Contents

List of Figures

Library Management Using Voice Assistant

1 Introduction

A student's life revolves around the library. It is also beneficial to academic researchers. Libraries may now be found in almost any location. It's a huge endeavor to locate a book at the library. While looking for the needed book, some people lose interest. It takes a long time to look for books on the computer.

With open source software products like Moodle for Virtual Learning gaining traction in related fields, many librarians are looking for OSS alternatives to their present Library Management Systems.

An Intelligent Personal Assistant (IPA) is a computer application that uses Artificial Intelligence (AI) to help people do tasks. The IPA maintains a continuous dialogue with its users while responding to their questions or carrying out measures to fulfill their demands. Modern IPAs are capable of a broad range of functions, from simple ones like opening an app or setting an alarm to more complicated ones like taking notes or making phone calls. Google Assistant from Google, Siri from Apple, and Alexa from Amazon are all instances of IPAs.

Although IPAs are not required in terms of communicating exclusively through voice, many current IPAs are pursuing Voice User Interface, which involves engaging with users only through voice, without the necessity of displays or physical interaction. This necessitates the IPA's ability to (A) listen to human speech, (B) comprehend what is being indicated, and (C) conduct an action or respond with their own synthesized voicection.

Natural Language Processing (NLP) is mostly about instructing machines to understand human languages and derive meaning from texts at their most basic level. Text mining, text categorization, text and emotion monitoring, and speech production and identification are just a few of its many applications. This is also why Natural language processing We constructed a working voice assistant for the purpose of this article, which can do simple tasks such as locating a book in the library. It can interpret "audio instructions" and obtain information from a database.

2 Proposed work

We employ speech recognition to discover books in our project, which makes this mammoth process a lot simpler. Even those with no prior understanding of the system can use it to discover books more quickly. It cuts down on time spent looking for books and makes the library more user-friendly. The system will connect with its users in real time as it responds to their questions. Inquiries include book locations based on a unique identifier or the title of the book.

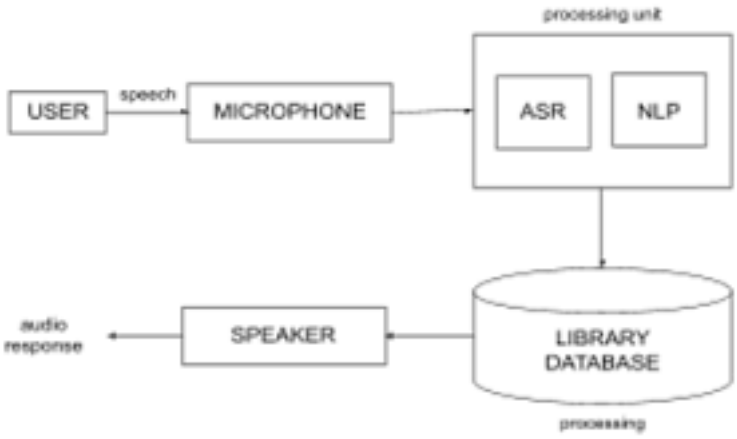

Figure 1: Architecture Diagram

The steps taken will be as follows :
Step 1 : User Speaks User speaks into the microphone and then these signals are then sent to the processor.
Step 2 : Audio Processing (ASR and NLP) Here, Automatic Speech Recognition (ASR) is used which takes an audio file and converts it into a sequence of words. The NLP module picks up keywords and identifies the book and its location in the database.
Step 3 : Relevant information gathered from the database Information is retrieved from the database (book and its location).
Step 4 : Response/Error is received The response is received through the speaker.

3 Modules

3.1 Automatic Speech Recognition (ASR) module

Speech recognition is defined as the automatic recognition of human speech and is recognized as one of the most important tasks when it comes to making applications based on Voice User Interfacing (VUI).

Python comes with several libraries which support speech recognition. We will be using the speech recognition library because it is the simplest and easiest to learn.

This module involves 2 steps -

(a) Loading Audio
To load audio the Audio File function is used. The function opens the file, reads its contents and stores all the information in an Audio File instance called source.

We will traverse through the source and do the following things:
- Every audio has some noise involved which can be removed using the adjust for ambient noise function.
- Making use of the record method which reads the audio file and stores certain information into a variable to be read later on.

