for Implementation”, *Indian Express*.

45 Outlook, “Progress of Accessible India Campaign Rather slow: Parl Panel.” *Outlook*, 6 August 2021, <https://www.outlookindia.com/newscroll/progress-of-accessible-india-campaign-rather-slow-parl-panel/2136476>.

4.3. More clarity from Personal Data Protection Bill about the regulation of voice data

The Indian Personal Data Protection Bill, in its 2019 version, defines biometric data as “facial images, fingerprints, iris scans, or any other similar personal data resulting from measurements or technical processing operations carried out on physical, physiological, or behavioural characteristics of a data principal, which allow or confirm the unique identification of that natural person.”⁴⁶ Although voice data has not been explicitly mentioned in this definition, it could fall under the processing of the physical characteristics of the data principal, which are unique to each individual. Biometric data is also considered sensitive personal data; hence, requirements such as the need for explicit consent to collect, share, store, and use such data, and the prohibition of processing such data outside India, are being established under the PDP Bill. The Bill also mentions an additional category of data fiduciaries called significant data fiduciaries,⁴⁷ which have more duties and responsibilities based on the volume of data processed, the sensitivity of that data, risk of harm, and the use of technologies. The Bill also states that if in the opinion of the Data Protection Authority, data processing by a fiduciary carries risk of significant harm to any data principal, then that fiduciary will be tasked with all or some of the responsibilities of a significant data fiduciary.⁴⁸

Although voice data can be considered biometric data and is in the ambit of sensitive personal data, it needs to be clearly included in the definition of biometric data in the Personal Data Protection Bill. This is becoming increasingly crucial as several services are including voice data, and certain institutions, such as banks, have also begun to use voice biometrics as a form of recognition.⁴⁹ This would mean that a person’s voice can be linked to their financial information, thus linking two types of sensitive information to a service or a company.

4.4. The need for more diverse voice datasets

Hindi was the first and is still the only Indian language available on some voice interfaces, both for virtual assistants and voice bots.⁵⁰ The mapping of voice interfaces in India revealed that out of the 27 companies covered, all provided voice features in English and Hindi. Hindi is also the Indian language of choice used in the most popular voice assistants, Amazon’s Alexa and Google Home. One of the reasons why Hindi is used so widely in voice interfaces is because it is one of the few high-resource languages in India with multiple voice datasets.

46 Section 3(7), The Personal Data Protection Bill, 2019, http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

47 Section 26(1), The Personal Data Protection Bill, 2019, http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

48 Section 26(3), The Personal Data Protection Bill, 2019, http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

49 “Making Voices Heard: Mapping Actors,” *Making Voices Heard*, accessed 02 February 2022, <http://voice.cis-india.org/mapping-actors.html>

50 Ahaskar, A. “Voice biometrics are Cleverer Now, But Still Need More Work”, *Mint*, 6 February 2020, <https://www.livemint.com/technology/tech-news/voice-biometrics-are-cleverer-now-but-still-need-more-work-11581011267941.html>.

Private companies develop voice interfaces for the most popular or most spoken languages as they are more profitable. The creation of voice databases for lesser spoken languages is left to volunteer-based organisations and public-funded projects. There is a need to look at how voice interfaces can be encouraged to support more Indian languages. Although there are several IVR systems in different Indian languages, their scope is limited to particular questions and answers.

4.5. The need for more funding towards community-led voice dataset collection

When a handful of companies are made responsible for collecting, processing, and creating speech datasets, the choice of languages is based on popularity and commercial viability. Even these systems, which work with data-rich languages, often fail to understand accents and voice modulations that are not present in the datasets.⁵¹ Additionally, as these datasets are owned by large corporations, they are protected by non-disclosure agreements, contracts, and Intellectual property rights. However, as stated by one of our interviewees, “language technology is an entry into a digital world”,⁵² especially in a country with widespread inequity in access to digital infrastructure. Community based voice data collection initiatives are attempting to bridge this gap by assembling open-access datasets.

In India, the Indic TTS consortium was created with the goal of making information available in regional languages. However, due to the scale and the resources required, the consortium could only collect data for 13 Indian languages. Common Voice⁵³ (a global open-access dataset of voice recordings in multiple languages that can be used to train speech-enabled applications) is another great example of how a community-driven and open-access collection of voice data can lead to a more inclusive internet. Common Voice now has over 13,905 hours of voice data across 76 different languages as of July 2021.⁵⁴ This was achieved by not only making the language available on Common Voice, but also by making the website available in that language. When adding a new language, the community localises 85% of the website, so that the local language community can easily navigate it without relying on English. When a language is active on the site, it is up to the community to present 5,000 sentences in that language that can be recorded. This indicates two things to Common Voice: one, that there is an active language community that can provide language recordings, and two, that the barrier to get the language into Common Voice is fairly low.⁵⁵

A recent example of community-driven voice data collection initiative was for

51 WP Company. “The Accent GAP: How Amazon’s and Google’s smart SPEAKERS Leave Certain Voices Behind”, *The Washington Post*, 19 July 2018, <https://www.washingtonpost.com/graphics/2018/business/alexas-does-not-understand-your-accent/>.

