

VOICE RECOGNITION AND DEVELOPMENT IMPACT

CASE STUDIES AND LANDSCAPE SURVEY FOR AFRICA & INDIA

UNIVERSITY OF WASHINGTON STRATEGIC ANALYSIS,
RESEARCH & TRAINING (START) CENTER

REPORT TO THE BILL & MELINDA GATES FOUNDATION

PRODUCED BY: AMLAND B, CHINE A, PANDEY S, & BADSHAH A.

DECEMBER 2020



**START
CENTER**

STRATEGIC ANALYSIS,
RESEARCH & TRAINING CENTER

Department of Global Health | University of Washington

Executive Summary

BACKGROUND

The expansion of broadband coverage and improved affordability of smartphones have increased internet penetration in low- and middle-income countries (LMICs). Despite increased internet access, illiterate and semi-literate people still face significant challenges in navigating text-heavy websites and applications. However, the rapid development and improvement of voice technologies offer the potential for illiterate and semi-literate people to use internet-enabled resources more effectively.

The University of Washington's Strategic Analysis and Research Training Program (START) team was tasked by the Bill and Melinda Gates Foundation's (BMGF) Financial Services for the Poor team to research voice technology use cases that are driving development in Africa and India and assess the potential of voice to impact development. The project goals were to understand each company's use of voice-based technologies, business models and markets for voice technology, and barriers to increasing the use of voice technology by illiterate or semi-literate people.

METHODS

The START team conducted interviews with subject matter experts and an extensive landscape review. Technology subject matter experts included eight voice technology subject matter experts, including graduate students, researchers, technology start-up founders, and professors. The landscape review was performed to collate information on voice technology and conduct a current state assessment of the industry with a focus on synchronous, free-form voice technology usage in Africa and India.

RESULTS

The START team produced eight case studies of enterprises operating in Africa and India. These enterprises use different business models to commercialize voice technology and reach customers that speak diverse languages. Through these case studies and interviews with sector experts, the START team found critical success factors for and significant obstacles to the use of voice technologies for improving the lives of semi-literate and illiterate people in Africa and India. The START team also made recommendations to expand voice technology use and development impact based on the research findings from case studies, interviews, and literature. Listed below are several findings of this research project.

Key Findings

- Methods of collecting voice data and of training algorithms for low resource languages are substantially different from the methods used for high resource languages. Across case studies, we found a software company that partners with its customers by utilizing the customers' own field workers to gather voice data in exchange for reduced prices on voice-based software. We also found that companies can adapt their voice algorithms to produce high accuracy with only hundreds of hours of voice data.
- The eight case studies had a wide range of business models and target customers. Four of the companies analyzed use natural language processing (NLP) technologies to help financial services companies reduce costs and expand financial services to their customers. However, only one company, Navana Tech, is focused on building its enterprise platform and NLP technology so that its customers (e.g. banks) can reach semi-literate and illiterate individuals.
- Barriers limiting the use of voice technology can be grouped into four major categories listed here from most to least important: infrastructure; cognitive learning and digital literacy; motivation; and cultural and privacy issues.

Recommendation

Voice technology offers a tremendous opportunity for illiterate and semi-literate individuals to benefit from internet technology, but significant barriers must first be overcome for voice to significantly impact development. Recommended actions to overcome these barriers fall into three categories: technology, business, and policy. The START team's recommendation is for BMGF to focus its effort on these three categories in the following concrete ways:

- **Technology/Business:** Utilize local partnerships to build speech corpora cost-effectively. As illustrated in the Navana Tech case study, local companies that would benefit from voice technologies may already have field workers who are able to gather voice data from target populations.
- **Technology:** Develop and leverage a suite of tools such as culturally relevant icons, images, and pre-recorded audio in addition to voice recognition. Combining voice technologies with these other tools will enable more individuals to better engage with mobile applications and internet-based services.
- **Technology/Policy:** Create open-source APIs and/or focus on incentivizing the development of more APIs to reduce the cost of telecom API integration for developers.

- **Policy/Business:** Accelerate the smartphone install base through such methods as asset financing so that developers have a larger obtainable market.
- **Policy:** Reduce voice data rates by partnering with telecom companies to establish better agreements on infrastructure sharing and with governments to enact policies such as universal broadband access.
- **Policy:** Build partnerships with governments that also want to reach and provide services to semi-literate or illiterate individuals. Voice recognition technology offers a promising way for governments to reach their underserved constituents.

Contents

- Executive Summary 1**
 - Background 1
 - Methods 1
 - Results 1
- Key Findings 2
- Recommendation 2
- Contents 4**
- Introduction 6**
 - Project Overview 6
 - Summary of the Problem 6
 - Voice Technology 7
 - Research questions 7
 - Methodology 8
 - Primary research 8
 - Secondary Research 8
- Case Studies 9**
 - Navana Tech 9
 - Teller 11
 - Aja.la Studios 13
 - AwezaMed 14
 - Niyo Bank 16
 - YeLo Bank 17
 - Kotak Mahindra Bank 17
 - HDFC Bank 19
- Discussion 21**
 - Barriers to Voice Recognition 21
 - Infrastructure 21

Cognitive learning and digital literacy.....	23
Motivation	24
Cultural and Privacy Issues.....	24
Opportunities to Accelerate Voice.....	26
Outlook on the Future of Voice in Africa and India	28
Overall assessment.....	28
Potential Impact of Voice Recognition in Development	29
Conclusion.....	30
Limitations	30
Appendix.....	31
Appendix A.....	31
References.....	34

Introduction

Project Overview

The intention of this report is to describe use cases of organizations in Africa and India using voice recognition technologies to drive development. Specifically, the START team focused on organizations using natural language processing (NLP)- based voice technologies in the digital financial services sector. Additionally, this report outlines key barriers preventing the proliferation of voice technology to impact development, describes opportunities for the Foundation to accelerate voice technology adoption, and gives an outlook on the future of this sector in India and Africa.

Summary of the Problem

Access to the internet is increasing dramatically around the world, especially in LMICs. Declining prices of entry-level smartphones, decreasing data rates, and the increasing availability of mobile broadband are bringing more people online for the first time. It is estimated that 300 million people will come online by 2025 in Sub-Saharan Africa alone(1). Those coming online primarily access the internet through smartphones. Smartphones in rural India increased from 9% in 2015 to 25% in 2018, and in Sub-Saharan Africa it is estimated that the number of smartphones will increase by 440M from 2017 to 2025(2)(1).

Many mobile internet users, especially women, are “stuck on application islands” lacking the digital skills to go beyond applications like Facebook – GSMA 2015

While access to the internet through smartphones is increasing, those who come online for the first time and are semi-literate or illiterate suffer from low internet usage and/or low quality of usage. The primary reason for low or poor internet usage is twofold. First, smartphone applications and internet content are highly text-mediated and full of complex information. Yet, over a billion people in the world are illiterate or semi-illiterate, and these people are disproportionately located in low-income countries(3). Second, traditional literacy skills (i.e., reading, writing, and counting) are insufficient in effectively using the internet and must be complemented with digital literacy skills. Digital literacy skills include, but are not limited to, digital safety, digital identity, digital security, digital rights, etc. In Africa, ‘34% of mobile users who are aware of the internet state digital skills are an important barrier’(4). Additionally, a prior study of 84 countries demonstrated that basic digital skills are present in less than 50% of the population in many countries(5).

As internet access in LMICs continues to increase over time, the barriers shift from access to utilization and the quality of that utilization. Low or no literacy is a significant obstacle to both use

and quality of use. Voice technology is one potential solution to enable semi-literate and illiterate individuals in LMICs to benefit more from the internet.

VOICE TECHNOLOGY

Voice technology offers users the ability to use their voice to control devices and applications for many uses such as: completing daily tasks like setting a timer; searching for information on the internet; making purchases; transferring money; and texting. Voice technology has rapidly expanded and improved over the last decade. Google, for example, reduced its speech recognition word error rate by over 18% from 2013 to 2017(6). Similar improvements in accuracy have been attained by other technology companies, such as Apple and Amazon.

The focus of this report is on voice technologies that use machine learning and natural language processing (NLP) to help the user find information and complete tasks on mobile or desktop applications. Illustrative examples of this are Amazon's Alexa and Google's Assistant because both can recognize a huge number of words, understand many user intents, and complete various tasks.

The advantage of advanced NLP-based voice models for semi-literate and illiterate individuals is that these people could speak and interact in their native languages to control their devices and applications, enabling greater participation in the digital economy and digital world. Advanced voice recognition and voice assistants have the potential for increasing financial inclusion, especially as developing countries shift more financial resources and services online. In India, the government has pushed for moving towards a cashless society and many stakeholders are invested in increasing digital transactions from three billion per month to one billion per day(7). Semi-literate and illiterate people will need a solution that enables them to participate in an increasingly digital financial sector.

Research Questions

The research questions which guided this research are listed below:

01

1A. How are companies or organizations using voice-based natural language processing technologies to drive development and reach low-literate or illiterate customers in Africa and India, especially in the financial services sector?

1B. What are these organizations' business models, market size, customers, products, and strategies?

02

2A. What are the barriers to further developing and implementing voice-based natural language processing services in these regions for these low-literate users?