(b) Reading data from audio
We can now invoke recognize_google() method and recognize any speech in the audio. After processing the method returns the best possible speech that the program was able to recognize from the first 100 seconds. The output comes out to be a bunch of sentences from the audio which turn out to be pretty good. The accuracy can be increased by the use of more functions but for now it does the basic functionalities.

3.2 Natural Language Processing (NLP) module

SpaCy is a Natural Language Processing package for Python that is open-source. It's primarily intended for use in the workplace, where it may create real-world initiatives and manage massive amounts of text data. Moreover, sixty languages are supported by spaCy, which has prepared pipelines for various languages and activities. Because this toolset is designed in Python and Cython, it is quicker and more accurate when dealing with massive amounts of text data.

3.3 Subprocess

Utilized to obtain information on system subprocesses, which are then used in different commands like shutdown, sleep, and so on. Python includes this module by default.

3.4 WolframAlpha

Wolfram's algorithms, knowledgebase, and AI technologies are used to compute expert-level responses.

3.5 Pyttsx3

In a software, this is used to convert text to speech.

3.6 Date time

Used to know date and time. It comes built in with Python.

3.7 Speech Recognition

Used to make the assistant understand your voice

4 Benefits of Implementing Voice Technology

Implementing voice assistant technology in library operations has several benefits. Some of the main advantages are:

- Enhanced accessibility: Voice assistants can help people with disabilities access library services easily.
- Improved efficiency: With voice assistants handling routine tasks like reserving books or checking out items, librarians have more time to focus on other important activities.
- Personalized experience: Voice assistants can offer personalized experiences based on user preferences such as recommending relevant books or suggesting upcoming events.
- Cost-effective solution: Using voice assistant technology is an affordable alternative to hiring additional staff.

5 Future of Voice Technology in Libraries

The future of libraries looks promising with the integration of voice assistant technology. As artificial intelligence continues its rapid development, there is room for innovative solutions that can make the library experience even better than before. For instance, future advances might include AI with facial recognition capabilities that will allow patrons easy access to secure areas within a library without needing physical identification cards.

6 Challenges and Limitations of Voice Assistant Technology in Libraries

As with any new technological introduction there always come some challenges which would be faced along the way:
- Privacy concerns regarding user data collection by third-party vendors like Amazon/Google requires vigilance when deciding between competing products/services involving sensitive information that could be compromised if not properly secured otherwise.
- Security issues associated with remote devices accessing local networks may require additional measures taken by IT departments overseeing network security protocols already established internally within institutional infrastructures.

7 Voice Assistants

7.1 What are Voice Assistants?

Voice Assistants have become synonymous with Google Home and Amazon Echo. This perception cannot be more incorrect. The underlying technology powering both these smart speakers are the actual voice assistants, Google Assistant and Amazon Alexa, respectively.

Today voice assistants are not limited just to smart speakers but are also available in cars, household devices, smartphones, and several apps.

Voice Assistants are a subset of Virtual Assistants or Intelligent Personal Assistants. These virtual assistants can take inputs in many different ways:

1. Text - Such Intelligent virtual assistants called chat bots are text-based assistants.

2. Voice:- These types of assistants are called Voice Assistants.

3. Image:- These assistants take an image as an input, e.g. Google Lens, Bixby Vision

A Voice Assistant is a virtual assistant that uses speech recognition, natural language processing and speech synthesis to take actions to help its users.

7.2 Essential Conditions for being a Voice Assistant:

Every voice solution is not a voice assistant, but every voice assistant is a voice solution. To be called Voice Assistants, a voice solution needs to match these conditions:

1. Voice as input - Primary mode of input for a Voice Assistant should be through Voice.

2. Conversational - Voice Assistants should be able to have natural and contextual two-way communication with the user.

3. Confirmational:- Voice assistants should be able to confirm, clarify and answer the user with context.

8 History of Voice Assistants:

8.1 A short history of the voice revolution

Voice assistant integration has taken two different routes - one is on the smartphone, and the other is on smart speakers (and other connected devices). Apple, Google and Amazon have been focused on making their Voice Assistants ubiquitous through general-purpose assistants.

Apple has taken the route of getting Siri on all Apple devices. They even launched the Homepod, a smart speaker without much traction.

Google is uniquely positioned to catapult the adoption of Google Assistant by making it available almost mandatory across all android smartphones and finding success in the smart speaker market with the Google Home lineup.