52 Interview, Anonymous, in person, Bangalore, March 3 2020.

53 “Making Voices Heard: Common Voice Case Study,” *Making Voices Heard*, accessed 02 February 2022, <http://voice.cis-india.org/case-studies/common-voice.html>

54 “Common Voice by Mozilla.” *Common Voice*, accessed January 4, 2022, <https://commonvoice.mozilla.org/en/datasets>.

55 “Making Voices Heard: Common Voice Case Study,” *Making Voices Heard*, accessed 02 February 2022, <http://voice.cis-india.org/case-studies/common-voice.html>

Kinyarwanda, a widely spoken language in Rwanda with over 12 million speakers.⁵⁶ In 2019, Mozilla and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) co-hosted an ideation hackathon in Kigali to create a data corpus for Kinyarwanda. A result of the hackathon was Digital Umuganda, a volunteer-driven start-up with the aim to build digital infrastructure such as voice data. Despite the challenges faced in mobilising the community, including poor access to mobile phones and the prohibitive cost of data plans, the startup managed to collect 1, 211 hours of Kinyarwanda voice data from a diverse set of over 420 contributors.⁵⁷ They are planning to set up a hybrid model involving both on-site and off-site recording through in-person and online events and by mobilising an expanding pool of volunteers. They hope that this process will hasten the contributions and be capable of withstanding any unforeseen circumstances.⁵⁸ One of the ways India could look at increasing the language reach of voice interfaces is to learn from the example of Rwanda, and have initiatives that bring together government agencies, startups and student volunteers to create voice data in languages from each state and community. In India, the CGNet Swara project is a great example of how voice can be used to help individuals of a particular community. CG Net Swara⁵⁹ is an Indian voice-based online portal that serves as a platform to discuss issues related to the Central Gondwana region in India. People in the forested regions of Chhattisgarh use it to report and share news in the Gondhi language through a phone call. Gondhi, which is spoken by almost 2 million people in different parts of northern and western India, can only be written by 100 people.⁶⁰ This is where a voice-based interface for people to report stories and listen to them in Gondhi helps. The portal is accessible through mobile phone or desktop; people can also listen to news reports and stories by giving a missed call. The CGNet Swara website helps the community preserve their language by participating online and via phone.

For government initiatives and private players, studying the approaches and best practices adopted by projects such as Common Voice and CGNet Swara could help expand their work and thereby the reach of the internet. Initiatives and projects such as these help reduce the language barrier, improve access to infrastructure and public services, provide services to people across languages and digital literacy, help people learn new skills, enhance adherence to privacy and accessibility guidelines, and help preserve low-resource and indigenous languages.

5. Conclusion

Voice interfaces have immense potential to make the internet accessible to people who are limited by purely text-based interfaces. However, in the case of India, there needs to be greater research and policy discussions on the challenges, possibilities, and dangers of voice interfaces. Currently, the discourse around

56 "How Rwanda is making voice tech more open", *Mozilla Foundation*, 16 September 2020, <https://foundation.mozilla.org/en/blog/how-rwanda-making-voice-tech-more-open/>.

57 "How Rwanda is", *Mozilla Foundation*.

58 "How Rwanda is", *Mozilla Foundation*.

59 "Welcome to CGNet Swara", *CG Net Swara*, <http://cgnetswara.org/>, 16 September 2020

60 Majumdar, M., "This Indian Language Can Be Written by Only 100 People", *The Hindu*, 31 March 2018, <https://www.thehindu.com/society/this-indian-language-can-be-written-by-only-100-people/article23384526.ece>.

voice interfaces has been sporadic, with announcements that certain government services will be accessible through voice but without much follow-up.⁶¹ There is also a need to look at how public and private services can be made universally accessible to people with varying accessibility needs. Additionally, accessibility should not be the sole responsibility of the government; private companies and start-ups should assess how accessible their services are, conduct user research, and have people with various accessibility needs on their teams. Along with the possibilities that voice interfaces bring, there is also a need to consider the privacy concerns and potential harm that they can cause. Given the possibility of widespread use of voice biometrics, there is a need to ensure that voice data is not used for profiling. Voice data should be given the same significance as facial recognition data, and how such technology is being deployed should be examined.

To sum up, voice interfaces and voice data have immense potential in India; however, greater attention needs to be given to development of policies directly related to these technologies. This would ensure that their full potential is reached without harming the individual using it or creating language erosion.

61 For example there have been numerous news reports about the Umang App being enabled with multilingual voice support, however at the time of writing this policy brief there have been no reports of its implementation and use.