2B. How are these barriers prioritized in order of importance?

Methodology

PRIMARY RESEARCH

The primary research process involved interviews with eight subject matter experts including graduate students, researchers, technology start-up founders, and professors. A combination of structured and discovery interviews was conducted for each interviewee. The interviewees' areas of expertise, with a contextual focus on semi-literate populations, cut across various fields such as voice-based technology, human-computer interaction, language and linguistics, financial technology, user interface design, artificial intelligence, and financial inclusion. The broad range of expertise provided a holistic view of voice technology usage in Africa and India within financial services and beyond. A full list of interviews, including names, positions/organizations, and key insights is included in the Appendix.

SECONDARY RESEARCH

The preliminary approach was secondary research to collate existing information on voice technology usage in Africa and India as well as conduct a current state assessment of the industry those regions. Another motivation for conducting desk research prior to primary research was to identify subject matter experts from relevant use cases who could provide additional insights on the current landscape. The secondary research involved an extensive literature review of over 30+ publications with an initial emphasis on synchronous, free-form voice recognition technology usage in the geographic regions of interest. Based on the availability of use cases, the research scope was subsequently revised to focus on the broad application of voice technology in digital financial services in Africa and India. In this process, significant attention was given to local languages spoken in the geographic regions of interest as this project is targeted at the use of voice recognition technology by the poor who are unlikely to be literate or able to communicate in English. The search strategy and resources used are detailed in **Table 1**.

Table 1 - Search strategy

Population	Illiterate, semi-literate
Concept	Voice recognition technology
Geography	India, Africa
Databases	University of Washington Library, Google Scholar, Google general search engine, referrals from subject matter experts

Case Studies

Navana Tech

Company Background: Navana Tech is a software company founded in Bangalore by two brothers, Raoul Nanavati and Jai Nanavati in 2018(8). In 2017, the brothers realized that a paradigm shift in user interface design was required to give low-literate Indian users access to services in diverse industries such as banking, healthcare, education, and e-commerce. This led to the development of Navana Tech’s flagship product, Bharat SDK, an artificial intelligence software that bridges the literacy gap on user interfaces for applications(9).

NAVANA TECH AT A GLANCE
Sector: Enterprise Software
Type of Company: For profit
Geographic Region: India
Languages: Hindi, Tamil, Bengali, Marathi, Punjabi, Telugu, Gujarati, Bhojpuri and Odia
TAM: 304M end users

Navana Tech’s end user base is approximately 500,000 and is expected to reach 2 million by the end of 2021 based on new customer engagements(9). They had 9+ employees and 4+ paying customers in 2019(8).

Target Market: Organizations whose applications are used by semi-literate mobile phone users or first-time smartphone users living in semi-urban or rural India(8). Navana Tech’s customer base includes banks, government agencies, and non-profits(9).

Total Potential Addressable Market (TAM): With 1.17 billion mobile phone subscriptions in India and a literacy rate of 74% in 2018, the estimated total addressable market for Navana Tech is over 304 million end users(10) (11).

Product Description and Operating Model: Navana Tech’s product, Bharat SDK, is an artificial intelligence software that enables businesses to add features to digital assistants on the applications they have developed(8). These features include animation of buttons on mobile applications and step by step explanation of application functionality to users in their native languages using pre-recorded voice notes(9). The product combines text-to-speech and audio files to provide voice-based assistance and to create descriptive images that enhance customer experience on mobile applications(9). The product also uses voice recognition technology to facilitate the process for making inquiries. The voice recognition model is industry-specific with financial services being the current focus; however, the company plans to expand the voice recognition model to cater to additional industries in the future(9). Bharat SDK is an additional layer on top of applications and does not require modifications to a company’s existing application interface(8).

Through grant funding and voice data collection efforts by their customers who have employees in the target regions, Navana Tech has been able to overcome the primary limitation faced by companies that use voice recognition technology, which is the unavailability of a robust voice database in different languages that is required to train voice recognition models(9). Navana Tech has a dedicated team and infrastructure for data collection and refining based on dialects(9). 10 Indian languages and 40 dialects are currently supported on Bharat SDK(9).

Ujjivan Small Finance Bank is one of Navana Tech’s first customers(9). The bank has a mobile banking application which is available in five languages and can be downloaded on Android, Apple, and Windows app stores(9). Customers are able to use the voice bot from Bharat SDK to request for services on the mobile application as well as receive instructions on how to use the application(9). Requests that are out of scope or not able to be completed over voice are directed to a customer service representative(9).

Business Model: Navana Tech uses a software-as-a-service model to generate revenue(8). The company currently charges its customers based on monthly or annual active users; however, their pricing strategy is still evolving(9).

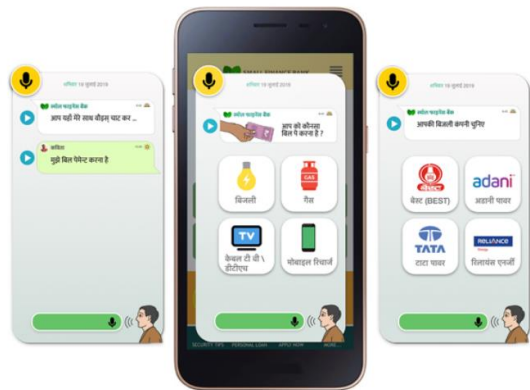
PRODUCT

BHARAT SDK VOICE DRIVEN NAVIGATION

Domain-specific, hyper localized speech to text to help users communicate, search, navigate and enter data using voice.

Building in partnership with Vivek Seshadri and Kalika Bali from Microsoft Research India.

Language set : Hindi, Tamil, Kannada, Bengali, Marathi, Punjabi, Telugu, Gujarati, Bhojpuri and Odia



Source: Navana Tech Website

Teller

Company Background: Teller was founded out of a project in Columbia Business School in 2016. After Facebook made their API public, there was an increased interest in leveraging this newly open technology to develop applications. Inspired by the rise in number of users in messaging apps, the founder, Sidharth Garg, combined messaging apps with existing NLP to build a channel for communication between users and financial institutions. The team initially launched their model in Brooklyn, United States by creating chatbots for credit unions. Their model did not gain traction in the market, possibly because, as Sidharth speculates, the market in the US already has multiple resources to learn about financial services and multiple ways to connect to these sources(12). Recognizing that the product may not be a great fit for the US, the team shifted their focus to developing countries, mainly to Francophone countries in East and West Africa through the help of DFS lab(12).

TELLER AT A GLANCE

Sector: Enterprise Software in Financial Services

Type of Company: For profit

Geographic Region: Francophone African countries

Languages Supported: French

TAM: 141M end users

Product Description: Teller, through an AI enabled chatbot model, allows financial institutions to communicate with their users in existing messaging applications, including WhatsApp, Facebook messenger, and SMS, as well as web chat. By utilizing these platforms, Teller enables financial institutions to communicate with their customers in a cost-effective way, especially with customers in areas where travel to a local branch is not feasible due to geographical or business constraints. Teller also has an educational project called Finance 150 which aims to increase financial literacy of interested users. Finance 150 consists of 150 pre-recorded explanations of financial terms that users can access. These recordings are available in Hindi, Swahili, English, French, Arabic, Chinese, Spanish, and Portuguese.

Voice Development: Teller has also received grants from the UN Capital Development Fund and BMGF to develop voice technology for two projects. The goals of these projects were to expedite payments and improve communications between workers and organizations(12).

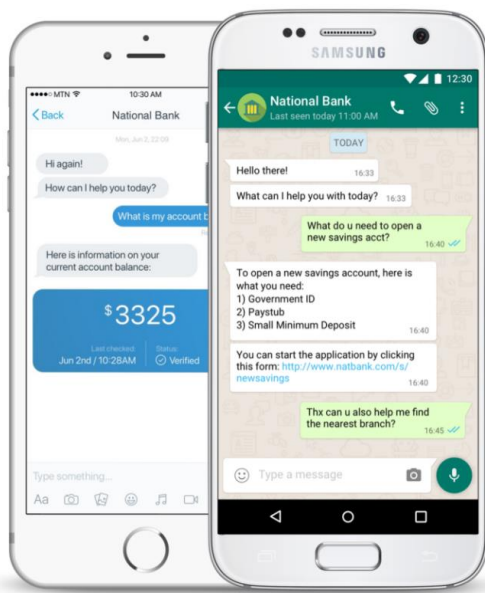
1. UN Capital Development Fund (UNCDF): Teller supported a UNCDF program that provides job training and skills training programs in Africa. This program works with training cohorts in industries, such as clean energy construction, to help interested workers find jobs. The UNCDF often struggled communicating with the participants about the timing of trainings, payment schedules, and other information about the program. Teller was invited to automate these communication processes, and Teller used automated voice messaging to

communicate information about the program and send updates regarding the payment amount based on the spreadsheet provided by the UNCDF(12). This voice messaging system synthesized the spreadsheet information and translated it into voice, which then generated messages that workers accessed through a keypad-driven menu system, Teller was able to expedite payments and build faster communication channels with the participants.