Amazon had an early mover advantage in the world of smart speakers where they were almost two years ahead of Google and found success with Alexa in Echo devices.

The latter two have quickly moved to offer voice assistants across various third-party devices in different form factors to appeal to different user preferences and contexts.

We don’t want to bore you with all the history stuff. We could have written a bunch of paragraphs going into details of what happened when, but it’s best if we give you a quick graphical snapshot with this fantastic timeline created by our friends at voicebot.ai.

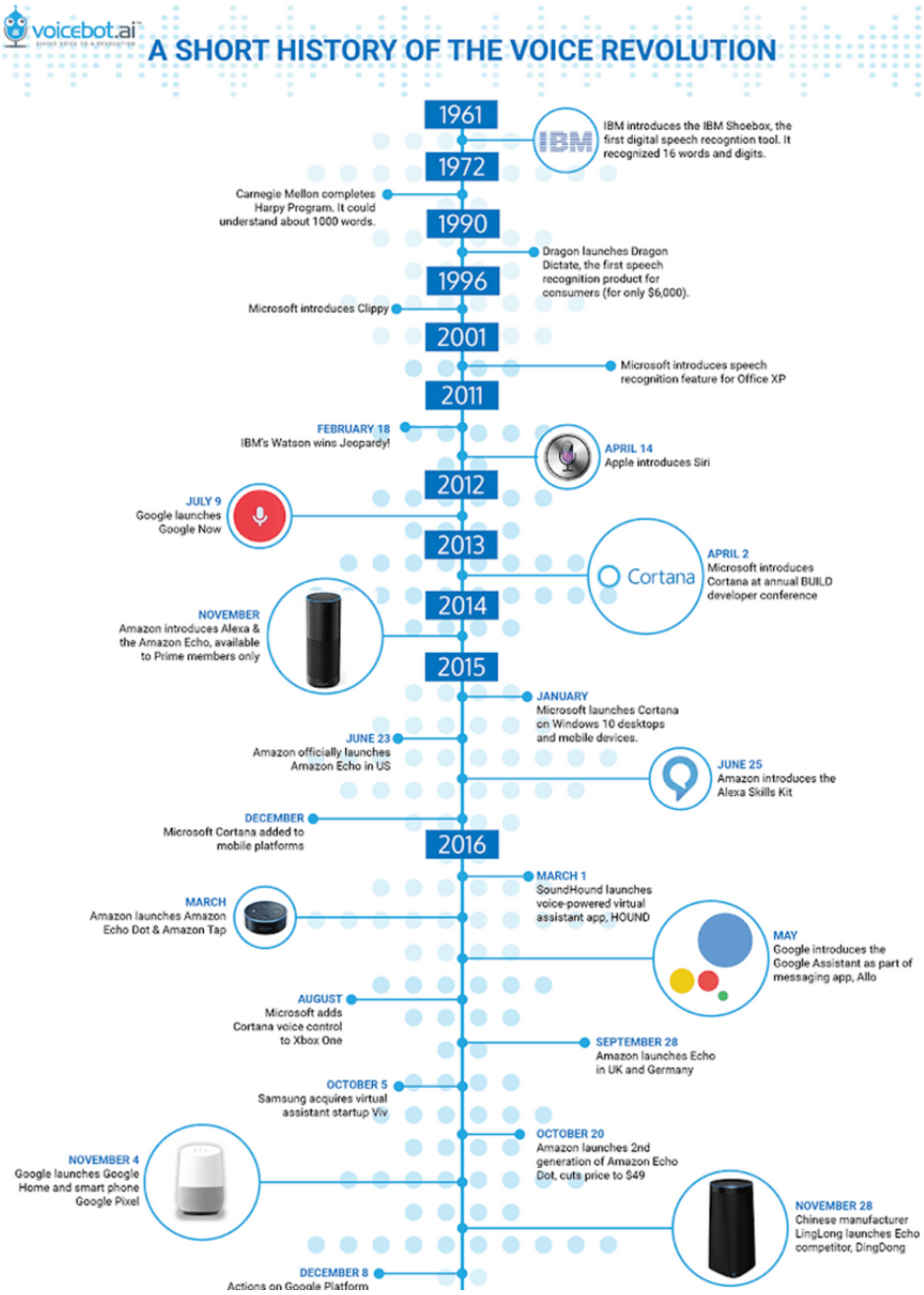

Figure 2: A short history of the voice revolution

8.2 Voice Assistant Timeline

- 1922 – First Voice activated consumer product hits store shelves as “Radio Rex”
- 1952 – Audrey, or the Automatic Digit Recognition Machine, is announced
- 1962 – IBM Shoebox is shown for the first time at the State Fair
- 1971 – Darpa funds five years of speech recognition research and development
- 1976 – Harpy is shown at Carnegie Mellon