Appendix – Timeline of key voice interface events

Government initiatives

Owing to the language diversity and low literacy rate of India, a number of studies and initiatives have studied IVR systems, including Avaj Otalo⁶² (a service for farmers to access relevant and timely agricultural information) and Sehat ki Vaani⁶³ (for the management of Type 2 diabetes and maternal health). In the year 2020 the Aarogya Setu IVRS service was set up to check the spread of COVID-19 and help people detect symptoms.⁶⁴

Although there have been no policies yet that directly regulate and encourage the uptake of voice interfaces, a few government initiatives encourage the development and adoption of voice technologies. One such initiative is the Indic TTS platform, sponsored by DeiTY, Ministry of Information Technology. The goal of this initiative is to develop a corpus of text-to-speech data in Indian languages. The consortium includes some of India's premier institutions, and the researchers have been able to collect a total of 40 hours of speech data in 13 Indian languages so far.

Umang

In 2018, the Indian government announced the inclusion of a multilingual voice search feature in the Unified Mobile Application for New-age Governance (UMANG) platform. Developed by the Ministry of Electronics and Information Technology and National e-Governance Division, UMANG provides easy access to an array of government services via smartphones and on their website.

Although the UMANG website and app are currently not enabled with voice technology, government tenders published in February 2020 reveal that the government intends to create a conversational chatbot and AI-based voice assistant.⁶⁵ They also emphasised the need to include more Indian languages to ensure inclusivity and widespread adoption. More recently, in 2021, the Ministry of Electronics and IT selected Senseforth AI as the firm to provide these services on the Umang platform. The first deployment will include voice bots and chatbots in English and Hindi, after which the service will expand to Malayalam, Tamil, and Telugu.⁶⁶

62 "Voice-based Social Media", *Awaaz.De*, <https://hci.stanford.edu/research/voice4all/>, 16 September 2020

63 Kazakos, K., Asthana, S., Balaam, M., Duggal, M., Holden, A., Jamir, L., Kannuri, N. K., Kumar, S., Manindla, A. R., Manikam, S. A., Murthy, G. V. S., Nahar, P., Phillimore, P., Sathyanath, S., Singh, P., Singh, M., Wright, P., Yadav, D., and Olivier, P., "A Real-time IVR Platform for Community Radio", proceedings of the 2016 CHI Conference on Human Factors in Computing System, 2016. <https://doi.org/10.1145/2858036.2858585>

64 "Aarogya Setu IVRS", <https://www.mohfw.gov.in/pdf/AAROGYASETUIVRS1921.pdf>

65 "Invitation to Bid for Appointment of Partner Agency (Vendor 5)", Umang, https://www.meity.gov.in/writereaddata/files/tender_upload/UMANG%20RFP_AI-Bot.pdf

66 Agarwal, Surabhi, "Move Over Alexa and Siri, 'Hey Umang' to Deliver Govt Services Through Voice Commands Soon", *Economic Times*, 05 April 2021, <https://economictimes.indiatimes.com/tech/technology/move-over-alexa-and-siri-hey-umang-to-deliver-govt-services-through-voice-commands-soon/articleshow/81916003.cms>

'AIRAWAT' (AI Research, Analytics, and Knowledge Assimilation platform)

In January 2020, NITI Aayog released an approach paper to set up India's first AI-specific cloud computing infrastructure, called 'AIRAWAT' (AI Research, Analytics and Knowledge Assimilation) platform. In the AI strategy paper released in 2018, Niti Aayog stated that the cloud-based platform would support AI-based speech recognition and natural language processing for research and development.

State initiatives

The Tamil Nadu government, under the Tamil Nadu e-governance agency (TNeGA), has expressed interest in creating a voice user interface in Tamil for availing of government services. Santosh Mishra, the chief executive officer of the Tamil Nadu e-Governance Agency (TNeGA), also stated at the summit on Responsible Artificial Intelligence for Social Empowerment (RAISE) that the voice interface would ensure that "the keyboard barrier to access technology is lifted".⁶⁷ With respect to existing voice services, the Madurai Kavalan app is a good example – the app allows individuals to record voice-based police complaints. The user study revealed that the voice API helped older people and those who found it hard to type and navigate the menu to access the app. The emergency feature also provides a 'women's safety' option, where a woman can either press the emergency button or request for help by saying "help me" in English or Tamil, which would trigger an SOS response.

The Bangalore Electricity Supply Company Ltd (BESCOM) has reportedly been working with the Machine and Language Learning (MALL) Lab at the Indian Institute of Sciences (IISc) to develop an "artificial intelligence-powered voice bot to attend to customer calls". This voice bot is being designed to allow people to seek answers to basic queries in English and Kannada.

⁶⁷ Shivakumar, C., "TN Agency to Develop First Voice User Interface by Government in Tamil", *New Indian Express*, 9 October 2020, <https://www.newindianexpress.com/states/tamil-nadu/2020/oct/09/tn-agency-to-develop-first-voice-user-interface-by-government-in-tamil-2208051.html>.