2. Rungwe Smallholders Tea Growers Association (RATGCA), Tanzania: Teller collaborated to send automated information to small tea farmers. These smallholder tea farmers sell their product to the cooperative and receive paper receipts which are then used as proof for receiving payment. Teller automated the process by providing an automated voice-call that contained the details of the receipt (the quantity of tea leaves sold, data, and expected payment date)(12). Teller sent the farmers a voice-based notification about the quantity of tea leaves delivered as well as the expectations on payment amount, all of which simplified the process and reduced the consequences of losing the physical receipt. The project was delivered to 20-30 farmers.

Operating and Business Model: Teller charges a licensing fee to institutions that use their platform and caps the number of messages that the institution can send without a fee increase. There is also an option to build customized solutions for clients with more robust needs.

Product



Source: Teller Website

Aja.la Studios

Company Background: The founder, Abake Adenle, received a PhD in theoretical machine learning before working for eight years at Morgan Stanley in its quantitative finance department. The decision to pivot towards developing voice technologies occurred in 2016 after learning the voice space in Africa had virtually no incumbent companies at the time(13). Around 2016, Adenle received a grant to work on two African languages, and then in 2018 she received another larger grant to work on more languages(13). Aja.la Studios focuses on building an enterprise voice platform that businesses in several African countries can use to automate their processes and to reach more customers who might be unbanked or less urban.

AJA.LA STUDIOS AT A GLANCE

Sector: Enterprise Software

Type of Company: For profit

Geographic Region: West/East Africa

Languages Supported: Kiswahili, Yoruba, Hausa, Igbo, Kinyarwanda. 7 more in the pipeline.

TAM: 266M speakers covered by 5 supported languages

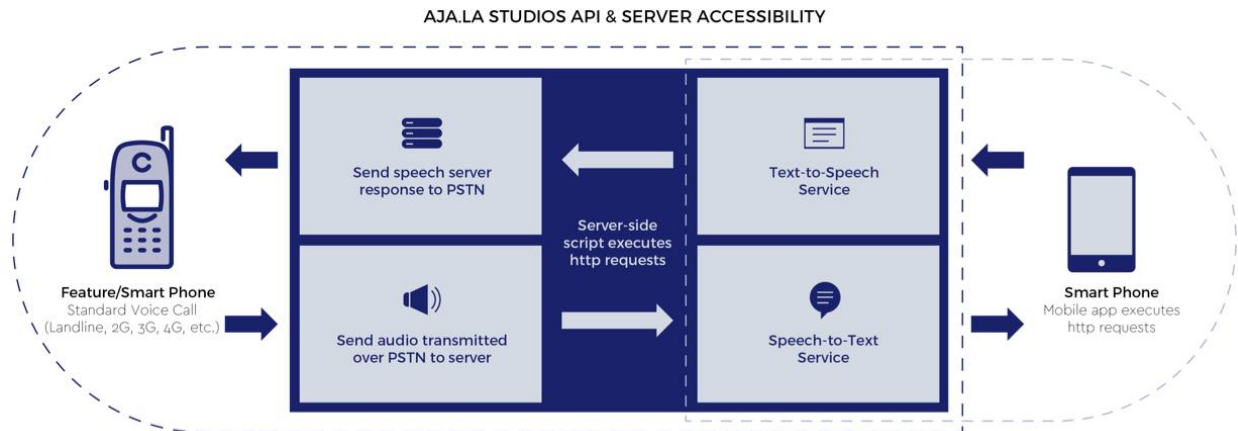
These grants were for training models in low-resource data contexts that could attain high accuracy with less data than traditionally required. For the first two languages, Adenle used the grants as a research opportunity to train high accuracy models and build an architecture that can be applied to other languages. According to her, the Yoruba model has better accuracy than what is reported in academic research for other languages(13). While this learning and development of a standard architecture does reduce some of the resources needed to train new languages, it still takes a lot of time to train a model for a new language because the manual text processing and coding is considerable. Currently, the company has about 2,000 hours of voice data across 11 different languages(13).

Target Market: The primary market for Aja.la Studios is the financial services sector in Nigeria, Rwanda, and other West and East African countries. The goal is building a broad enterprise platform so that its business customers can use Aja.la's API to develop their own software. Over 266 million speakers are covered by Aja.la's five currently supported languages, but this will greatly increase as the company has seven more languages in the pipeline.

Product Description, Operating and Business Model: Aja.la builds an ASR and speech synthesis technology platform for African enterprises. The primary cost-driving activities Aja.la performs in house are: collecting and processing voice data; building the NLP model; customizing the platform for individual enterprise customers; and marketing and selling efforts(13). Notably, the company does not pay external companies to collect voice data but hires and manages people to do that

instead. Quotes from external companies to collect voice data for 10 languages were around \$1 million and Adenle believed she do the same internally for less money(13).

The primary value proposition identified so far by Aja.la is the cost-cutting potential of voice technologies when used by financial institutions to automate customer service. Financial institutions are the primary potential revenue generating service for the company, as it attempts to expand into other sectors and identify other use cases, such as: insurance, healthcare, agriculture, telecoms, and recruiting.



Source: Aja.la Studios Website

AwezaMed

Company Background: Founded by Glenn Stein in 2014, the company first focused on overcoming the language barriers faced by patients and providers in South Africa’s health system. South Africa has 11 official languages, and it is common for patients and providers to encounter language barriers when interacting. In the early 2000s, the South African government funded voice data collection to build a speech corpus that contains 50 hours of speech data for each of the 11 official languages(14). Leveraging this resource, AwezaMed is focusing on machine translation, voice-enabled search, speech-to-text, and text-to-speech in healthcare and education.

AWEZAMED AT A GLANCE

Sector: Health and education tech

Type of Company: Nonprofit

Geographic Region: South Africa

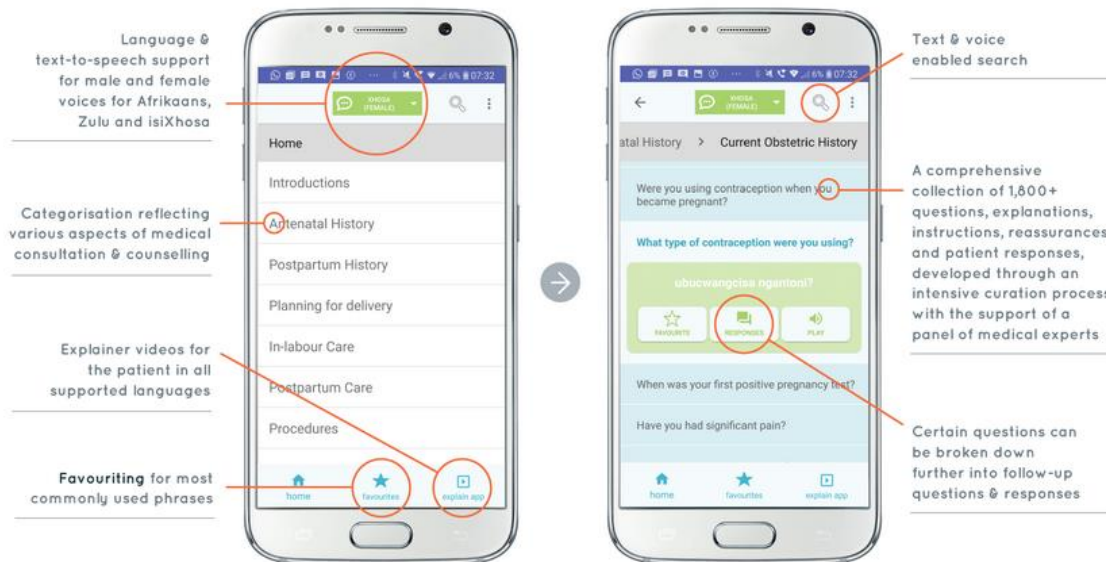
Languages Supported: 11 official South African languages

TAM: 27K healthcare providers, 12.7M students

CASE STUDIES

Target Market: Only 8.1% of South Africans speak English at home. The most common languages are isiZulu, IsiXhosa, and Afrikaans(15). In healthcare, AwezaMed is a free software application that providers and patients can download to and use on their smartphones. In South Africa, anywhere from 20 to 22 million people have smartphones and 27,000 of those are healthcare providers(16). Aweza’s machine translation and speech recognition software is also used in education. The company has an education product to help students understand and learn in the language most comfortable to them. The education market has 12 million students in schools and 700,000 in vocation training programs(17) (18).

Product Description, Operating and Business Model: AwezaMed is a smartphone application available for Android operating systems that offers speech-to-speech translation, specializing in maternal healthcare. The technology pipeline begins with a large content repository of health-related questions and answers, automatic speech recognition for speech to text, machine translation into the target language, and then a text to speech synthesizer in male or female voices. The company owns and operates its smartphone app, which is offered pro-bono with funding by AwezaMed founder Glenn Stein and by the South African Departments of Sports, Arts, and Culture and Science and Innovation.



The AwezMed Application

Source: AwezaMed Website

Niyo Bank

Company Background: Founded in 2015, Niyo Bank is a digital bank that exists only in the cloud. Niyo must still partner with traditional banks to comply with Indian banking regulations, so it has set up partnerships with IDFC First Bank, DCB Bank, and YES Bank. Niyo currently has over 800 employees, over 1.5 million customers, and received \$49 million in funding(19).