- 1984 – IBM releases “Tangora” the first voice activated typewriter
- 1990 – Dragon Dictate is released
- 1994 – Simon by IBM is the first modern voice assistant released
- 2010 – Siri is released as an app on the iOS app store
- 2011 – IBM Watson is released
- 2012 – Google Now is released
- 2014 – Amazon Alexa and Echo are released
- 2015 – Microsoft Cortana is released
- 2017 – Alan is developed and released with the Alan Platform

8.3 Phases of voice assistant journey

Phase 1 was all about getting consumers introduced to the idea of using voice to perform tasks.

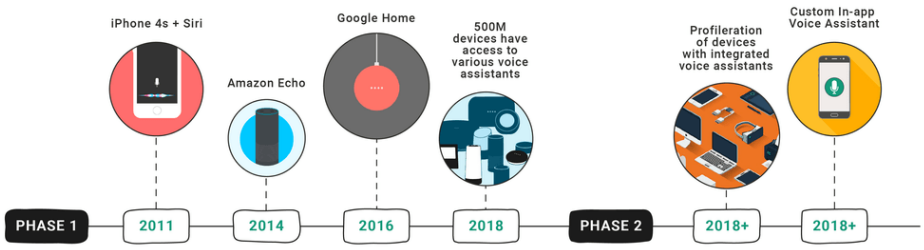

Figure 3: Phases of voice assistant journey

Phase 2 is about voice becoming a pervasive interaction mode with more capabilities which is used more frequently on more devices, in apps, and in different contexts.

9 Technology behind Voice Assistants

Voice assistants use Artificial Intelligence and Voice recognition to accurately and efficiently deliver the result that the user is looking for. While it may seem simple to ask a computer to set a timer, the technology behind it is fascinating.

9.1 Voice Recognition

Voice recognition works by taking an analog signal from a users voice and turning it into a digital signal. After doing this, the computer takes the digital signal and attempts to match it up to words and phrases to recognize the users intent. To do this, the computer requires a database of pre-existing words and syllables in a given language to be able to closely match the digital signal with. Checking the input signal with this database is known as pattern recognition, and is the primary force behind voice recognition.

9.2 Artificial Intelligence

Artificial intelligence is using machines to simulate and replicate human intelligence.

In 1950, Alan Turing (The namesake of our company) published his paper "Computing Machinery and Intelligence" that first asked the question, can machines think? Alan Turing then went on to develop the Turing Test, a method of evaluating a computer to test its capability of thinking like a human. There were four approaches later developed that defined AI, Thinking humanly/rationally, and acting humanly/rationally. While the first two deal with reasoning, the second two deal with actual behavior. Modern AI is typically seen as a computer system designed to accomplish tasks that typically require human interaction. These systems can improve upon themselves using a process known as machine learning.

9.3 Machine Learning

Machine learning refers to the subset of Artificial Intelligence where programs are created without the use of human coders manually creating the program. Instead of writing out the complete program on their own, programmers gives the AI "patterns" to recognize and learn from and then gives the AI large amounts of data to sift through and study. So instead of having specific rules to abide by, the AI searches for patterns within this data and uses it to improve its already existing functions. One way machine learning can be helpful for Voice AI, is by feeding the algorithm hours of speech from various accents and dialects.

While traditional programs requires an input and rules to develop an output, machine learning tools are given an input and an output and use that to create the program itself. There are two approaches to machine learning, supervised learning and unsupervised learning. In supervised learning, the model is given data that is already partly labeled, this means some of the data given will be already tagged with the correct answer. This helps guide the model into categorizing the rest of the data and developing a correct algorithm. In unsupervised learning, none of the data is labeled, so it is up to the model to find the pattern correctly. One of the reasons this is very useful is because it allows the model to find patterns that the creators might have never found on their own, but the data is much more unpredictable.