Target Market: Niyo has a few products aimed at different target markets. Its Bharat Payroll solution is designed for blue-collar workers earning between Rs 8,000 to Rs 12,000 (\$109 to \$163) per month. The total market size for blue-collar workers in India is approximately 170 million(20). In the next two years, the company is aiming to add 5 million more customers to its zero balance Bharat Payroll solution. At the price of Rs 200 annually, the total addressable market is Rs 34 billion per year, or \$461 million(19).

Product Description and Operating Model: While Niyo offers a global card to enable its higher-income travelling customers to pay in Rupees anywhere that Visa is accepted, its Bharat app and card is aimed specifically at blue-collar workers. Niyo works with employers to open salary accounts for their workers so that employers can automatically deposit wages into workers' Bharat mobile wallet. From there, blue-collar customers can view transaction history, transfer money, attach receipts, track savings, request cash advances, apply for small ticket loans, and even buy gold. Niyo offers a voice assist feature in the app available in multiple languages that is powered by Navana Tech's Bharat SDK platform. However, Niyo is only using Navana's pre-recorded contextual audio help for its users as of this point.

NIYO BANK AT A GLANCE

Sector: Fintech

Type of Company: For profit

Geographic Region: India

Languages Supported: 10 languages

TAM: 170M blue-collar workers, \$461M per year

YeLo Bank

Company Background: Founded in 2019, YeLo is a neo-bank existing only in the cloud. The purpose of the company is to improve financial inclusion for gig economy and blue-collar workers. The company completed a seed round of investing in late 2019 for an undisclosed amount.

Target Market: The market YeLO targets consists of 100 million customers who earn less than Rs 30,000 (\$408) per month and typically work for small companies or as gig workers(21). However, the number of gig workers in India is about 3 million and the number of people earning less than Rs 30,000 is projected to continue declining through 2025(22).

Product Description, Operating, and Business Model: YeLo hopes to offer savings, remittance, and micro-credit services to its customers in a branchless, completely digital way. Customers would also receive a prepaid card that is linked to their app and online account. The funds raised in the seed round are being used to build a local language-based experience necessary to reach this customer segment. YeLo uses Amazon Web Services for its banking architecture and security(23).

YELO BANK AT A GLANCE
Sector: Fintech
Type of Company: For profit
Geographic Region: India
Languages Supported: Unknown
TAM: High estimate – 100M customers

Kotak Mahindra Bank

Company Background: Kotak Mahindra Bank is a subsidiary of a full-service financial institution, the Kotak Group. The bank obtained their banking license from the Reserve Bank of India (RBI) in February 2003 and has four key business segments: consumer banking, corporate banking, commercial banking, and treasury, and the bank serves both retail and corporate customers in urban and rural India(24). Kotak has an active customer base of over 23 million and over 71,000 employees work for the Group.

They have 1,600 branches and 2,519 ATMs distributed across all the regions in India(25). In 2018, it launched the first AI-powered bilingual voice bot for the Indian banking sector, Keya, and its advanced successor, Keya 2.0, in 2019(26) (27).

Target Market: Based on the wide range of services Kotak Mahindra Bank provides, their target markets are businesses and middle-to high-income individuals.

KOTAK AT A GLANCE
Sector: Financial Services
Type of Company: For profit
Geographic Region: India
Languages Supported: Hindi, English
TAM: 516M customers

Total Potential Addressable Market: India has a population of 1.3 billion people, poverty rate of 13.4%, with 43.63% of the population speaking Hindi, and a high likelihood that the middle- and upper-class speak English(28) (29) (30) (31). Based on this, the Bank's total potential addressable market is estimated to be 516 million customers.

Product Description and Operating Model: Keya uses automatic voice recognition, natural language processing and text-to-speech technology to supplement the bank's traditional interactive voice response (IVR) system. It is also integrated with the bank's mobile banking helpline(26). Keya 2.0 has taken their digital banking services a step further with additional features such as conversational banking enabled by artificial intelligence. Keya 2.0's database has over 100 use cases and over 150,000 utterances, both of which have enhanced call accuracy and reduced call duration(27). The voice bot effectively determines the purpose of customer service calls in over 75% of the cases and can complete over 11% of all customer service without any human involvement. Keya 2.0 is also able to singlehandedly resolve 1.7 million customer calls monthly which allows the bank to allocate representatives to more advanced queries, leading to a higher level of efficiency and customer satisfaction at their call center. Keya 2.0 is accessible on the bank's net banking and mobile banking platforms as well as their website.

Business Model: Kotak Mahindra Bank uses the traditional business model for banks to generate revenue which includes services charges, fees, and interest from loans. Keya 2.0 is available to all their customers at no additional cost.



Source: Kotak Mahindra Website

HDFC Bank

Company Background: Housing Development Finance Corporation (HDFC) was established in Mumbai, India in August 1994 and started operations on January 1995. The bank began as a commercial bank and received authorization from the Reserve Bank of India (RBI) to operate as a private sector bank in 1994. The Bank has three main business segments: retail banking, wholesale banking, and treasury(32). They are one of the leading private banks in India with a branch network spanning 5,430 locations and 15,292 ATMs in

2,848 sites. The bank has a customer base of over 50 million and approximately 117,000 employees(32). In 2017, HDFC launched an Interactive Robotic Assistant to improve in-branch customer service and launched India's first AI-powered banking chatbot, EVA(33).

Target Market: Based on the wide range of services HDFC Bank provides, their target market is businesses and middle to high income individuals.

Total Potential Addressable Market: India has a population of 1.3 billion people, poverty rate of 13.4%, with 43.63% of the population speaking Hindi, and a high likelihood that the middle- and upper-class speak English(28) (29) (30) (31). Based on this, HDFC's total potential addressable market is estimated to be 516 million customers.

Product Description and Operating Model: HDFC's Interactive Robotic Assistant (IRA) 2.0 is an advanced descendant of the IRA which has built-in voice recognition technology and combines the capabilities of their chatbot (EVA) with speech recognition technology to communicate with customers in a branch(34) (35). IRA 2.0 facilitates customer service in the branch by responding to questions about the bank's services. It also uses voice-based navigation to direct customers to appropriate service stations within the branch.

EVA is also available on Amazon Alexa and Google Assistant to provide customized voice banking services using NLP. This gives HDFC's customers the option of interacting with EVA from any location and without touching any screens. The languages supported included Hindi and English; however, the bank has plans to include additional languages.

Business Model: HDFC uses the traditional business model for banks to generate revenue which includes services charges, fees and interest from money loaned to customers. Voice-enabled EVA and the Interactive Robotic Assistant (IRA) 2.0 are available to all HDFC customers at no additional cost.

HDFC BANK AT A GLANCE

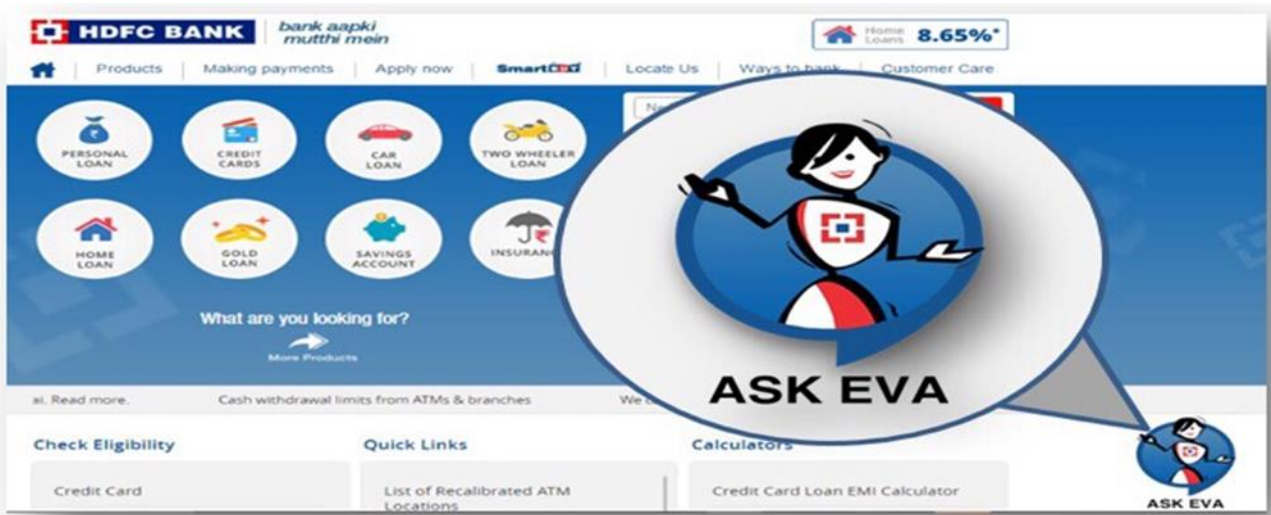
Sector: Financial Services

Type of Company: For profit

Geographic Region: India

Languages Supported: Hindi, English

TAM: 516M customers



Source: HDFC Website

Discussion

Of the eight case studies presented, only one company (Navana Tech) is building NLP-based voice technologies specifically for the poor and semi-literate or illiterate populations. In Africa and India, the barriers to using voice for development are present and significant. Furthermore, the approach to building consumer and enterprise voice technologies in the U.S. and other developed countries has relied on several key factors that are less present in LMICs, including:

1. Massive datasets for relatively few (~30) languages
2. Comparatively lower data rates (average mobile data rates in the U.S. were 0.53% of monthly income compared to 6.8% in Sub-Saharan Africa)(36) (37)
3. High prevalence of non-tonal languages and low prevalence of tonal languages which require special treatment(38)
4. Low prevalence of code-mixing, or speaking more than one language in the same utterance
5. High smartphone penetration (80%) and broadband coverage(39)
6. Greater cultural trust in technology along with the required skills to use it

These six factors are not all present to the same degree in LMICs, and absent in some cases, especially when the target market is semi-literate or illiterate populations. Discussed below are the most significant barriers to building, implementing, and using NLP-based voice recognition technologies for development in these contexts.

Barriers to Voice Recognition

INFRASTRUCTURE

Speech Data Corpora: One of the major barriers for the development and adoption of voice technology in the LMICs is a lack of data corpora required to train existing voice models. Multiple languages are spoken in India and countries in Africa, which further complicates the infrastructure and incentive required to build data corpora. The dataset required to train models are “much smaller for languages such as Swahili and Hindi and relatively non-existent for a long tail of languages spoken in much of Africa and Asia”(6). For example, in India, there are 22 constitutionally recognized languages and almost 100 mother tongues are reported(40). Similarly, in Africa, there are more than 2000 languages spoken. As such, the data corpora required to expand voice technology in these geographies requires simultaneous development of data corpora in multiple languages.

This lack of language representation is reflected in the voice corpora that are open source as well. For example: the Open Language Archives Community (OLAC), is an initiative to create a unified

means of searching online databases of language resources for linguistic research(41). For countries like the United States, they have catalogued 15,050 and the data catalog for English language is 84,129. For comparison, Nigeria has the highest catalog available among African countries with only 4,294 data catalogs. Goimai with 214 catalogues had the highest language catalog available within Nigeria.

Countries like India have recognized that diversity in their languages and lack of voice corpora to incorporate that diversity is hindering their citizens' ability to fully participate in technological advancements and human-machine interaction. Multiple steps are being taken at the government level to counter that. For example, the Department of Electronics and Information Technology in India initiated the Technology Development of Indian Languages program. One of its major initiatives is to develop a corpora project under the Indian Language Corpora Initiative (ILCI) program. Its current phase includes a corpus of 12 languages: Hindi, Bangla, Punjabi, Oriya, Marathi, Gujrati, Konkani, Urdu, Tamil, Telugu, Malayalam, and English. The program's second phase is to expand to another 11 languages that are recognized by the Indian constitution(42). As a part of its Research Open data initiative, Microsoft India has also opened its Indian language speech corpus, which is available in Telegu, Tamil, and Gujrati (43). Hence, investments in building data corpora are taking place in India which may build a stronger infrastructure for voice technology development.

"For automatic speech recognition (ASR) technology to impact the developing world more significantly, a number of hurdles must first be overcome. One of these hurdles is the collection or expansion of ASR corpora. Currently [2014], only about 30 of the world's 6900 languages have significant quantities of digitized data, and much of this data is only in textual form" – Kleynhans N.T. "Efficient Data Selection for ASR" (2015)

Code Mixing and Accents: To add to the complexities in the data infrastructure, NLP voice technologies in developing countries are difficult because of code-mixing - where speakers use more than one language in a single conversation. The voice technology and the AI machine are not developed and trained to capture such nuances in languages and identify these patterns in their speech recognition(44) (45).Currently, Microsoft Research India has Project Melange, which aims to understand and build tools around code mixing(46). There are also concerns about accents and different phonetic pronunciation which further complicates the process to training the machine, even in cases where a data corpus might be robust.

High cost of transactions and Setup: High cost of using APIs for voice relative to SMS and WhatsApp messages has also hindered the utilization of voice technology. Africa Talking is one of the companies in Africa enabling developers to access telecom APIs across SMS, voice, payment,

and airtime services(47). It currently operates in 20 countries in Africa, supporting almost 5,000 organizations(48).

The prices, however, that this major API provider charges are relatively high. For example, in Kenya, the price of 'voice regular outgoing' is 400% more expensive than average price of the highest available SMS bulk plan. Similarly, in Uganda and Nigeria, the price of regular voice outgoing is 1000% and 302% more expensive than similar SMS plans(49). While the cost difference decreases for voice SIP outgoing and incoming per minute in comparison to SMS, the cost for using voice is still significantly higher for developers. Moreover, Africa Talking provides the voice options to only three of these countries while SMS options were available to 20 countries. More details about the pricing can be found on the Africa Talking website.

Comparing prices across data types, WhatsApp messaging costs only \$0.005 flat rate for all countries, which partially explains the growing popularity of WhatsApp among developers. The cost discrepancy as well as availability of API for interested developers contributes to the barriers in the development of speech recognition technology.

COGNITIVE LEARNING AND DIGITAL LITERACY

Internet penetration is high in many rural areas of India, but a significant portion of the rural population are still novice users. Of the rural internet users in India, only 14.9% access the internet through a desktop computer while 99% access the internet through mobile phones only(50).

According to a 2019 GSMA survey, "33% of mobile users in South Asia stated digital skills as the single most important barrier to internet use"(4). In such scenarios, the low rate of digital literacy exacerbates the difficulties in adopting a technology, which in turn creates both a barrier to and an opportunity for voice technology.

Lower literacy is also associated with a lower cognitive ability. According to an interview with Indrani Medhi Thies, a principal researcher in the Technology for Emerging Markets Group at Microsoft Research India, "researchers are beginning to look at how non-literacy is not just about the inability to read text, but is correlated with cognitive skills such as the ability to transfer learning in video-based skills training, and the ability to navigate the hierarchical organization of information" (51). This illustrates how the cognitive ability of semi-literate and illiterate users demands simpler interfaces and innovative methods to make these technology work for them. Therefore, voice will likely need to be integrated with other modalities that cater to users with lower literacy. Using large icons and culturally-relevant images in coordination with voice could help integrate these semi-literate and illiterate users with access to the digital world(44).

Internet and mobile users with lower literacy tend to engage only in with certain parts of a smartphone application and will not utilize its entire functionality. Navana Tech co-founder, Raoul Nanavati, detailed its WhatsApp user segmentation model based on internal research they gained from users with literacy levels comparable to an eighth grade student(9). The model segments users into the following five buckets based on complexity of the WhatsApp function: consuming what others sent, forwarding content to other users, navigating photos and gallery, curating and sharing content through photos, and using voice notes and creating content.

Moreover, barriers in using voice technology start as early as the search and download phase. Semi-literate and illiterate smartphone users in rural India typically download applications by enlisting help from local 'experts' (children, local shopkeepers, etc.)(9). Navana, however, found that a list of 10-15 pre-recorded audio steps brought similar results as enlisting help from others.

Overall, barriers to voice-based technology and applications for semi-literate and illiterate individuals stem from a lack of digital skills and reduced cognitive ability that manifest in a variety of ways, such as the inability to download new apps, to use the full functionality of an app, and to create and share content.

MOTIVATION AND TRUST

One of the barriers to the expansion of voice technology is the lack of motivation from users. Even if the underlying infrastructure progresses and the willingness of developers to create applications utilizing voice technology increases, the wide-spread implementation of voice technology will not be possible without more motivation from users. Most of the internet users in rural areas use their devices for basic communication and entertainment. From a motivation standpoint, "it is much harder to make users excited about financial services and how voice technology can be leveraged to expand their access to financial services"(44).

Furthermore, trust is a major factor that needs to be addressed when focusing on financial services. For someone with relatively low income, their preferred form of financial payment is still cash since it is a tangible form of money. Thus, moving from a tangible form of money to a digital transaction that is operated through voice is a major leap. During the interview with Thies, she mentioned "workers would often ask for a receipt as a proof of transaction and it was difficult to explain the concept of digital currency"(44). Thus, trust can also hinder the adoption of voice technology.

CULTURAL AND PRIVACY ISSUES

In India, it is common to have one smartphone per household(9). In such scenarios, concerns of privacy can also inhibit the use of voice to make financial decisions. Moreover, in developing nations, women face gender disparity in accessing technology and making household decisions,

including financial decisions. Several studies conducted in India confirm the gender disparity in owning a smartphone. In urban slums, 90% of men had smartphones and 59% were internet users whereas only 22% of women had smartphones and 18% were internet users(52). One study conducted in Aligarh, India found that none of the female respondents in their study knew about app stores or how to download applications(53). Also, cultural practices can create additional challenges for female users. In many cultures, women are expected to not speak loudly and to speak rarely in public settings, so using voice-enabled applications in public spaces would be difficult(54). All of these cultural barriers and privacy concerns can slow the penetration of voice technology in LMICs.

Opportunities to Accelerate Voice

Increasing the use and impact of voice in development requires investments in technology, business, and policy, which all coalesce to enable and incentivize private enterprises to enter the voice market and consumers to buy or use voice services. Based on information gathered from entrepreneurs, subject matter experts, and desk research, our team learned of several opportunities for investment that could accelerate the development of voice recognition technologies in Africa and India that benefit semi-literate and illiterate populations. These opportunities address the technology, business, and policy aspects of voice recognition development.