10 Different Voice Assistant approaches

Many conversational assistants today combine both a task-oriented and knowledge-oriented workflow to carry out almost any task that a user can throw at it. A task-oriented workflow might include filling out a form, while a knowledge-oriented workflow includes answering what the capital of a state might be or specifying the technical specifications of a product.

10.1 Task-oriented approach

A **task-oriented approach** is using goals to tasks to achieve what the user needs. This approach often integrates itself with other apps to help complete tasks. For example, if you were to ask a voice assistant to set an alarm for 3PM, it would understand this to be a task request and communicate with your default Clock application to open and set an alarm for 3PM. It would then communicate with the app to see if anything else was necessary, such as a name for the alarm, then it would communicate

this need back to you. This approach does not require an extensive online database, as it is mainly using the knowledge and already existing skills of other installed applications.

10.2 Knowledge-oriented approach

A **knowledge-oriented approach** is the use of analytical data to help users with their tasks. This approach focuses on using online databases and already recorded knowledge to help complete tasks. An example of this approach is anytime a user asks for an internet search, it will use the online databases available to return relevant results and recommend the highest search result. If someone is searching up a trivia question, this would use a knowledge-oriented approach as it is searching for data instead of working with other apps to complete tasks.

11 How do Voice Assistants work?

Figure 4: working of a voice assistant

Do you ever wonder how a single command like 'Alexa, how is the weather outside ?' is interpreted by your smart speaker. Don't worry. We will break it down for you to understand how this magic happens. We have abstracted out the nuances of this works and simplified it to help you understand:

- Automatic Speech Recognition (ASR)
- Natural Language Processing (NLP)
- Desired business logic via hooks
- Text To Speech (TTS)

12 Popular Voice Assistants

Voice Assistant adoption by platform, from Voicebot.ai

12.1 Siri

Siri is the most popular voice assistant today. Created in 2010 by SRI Inc, and purchased in 2011 by Apple, Siri has quickly become an integral part of the Apple ecosystem in bringing all the Apple devices and applications together to use in tandem with one another.

12.2 Alexa

Created by Amazon in 2014, Alexa was named due to its similarity to the Library of Alexandria. Alexa was originally inspired by the conversational voice system found on board the U.S.S. Enterprise in Star Trek. Alexa was released alongside The Amazon Echo, a smart speaker intended for consumers to dive into the world of home automation, uses the Alexa platform to allow users to interact with the Amazon ecosystem and allow for a plethora of smart devices to be connected.

12.3 Google Assistant

Originally unveiled in 2016, Google Assistant was the spiritual successor of Google Now, with the main improvement being the addition of two-way conversations. Where Google now would return answers in the form of a search results page on Google, Google Assistant gives answers in the form of natural sentences and returns recommendations in the form of Feature cards.

12.4 Cortana

Beginning in 2009, Cortana by Microsoft has had one of the longest visions of giving people access to voice assistants in their daily lives. Microsoft began shipping Cortana with all Windows 10 and Xbox devices, leading to a huge increase in the amount of registered Cortana users. In 2018 it was reported that Cortana had over 800 Million users.

12.5 Alan

In 2017 Alan set out to take voice assistants to the next level, by enabling voice AI for all applications. Using domain specific language models and contextual understanding, Alan is focused on creating a new generation of Enterprise Voice AI applications. By using the Alan Platform, developers are able to take control of voice, and create an effective workflow that best fits their users with the help of vocal commands.

13 Voice Assistants gaining popularity

Here is useful data point: According to Voicebot.ai, There are twice as many monthly active voice assistant users on smartphones as smart speakers, and voice usage in cars also exceeds the use of smart speakers.