Partner to Build Speech Corpora - Develop diverse speech databases through private sector partnership with companies who have operations in communities where low-literate customers live. The Foundation can replicate the relationship between Navana Tech and Ujjivan Small Finance Bank by collaborating with businesses to gather voice data in local languages during a business's regular interaction with their customers in rural or semi-rural locations. The Foundation can incentive businesses to participate in the partnership through grants.



Develop Culturally Relevant Icons/Images/UX - Communicate with semi-literate population in India and Africa through icons. Icons are a relatively less niche channel that can be used to bridge the literacy barrier in addition to voice recognition technology. An expert in user interface for inexperienced technology users referenced icons as an effective way to interact with lower literate mobile phone users. Icons are interpreted based on shared cultural experiences and do not have the added complication of dialects. Therefore, it is likely that they will reach a broader range of semi-literate people in the same country who may speak different languages. The foundation can also use private sector partnerships to build a database of relevant icons/images that can be used in mobile apps to drive financial inclusion.



Build Better APIs: From an interview with Sid Garg at Teller, one opportunity for investment is helping to create more application programming interfaces (API) that can be used to integrate software with major African telecom carriers. The current API barrier is threefold. First, it takes 9-12 months for Teller to integrate with a telco(15). Second, only a few aggregators (e.g. Africa's Talking) build APIs in Africa and "they charge a lot of money"(15). Third, these aggregators do not focus on the user base (i.e. low income) that Teller is focused on.



Policy/Business

Increase Entry-Level Smartphone Install Base: Since the target populations of voice technologies for development are the low-literate and illiterate poor, these populations are less able to afford smartphones. About 689 million people live below the \$1.90/day global poverty line. This makes even one of the cheapest full smartphones available, the \$29 Mobicel Astro, cost over 50% of their monthly income(58). Reducing smartphone prices will improve affordability, increase the install base, and increase the incentives for software developers. This, according to the CEO at Teller, will help incentivize developers and telcos to incorporate higher cost functionalities, such as voice, into applications.

Potential Solution: In a 2017 report, GSMA describes ‘asset financing’ as a way for low-income individuals to access smartphones through loans from carriers, banks, and finance groups(59). A number of organizations (e.g. Vodafone) have done pilot financing programs.



Policy

Reduce Data Rates: WhatsApp is so successful partly because it costs as low as 0.002 cents to send a message. Based on an interview with Teller, voice message however can be as high as 30-40 cents per minute. Since telcos in some larger countries only earn \$10-15 per user per year, voice data rates are significant for both telcos and developers. The Alliance for Affordable Internet (A4AI) has a few suggestions to reduce these rates. One supply-side intervention is better regional agreements to share infrastructure across operators and “establish agreements on interconnection and outbound rates”(57). At the governmental level, policies around universal access and spectrum management encourage affordable broadband(57).



Policy

Create Partnerships with Governments: Governments wanting to provide certain services to their citizens can benefit from voice technologies. For example, the Pakistani government sends pension checks to citizens living in remote areas, but a certain amount of fraud occurs because the government cannot verify if the right person is alive and receiving the money(57). Know your customer (KYC) is central to banking and officials want a voice biometric system to identify users based on their voice. Though opportunities such as this might start with voice authentication, they could lead to a broader use of voice by government.

Outlook on the Future of Voice in Africa and India

OVERALL ASSESSMENT

The START team is confident about the potential of voice technology to impact development and increase financial inclusion for semi-literate and illiterate populations. We are also cautiously optimistic about the growth of voice technology to reach these populations. However, the digital access and use divide is a significant barrier for these populations. For example, in developing countries, women are 14% less likely to own a mobile phone than men, which translates into 200 million fewer women owning phones⁽⁵⁵⁾ ⁽⁵⁶⁾. Women are 50-60% less likely to own a mobile phone in India and 10-40% less likely in Sub-Saharan Africa⁽⁵⁵⁾. Still, voice technology has the capacity to reach millions of users in developing nations that may lack the literacy competency required to participate in the current digital landscape. The reduced ability to read and write languages has alienated many rural users. For example, in India although rural users (277 million) outnumber urban users (227 million), the rural population is less likely to fully benefit from internet access due to lower rates of literacy⁽⁵⁷⁾. Voice technology has the potential to change that.

Furthermore, the use of voice technology can increase access to financial services among the poor, especially those living in areas where there is little incentive for financial institutions to open a branch. Financial institutions can leverage voice technology to reach previously unbanked customers and drive the adoption of healthy financial practices, as demonstrated by Ujjivan Small Finance Bank described in the Navana Tech case study.

The endeavors taken by many research institutions, universities, and private organizations to expand the data corpora available as well as the growth of open-source data are creating infrastructure required to train the NLP models. Although some private organizations, such as Apple, maintain proprietary voice data, open-source data initiatives described earlier are promising and will lead to faster technological developments⁽⁵⁸⁾. Moreover, certain trends, such as the increasing attention paid by African and Indian governments to expand internet connectivity and the reduction in the average cost of smartphones and of data-plans, all suggest that the infrastructure required for voice technology is heading in the right direction.

Nevertheless, the START team is apprehensive about the timeline of the adoption of voice-based technology. Our assessment is that voice technology is currently working for mature economies that have stronger economies, have higher rates of digital literacy, and are at later stages of technology development. So, it would require large investments and significant time to translate the use of voice technology to other regions and populations that use multiple languages, have differences in accents and tonality, and have reduced digital skills and lower levels of literacy.

POTENTIAL IMPACT OF VOICE RECOGNITION IN DEVELOPMENT

The top 20 languages spoken in the world cover over 6.1 billion people(59). However, **Table 2** shows that eight of those 20 languages, which cover 1.2 billion people, are not available on Google’s or Amazon’s speech technology(60) (61). Also, six of these eight languages are mainly used in three countries: India, Bangladesh, and Pakistan. Although literacy rates for specific language populations do not exist, the illiterate population for these eight languages can be estimated by averaging the adult literacy rates in the three above countries and applying that to the 1.2 billion people that speak these eight languages. **Table 2** shows that an estimated 357 million illiterate people speak these eight languages that are not even on the list of speech capabilities for Google or Amazon. This is also a highly conservative estimate because it does not count semi-literate individuals.

Table 2 – Top 8 Languages Not Covered by Google or Amazon

Language	Pop. (m)	Google DialogFlow ^b	Alexa Skills ^c
Standard Arabic	274	N	N
Bengali	265	N	N
Urdu	171	N	N
Swahili	99	N	N
Marathi	95	N	N
Telegu	93	N	N
Tamil	84	N	N
Western Punjabi	83	N	N
Total^a	1164		
Total Illiterate in 8 Languages	357		
Avg. Literacy Rate Estimation for 8 Languages ^d			
India Adult Literacy Rate	74.3%		
Pakistan Adult Literacy Rate	59.0%		
Bangladesh Adult Literacy Rate	74.7%		
Average Literacy Rate	69.3%		

^aEthnologue Top 200 Languages (2018) ^bGoogle DialogFlow Language Reference (2020) ^cAlexa Skills Kit (2020) ^dWorld Bank (2020)

Conclusion

The overall objectives of this work were to (1) understand how existing companies in Africa and India are leveraging NLP-based voice technologies to reach the semi-literate and illiterate poor with new services, especially digital financial services, and (2) describe opportunities to accelerate the use of voice in development. From the case studies and interviews with experts, barriers to using voice recognition to drive development are substantial, but not insurmountable. Strategic initiatives and partnerships have enabled organizations to overcome these barriers and effectively use voice to reach semi-literate and illiterate users. The South African government's initiative to fund the collection of 50 hours of speech data for each of its 11 languages and Navana Tech's partnership with Ujjivan Bank to use the bank's field workers to gather voice data are just two examples of successful strategic initiatives.

Furthermore, voice technology is not the only tool that can be used to expand financial services and drive development for semi-literate and illiterate populations in Africa and India. The future of voice might lie in combining it with culturally appropriate icons and images, pre-recorded audio and video, and intentional user design for these populations. With this suite of tools embedded in mobile apps, the divide in digital use and quality of use will shrink.

Limitations

Our findings are subject to some limitations which may affect our current assessment and recommendations for the voice recognition technology landscape in Africa and India.

- 1. Limited access to data:** Given that voice recognition technology is a niche field in Africa and India, and the target regions are LMICs, access to credible and relevant data about this technology was limited. Furthermore, several subject matter experts mentioned a dearth of publicly available information on voice recognition technology used in the target countries, especially about information on low-resource languages in these regions.
- 2. Number of case studies and subject matter experts:** The number of case studies reviewed, and subject matter experts interviewed were restricted due to time constraints. Therefore, our findings may not be sufficient to provide a balanced and holistic view of the current landscape for voice recognition technology use in India and Africa.

These limitations can be overcome in future studies through a larger sample size of use cases, additional interviews with subject matter experts, and a longer research period, all of which will support a more holistic investigation into the use of voice recognition technology in Africa and India.

Appendix

Appendix A

Organization/Name & Title	Type	Key Insights
<p>University of Washington Computer Science</p> <p>Sudheesh Singanamall</p> <p>PhD student</p>	<p>Research</p>	<ol style="list-style-type: none"> 1. Allens School focuses on NLP for justice and racial bias; GPT3 deep learning; open machine learning and not really on speech recognition for low-resourced languages 2. Primary challenges in voice tech space are speech corpus and local accents/variations of languages 3. Harvesting and storing voice data to train/refine models can be tricky because consent is required, which can be difficult to get from semi-literate people
<p>University of Washington Computer Science</p> <p>Samia Ibtasam</p> <p>PhD student</p>	<p>Research</p>	<ol style="list-style-type: none"> 1. One way to collect data is through free IVR systems deployed in low-resourced settings; this is the approach taken by Agha Ali Raza in Pakistan who now may have enough data to train an NLP model. This requires human moderators and translators though to label the data 2. Voice recognition for proof of identity and proof of life is interesting to government of Punjab to verify if pensioners are alive and prevent fraud 3. Privacy is a significant barrier and concern for users, especially in public spaces; speech differentiation from background noise is challenging in these environments; women having financial agency is still significant barrier even with women who have phones 4. Using voice to guide through simple prompts is more useful than through complex use cases
<p>Microsoft Research India</p> <p>Indrani Medhi Thies</p> <p>Principal Researcher</p>	<p>Corporate Research</p>	<ol style="list-style-type: none"> 1. Primary smartphone uses in low-income communities are for communication, WhatsApp, entertainment and social media. People show extra caution for financial and health apps 2. Pure voice is less feasible than hybrid model complementing voice with images, icons, videos, and pre-recorded audio 3. Does not know of any case examples of companies using speech recognition for financial services

		<p>4. Semi-literate users are more comfortable with short, simple voice input phrases (e.g. video for tomato sewing)</p> <p>5. Other ways to reach semi-literate users, such as <u>Eko's business model</u> in India.</p> <p>6. Low-literacy is correlated with lower cognitive ability, which means these users struggle more with digital apps and frameworks (e.g. how mobile money works)</p>
<p>University of Ibadan Seyl Osunade Associate Professor</p>	<p>Research</p>	<p>1. Yoruba dictionary is still in research and development stage; highly difficult to digitize tonation and place accent marks correctly</p> <p>2. Purpose of the dictionary is a starting point for translation and to detect plagiarism</p> <p>3. Obafemi Awolowo University in Nigeria has a speech corpus for Yoruba, but they are in the middle of transcribing that to text. Has not seen any other work using NLP speech recognition in low-income communities.</p>
<p>University of Washington Computer Science, Information and Communication for Development Kurtis Heimerl Assistant Professor</p>	<p>Research</p>	<p>1. Big tech companies (e.g. Google) see value in the long tail of languages and are working on solutions (voice search) for people coming online for first time. Startups tend not to focus on new online users who might be semi-literate or low-income</p> <p>2. Story based engagement in apps is more successful for semi-literates than straightforward task flows</p> <p>3. Spoken dialogue systems on IVR use voice as input instead of DTMF but results on user engagement are mixed and these systems are very limited</p> <p>4. Does not know of other groups using NLP models in speech recognition for developing areas</p>
<p>Teller Sid Garg Founder</p>	<p>Financial Services</p>	<p>1. Profitable company now because clients are large telecom providers willing to pay to reach customers through messaging channels</p> <p>2. Some grant work with UW Capital Development Fund has been voice-based. Teller built a system to allow for DTMF and voice input for workers to access information about when to expect payment, how much, etc. Voice system could synthesize text in Excel to speech output</p> <p>3. User studies he has done in low-income countries show that users won't say they need voice. Voice is pushed more by UN and other programs.</p>

		<p>5. One barrier is that voice data is 10x more expensive than SMS</p> <p>6. Barriers that need to be lowered are: install base of smartphones needs to increase; easier integration with African telcos because APIs are not widespread enough (takes 9-12 months for Teller to build voice platform for a telco).</p>
<p>Navana Tech Raoul Nanavati Co-founder</p>	<p>Enterprise Software</p>	<p>1. Voice data gathered by Navana’s customers (e.g. Ujjivan Bank) for free; Navana provides discounted rates in the future to these customers for its SDK</p> <p>2. Voice NLP model must be highly accurate otherwise will lose customers because new technology handling money is scary to semi-literate users</p> <p>3. Combines illustrative iconography, pre-recorded audio help in the app, and voice bot for more conversation-based help</p> <p>4. Not VC funded so is not pressured to focus on higher income groups</p> <p>5. Data corpus with open datasets is a huge limitation. So, they always have a deep partnership with a firm on the ground to gain data and insights to build the app.</p>
<p>Aja.la Studios Abake Adenle Founder</p>	<p>Enterprise Software</p>	<p>1. Biggest barrier Adenle sees is figuring out the specific use cases for various sectors and creating sustainable revenue-generating models for speech</p> <p>2. Received grant funding for 5 speech enabled languages so far and has 7 more languages in the pipeline. Yoruba word error rate is higher than what is reported in literature for other languages, according to Adenle.</p> <p>3. Primary use case identified so far is customer service automation in financial sector through voice and is developing that for a bank now.</p> <p>4. Specialization of Aja.la is developing speech models with limited data. Currently has 2,000 hours of data across 11 languages</p>

References

1. The Mobile Economy Sub-Saharan Africa 2019 [Internet]. GSMA Sub Saharan Africa. [cited 2020 Dec 15]. Available from: <https://www.gsma.com/subsaharanafrica/resources/the-mobile-economy-sub-saharan-africa-2019>
2. Indian to have 820 million smartphone users by 2022 - The Economic Times [Internet]. [cited 2020 Dec 15]. Available from: <https://economictimes.indiatimes.com/industry/telecom/telecom-news/indian-to-have-820-million-smartphone-users-by-2022/articleshow/76876369.cms?from=mdr>
3. Literacy - Our World in Data [Internet]. [cited 2020 Dec 15]. Available from: <https://ourworldindata.org/literacy>
4. Latest Consumer Insights Survey: Mapping Mobile Internet Use [Internet]. GSMA Intelligence. 2020 [cited 2020 Dec 15]. Available from: <https://www.gsmainelligence.com/product-news/latest-consumer-insights-survey-mapping-mobile-internet-use/>
5. FactsFigures2019.pdf [Internet]. [cited 2020 Dec 15]. Available from: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf>
6. Wills A. Conversational interfaces and the long tail of languages in developing countries. :28.
7. Ghosh S. India sets target of a billion digital transactions per day [Internet]. mint. 2020 [cited 2020 Dec 15]. Available from: <https://www.livemint.com/news/india/india-sets-target-of-a-billion-digital-transactions-per-day-11591963111515.html>
8. [Tech30] With Navana Tech, these brothers are building a mobile SDK to make the Internet Bharat-friendly [Internet]. [cited 2020 Dec 15]. Available from: <https://yourstory.com/2019/11/tech30-navana-tech-billion-users-internet-literacy-cornell>
9. Nanavati, R. Voice Recognition in Navana Tech. [Personal Interview, 19 Nov] Online; 2020 (unpublished).
10. MGI-Digital-India-Report-April-2019.pdf [Internet]. [cited 2020 Dec 15]. Available from: <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20India%20Technology%20to%20transform%20a%20connected%20nation/MGI-Digital-India-Report-April-2019.pdf>
11. Literacy rate, adult total (% of people ages 15 and above) | Data [Internet]. [cited 2020 Dec 15]. Available from: <https://data.worldbank.org/indicator/SE.ADT.LITR.ZS>
12. Garg, S. Voice Recognition in Teller. [Personal Interview, 17 Nov] Online; 2020 (unpublished).
13. Adenle, A. Voice Recognition in Aja.la Studios. [Personal Interview, 11 Dec]; Online (unpublished).
14. Barnard E, Davel MH, van Heerden C, de Wet F, Badenhorst J. The NCHLT Speech Corpus of the South African languages. St Petersburg. 2014;7.
15. These are SA's most spoken languages in 2019 – The Citizen [Internet]. [cited 2020 Dec 15]. Available from: <https://citizen.co.za/news/south-africa/general/2139086/these-are-sas-most-spoken-languages-in-2019/>