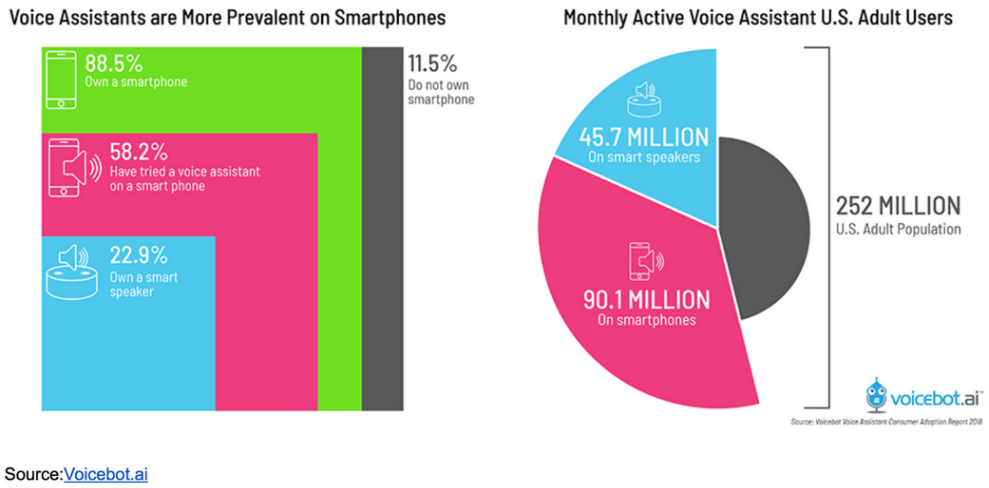

Figure 5: popularity

14 Usage of Voice in the Developing world

Smart speakers are yet to take off in the developing world, they might be all the rage in the West, but Voice Assistants on smartphones is still the king here. In India, masses cannot afford a stand-alone Echo dot or a Google Home, but they can use the same voice assistant through their existing smartphone.
According to the this estimate by eMarketer, a quarter of India's population will use smartphones. Here are a few points from the report published in 2018.

- There were 291.6 million smartphone users in India by the end of 2017.
- The number of smartphone users in India is estimated to hit 337 million by the end of 2018.
- The number of smartphone users in India would reach 490.9 million by 2022.

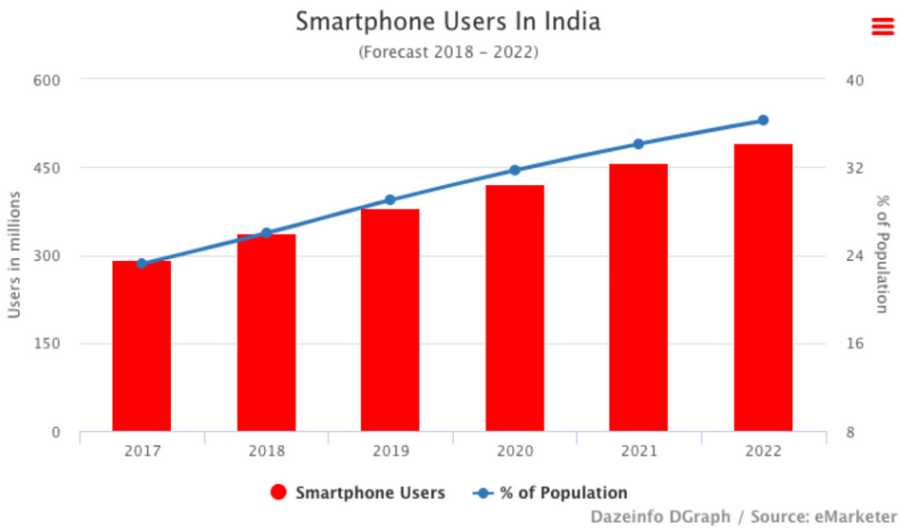

Figure 6: smart phone user in India

India has a vast potential to be the biggest market for voice search with an expected YoY growth of 270%, as revealed by Google.

According to MMA and Isobar report, globally over 500 million people use Google Assistant every month already, with Hindi second only to English as the most commonly used language.

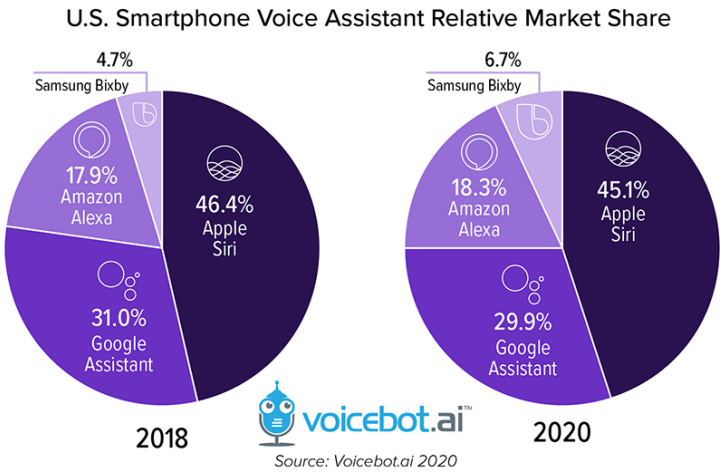

Figure 7: U.S. smart phone voice assistant relative market share

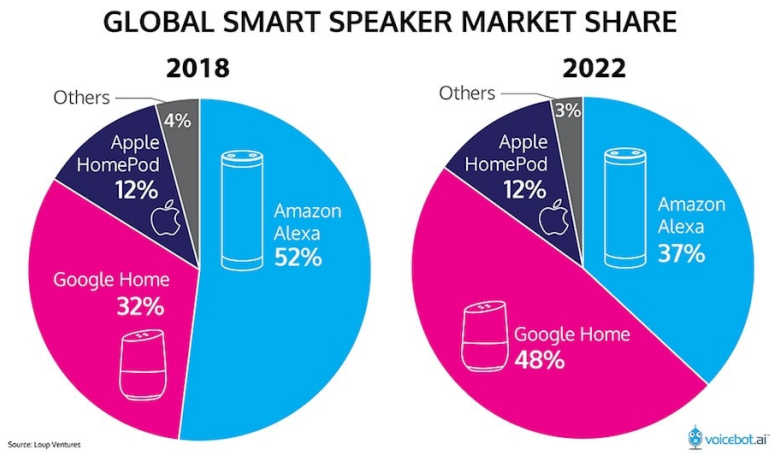

Figure 8: Global smart speaker market share

15 Types of Voice Assistants

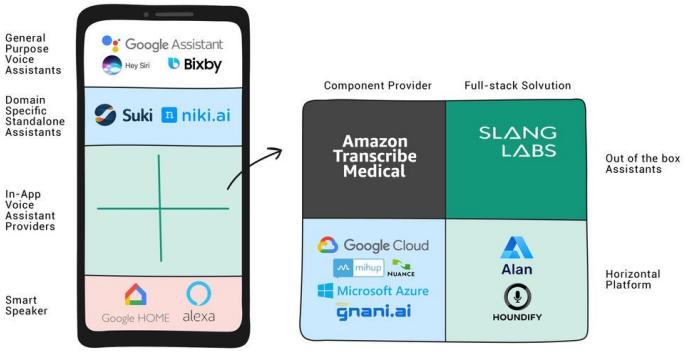

Figure 9: Types of Voice Assistants

If we try to understand the type of VA's available, there are broadly three categories into which we can divide them.

15.1 General Purpose Voice Assistants

Google Assistant, Siri, Bixby in Android, iPhone and Samsung devices, respectively, are great examples of General purpose Voice Assistant present on smartphones, smart speakers and other smart devices.

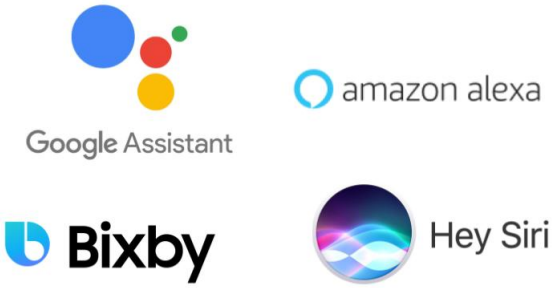

Figure 10: General Purpose Voice Assistants

All of these Voice assistants help you with general things like setting the alarm, scheduling events, making calls, launching apps, amongst others.

Google Assistant also has a feature called App Actions which can trigger actions inside specific apps using deep links. Recently, Alexa announced the same with Alexa on apps.

Unfortunately, Siri doesn't have any such feature currently and has poor performance compared to both Alexa and Google Assistant.
Increasingly, Google Assistant and Alexa are shifting strategy and focusing on doing tasks inside apps encroaching on the niche carved by in-app Voice Assistants.

15.2 In-app Voice Assistants

Witnessing the popularity of Voice Assistants, some brands have started adding Voice Assistant to the primary mode of communication - their apps and websites. Trainman, Flipkart,Amazon, Bank of America have added in-app voice assistants, and yuyiii.com has added a Voice Assistant to their website.

These voice assistants are present in an app to ease or elevate the customer experience.