16. • Smartphone users in South Africa 2014-2023 | Statista [Internet]. [cited 2020 Dec 15]. Available from: <https://www.statista.com/statistics/488376/forecast-of-smartphone-users-in-south-africa/>
17. Tjønneland EN. Crisis at South Africa's universities – what are the implications for future cooperation with Norway? CMI Brief [Internet]. 2017 [cited 2020 Dec 15];3. Available from: <https://www.cmi.no/publications/6180-crisis-at-south-africas-universities-what-are-the>
18. How big are South Africa's schools? [Internet]. [cited 2020 Dec 15]. Available from: <https://passmark.org.za/howbig/>
19. Niyo Bharat aims at 50 lakh customer base in two years [Internet]. [cited 2020 Dec 15]. Available from: <https://www.outlookindia.com/newscroll/niyo-bharat-aims-at-50-lakh-customer-base-in-two-years/1752898>
20. Over 120 million blue collar employees go unpaid as operational restrictions impact businesses | Business Insider India [Internet]. [cited 2020 Dec 15]. Available from: <https://www.businessinsider.in/careers/news/over-120-million-blue-collar-employees-go-unpaid-as-operational-restrictions-impact-businesses/articleshow/75421501.cms>
21. September 2019 9th. Indian neobank Yelo raises seed round [Internet]. FinTech Futures. 2019 [cited 2020 Dec 15]. Available from: <https://www.fintechfutures.com/2019/09/indian-neobank-yelo-raises-seed-round/>
22. India - households by annual income brackets 2010-2025 [Internet]. Statista. [cited 2020 Dec 15]. Available from: <https://www.statista.com/statistics/482584/india-households-by-annual-income/>
23. येLo (Yelo Bank): Using AWS to Provide Financial Solutions for the 'Next Half Billion' [Internet]. Amazon Web Services. 2020 [cited 2020 Dec 15]. Available from: <https://aws.amazon.com/blogs/startups/yelo-bank-using-aws-to-provide-financial-solutions-for-the-next-half-billion/>
24. An Overview of Our Businesses - Kotak Mahindra Bank [Internet]. [cited 2020 Dec 15]. Available from: <https://www.kotak.com/en/about-us/our-business.html>
25. Kotak_Mahindra_Bank_Limited_FY20.pdf [Internet]. [cited 2020 Dec 15]. Available from: https://www.kotak.com/content/dam/Kotak/investor-relation/Financial-Result/Annual-Reports/FY-2020/kotak-mahindra-bank/Kotak_Mahindra_Bank_Limited_FY20.pdf
26. www.ETCIO.com. Kotak Mahindra Bank brings first AI powered bilingual voicebot Keya - ET CIO [Internet]. ETCIO.com. [cited 2020 Dec 15]. Available from: <https://cio.economictimes.indiatimes.com/news/business-analytics/kotak-mahindra-bank-brings-first-ai-powered-bilingual-voicebot-keya/63582501>
27. Bank KM. Kotak Mahindra Bank Launches Keya 2.0 Indian Banking's First Voicebot gets Upgraded with AI-powered Conversational Banking And Keya is Now Available as a Chatbot too. :2.
28. Population, total - India | Data [Internet]. [cited 2020 Dec 15]. Available from: <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=IN>
29. Overview [Internet]. World Bank. [cited 2020 Dec 15]. Available from: <https://www.worldbank.org/en/country/india/overview>

30. Hindi mother tongue of 44% in India, Bangla second most spoken | India News - Times of India [Internet]. [cited 2020 Dec 15]. Available from: <https://timesofindia.indiatimes.com/india/hindi-mother-tongue-of-44-in-india-bangla-second-most-spoken/articleshow/64755458.cms>
31. In India, who speaks in English, and where? [Internet]. [cited 2020 Dec 15]. Available from: <https://www.livemint.com/news/india/in-india-who-speaks-in-english-and-where-1557814101428.html>
32. HDFC 2019 Annual Report [Internet]. [cited 2020 Dec 15]. Available from: <https://www.hdfcbank.com/content/api/contentstream-id/723fb80a-2dde-42a3-9793-7ae1be57c87f/1bcf4f2c-17cc-4759-9081-dcc0f5beeb60>
33. www.ETCIO.com. HDFC Bank's EVA becomes India's largest banking Chatbot - ET CIO [Internet]. ETCIO.com. [cited 2020 Dec 15]. Available from: <http://cio.economictimes.indiatimes.com/news/strategy-and-management/hdfc-banks-eva-becomes-indias-largest-banking-chatbot/60421681>
34. www.ETCIO.com. HDFC Bank launches IRA 2.0, the advance version of its interactive humanoid - ET CIO [Internet]. ETCIO.com. [cited 2020 Dec 15]. Available from: <https://cio.economictimes.indiatimes.com/news/enterprise-services-and-applications/hdfc-bank-launches-ira-2-0-the-advance-version-of-its-interactive-humanoid/63936738>
35. www.ETCIO.com. HDFC Bank launches interactive humanoid IRA - ET CIO [Internet]. ETCIO.com. [cited 2020 Dec 15]. Available from: <http://cio.economictimes.indiatimes.com/news/corporate-news/hdfc-bank-launches-interactive-humanoid-ira/56825673>
36. • Average monthly GB per smartphone in US 2010-2017 | Statista [Internet]. [cited 2020 Dec 15]. Available from: <https://www.statista.com/statistics/801229/mobile-data-usage-per-smartphone-in-the-us/>
37. Cellular data average price in US 2018-2023 [Internet]. Statista. [cited 2020 Dec 15]. Available from: <https://www.statista.com/statistics/994913/average-cellular-data-price-per-gigabyte-in-the-us/>
38. Chang E, Zhou J, Di S, Huang C, Lee K-F. LARGE VOCABULARY MANDARIN SPEECH RECOGNITION WITH DIFFERENT APPROACHES IN MODELING TONES. :4.
39. cycles T text provides general information S assumes no liability for the information given being complete or correct D to varying update, Text SCDM up-to-DDTR in the. Topic: US smartphone market [Internet]. Statista. [cited 2020 Dec 15]. Available from: <https://www.statista.com/topics/2711/us-smartphone-market/>
40. Census of India 2011. [Internet]. [cited 2020 Dec 15]. Available from https://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf.
41. Resources for languages in United States [Internet]. [cited 2020 Dec 15]. Available from: <http://www.language-archives.org/country/US>
42. Jha GN. The TDIL program and the Indian Language Corpora Initiative (ILCI). :4.
43. Microsoft releases largest publicly available speech data for three Indian languages to aid researchers [Internet]. OnMSFT.com. 2018 [cited 2020 Dec 15]. Available from:

<https://www.onmsft.com/news/microsoft-releases-largest-publicly-available-speech-data-for-three-indian-languages-to-aid-researchers>

44. Thies, I M. Voice Recognition in India. [Personal Interview, 6 Nov] Online; 2020 (unpublished).
45. Singanamall, S. Voice Recognition. [Personal Interview, 27 Oct] Online; 2020 (unpublished).
46. Project Mélange [Internet]. Microsoft Research. [cited 2020 Dec 15]. Available from: <https://www.microsoft.com/en-us/research/project/melange/>
47. About Us – APIs for developers in Africa | Africa’s Talking [Internet]. [cited 2020 Dec 15]. Available from: <https://africastalking.com/about>
48. A platform for Africa’s mobile innovators [Internet]. MIT News | Massachusetts Institute of Technology. [cited 2020 Dec 15]. Available from: <https://news.mit.edu/2019/africas-talking-mobile-0611>
49. Africa’s Talking – Pricing [Internet]. [cited 2020 Dec 15]. Available from: <https://africastalking.com/pricing>
50. Digital Literacy in India: Structural Constraints and the NEP 2020 [Internet]. [cited 2020 Dec 15]. Available from: <https://www.sprf.in/post/digital-literacy-in-india-structural-constraints-and-the-nep-2020>
51. Medhi Thies I. User Interface Design for Low-literate and Novice Users: Past, Present and Future. *FNT in Human–Computer Interaction*. 2015;8(1):1–72.
52. Final Report Digital literacy, Institute of Home Economics.pdf [Internet]. [cited 2020 Dec 16]. Available from: <https://wcd.nic.in/sites/default/files/Final%20Report%20Digital%20literacy%2C%20Institute%20of%20Home%20Economics.pdf>
53. Digital Literacy among Rural Youths at Jarauthi Village, Aligarh: Status and Intervention [Internet]. SKH. [cited 2020 Dec 16]. Available from: <http://www.socialworkfootprints.org/6/post/2018/12/digital-literacy-among-rural-youths-at-jarauthi-village-aligarh-status-and-intervention.html>
54. Ibtasam, S. Voice Recognition in Africa and India. [Personal Interview, 6 Nov] Online; 2020 (unpublished).
55. GSMA-The-Mobile-Gender-Gap-Report-2019.pdf [Internet]. [cited 2020 Dec 16]. Available from: <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf>
56. The State of Mobile Internet Connectivity Report 2019 [Internet]. Mobile for Development. [cited 2020 Dec 15]. Available from: <https://www.gsma.com/mobilefordevelopment/resources/the-state-of-mobile-internet-connectivity-report-2019/>
57. Digital in India [Internet]. [cited 2020 Dec 16]. Available from: <https://cms.iamai.in/Content/ResearchPapers/2286f4d7-424f-4bde-be88-6415fe5021d5.pdf>
58. The Future of American Industry Depends on Open Source Tech | WIRED [Internet]. [cited 2020 Dec 16]. Available from: <https://www.wired.com/story/opinion-the-future-of-american-industry-depends-on-open-source-tech/>

59. What are the top 200 most spoken languages? [Internet]. Ethnologue. 2018 [cited 2020 Dec 15]. Available from: <https://www.ethnologue.com/guides/ethnologue200>
60. Language reference | Dialogflow ES [Internet]. Google Cloud. [cited 2020 Dec 15]. Available from: <https://cloud.google.com/dialogflow/es/docs/reference/language>
61. Develop Skills in Multiple Languages | Alexa Skills Kit [Internet]. [cited 2020 Dec 15]. Available from: [/alexa/techdoc-template](#)