With in-app voice assistants gaining popularity, we decided to do an in-depth analysis of some of them which are present in the most popular apps in India - Udaan, MyJio, Amazon, YouTube, Trainman, Gaana and JioMart.

15.3 Stand Alone Voice Assistants

These voice assistants don't sit inside an app, but they are the primary communication channel with the users in a stand-alone app. They are self-sufficient voice assistants usually built for limited use cases and specific domains.

Some examples of such assistants are Suki and Niki.ai. Suki is a stand-alone voice assistant in an app built explicitly for doctors. Suki used component providers like Amazon Transcribe Medical to create this Voice Assistant.

Figure 11: Stand Alone Voice Assistants

15.4 Smart Speakers and Smart Devices

Figure 12: Smart Devices

These devices are sold by the companies which make these voice assistants in most cases - Amazon and Google. They have Google Assistant and Alexa built-in.

15.5 Domain Specific Voice Assistant

As the name suggests, Domain-specific voice assistants are specific to a particular domain, e.g., eCommerce, Travel, Healthcare or Hospitality.

They are either purpose-built for these domains and optimised for them. These voice assistants have much higher accuracy and support use cases relevant to the domain leading to a better user experience. This is possible due to the boundaries of information that come with a particular domain.

Figure 13: Domain Specific Voice Assistant

16 Benefits Of Voice Assistants

Figure 14: Benefits Of Voice Assistants

1. More convenience to users
   Using voicebots is an intuitive way to help answer customer queries that offer the convenience of talking to a person in a natural language.

   Customers can talk to the voicebot just as they would with a live agent to seek support. All without the need for constant human intervention for repetitive questions.

2. Faster and tailored resolutions
   Voicebots can be deployed to answer queries and assist customers with making informed decisions around the clock. They can pick up from previous conversations and devise personalised answers and resolutions that fit the user query the best.

   Users can access AI voice assistants at any time to get query resolutions instantaneously. This does away with the wait times and support queues creating happier customer experiences.

3. Reduced handling time
   One of the most important metrics in customer support is the average handling time or AHT. AI voicebots can help reduce the average handling time drastically. Why is that? Because voice input and output time is lesser than the time needed to type out a query and read the response on a chatbot. This eases the load off your support staff and results in quicker resolution cycles.

4. Lowered costs
   An AI voicebot can answer multiple users at the same time without training and equipping support staff to help manage high support ticket volumes. The reduced need for human resources can save you a lot of money in the long run.

5. Precision and accuracy
   Voicebots are powered by artificial intelligence which constantly learns from previous interactions. This makes voicebots better at answering a higher volume of search queries every time. Built on sharp AI algorithms, voice assistants are extremely precise and accurate with their answers.

6. More conversions
   An AI voicebot can boost conversions by engaging with users before they abandon their carts. It can quickly ask and gauge the user's requirements to come up with a directed solution best for their use case.

17 Challenges faced by Voice Assistants

17.1 Privacy

Privacy is a huge concern, especially when it comes to smart speakers; they always listen for their wake word, posing a huge privacy concern.

The crucial detail that is often missed out on is the difference between listening and recording. Once these speakers or voice assistants get activated using the wake word, they start recording the audio.

17.2 Accuracy

Voice Assistants don't always understand what's spoken. There could be many reasons behind these- sometimes it could be because of how we say, our accent can cause that. Sometimes, it could be because the voice assistant simply doesn't know what to do with your question. After all, it doesn't have any instructions related to your query.

17.3 Lack of vernacular Support

Speech recognition, perhaps the most critical component of a Voice Assistant, is not available for a lot of languages spoken around the world. The problem is not only limited to speech recognition but also extends to other critical functional areas of Voice Assistants.

18 Future of Voice Assistants

Advancements in Artificial Intelligence and Machine Learning are truly revolutionizing how we use voice assistants in our daily lives.

With voice now establishing itself as an ultimate mobile experience, businesses are only beginning to understand how they can integrate voice in all their activities. A recent report by PwC reveals that the adoption of voice assistant technology is highest among 18-24-year-olds. But the group that uses voice assistants most frequently is the 25-49-year-old group.

The coming time presents many opportunities for voice to grow by leaps and bounds, but a lack of skills and knowledge makes it difficult for businesses to get on board with a voice strategy. If one is in it for the long run, the voice will present an opportunity to understand and provide experiences to your consumers like never before.