

Voice Enabled Application for Physically Challenged Individuals

Poorva Datar¹, Prachi Malewar², Yuktee Ramteke³

^{1,2,3}Student, Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India

Abstract: Physically disabled and mentally challenged people are an important part of our society that has yet not received the same opportunities as others in their inclusion in the society. Therefore, it is necessary to develop easily accessible web applications to help them in transportation because this is the most required field to work upon. The proposed voice enabled web system integrates multiple channels and provides users with choices for accessing the service of transportation. It gives users options to meet their needs, fit their preferences, or overcome environmental constraints. This project presents a speech to text conversion that will extract domain specific information from a user oriented web application based on transportation. Firstly, the user will provide its voice or speech to the system. The speech to text conversion can be carried out through an API like Google Cloud API. Google Cloud API interface allows you to easily add the everything from storage access to machine learning analysis and helps to build the communication through various components. And then the supervised machine learning classifier will perform various actions based upon the words extracted from the Google API. This will help the physically challenged people move from one place to another.

Keywords: Web applications, voice enabled web system, Google Cloud API, machine learning analysis.

1. Introduction

A. What is voice enabled application?

A voice-user interface (VUI) makes spoken human interaction with computers possible, using speech recognition to understand spoken commands and questions, and typically text to speech to play a reply. Voice user interfaces have been added to automobiles, home automation systems, computer operating systems, home appliances like washing machines and microwave ovens, and television remote controls. They are the primary way of interacting with virtual assistants on smartphones and smart speakers. Newer VCDs are speaker-independent, so they can respond to multiple voices, regardless of accent or dialectal influences. They are also capable of responding to several commands at once, separating vocal messages, and providing appropriate feedback, accurately imitating a natural conversation.

A VUI is the interface to any speech application. Controlling a machine by simply talking to it was science fiction only a short time ago. Until recently, this area was considered to be artificial intelligence. However, advances in technologies like text-to-speech, speech-to-text, Natural Language Processing,

and cloud services, in general, contributed to the mass adoption of these types of interfaces. VUIs have become more commonplace, and people are taking advantage of the value that these hands-free, eyes-free interfaces provided in many situations.

B. Why for physically challenged individuals?

Sections in our society still persist, not just on the basis of traditional autocracies but also on physical capabilities. We propose to blur that line a little bit more. Here we present a go to application which can be used well by around 56.7 million of 'just physically impaired' and the even more 'lazy lot' of us. This corresponds to about 15% of the world's population. In reference to India's population, around 19% in seeing, 20% in movement and 8% in multiple disabilities persist and are seen to be growing in number. It's not easy for them to access web applications. Characters become sounds for the blind. Sometimes layouts are meaningless to the mentally ill people. Those who don't have hands can't use the mouse or keyboard. We cannot understand the global picture of people with disabilities, their needs, and the barriers they face to participate completely in their societies. Mainly people with disabilities face barriers in accessing services like health, education, daily needs, transportation. We believe that the most required and essential thing they might be a facing a lot is transportation and also nothing much is done in this field till date and therefore we are taking a step forward to build such a web application that will recognize their voice and will help them moving from one place to another. The people who cannot see, move, don't have hands or have Dementia or Autism or suffering through various other movements and mental disorders would not depend on anyone and can easily move from one place to another. And we are trying to make this possible by building a voice based web application.

C. Issues

Voice interfaces pose a substantial number of challenges for usability. In contrast to graphical user interfaces (GUIs), best practices for voice interface design are still emergent.

1) Discoverability

With purely audio-based interaction, voice user interfaces tend to suffer from low discoverability: it is difficult for users to understand the scope of a system's capabilities. In order for the system to convey what is possible without a visual display,

it would need to enumerate the available options, which can become tedious or infeasible. Low discoverability often results in users reporting confusion over what they are “allowed” to say, or a mismatch in expectations about the breadth of a system’s understanding.

2) *Transcription errors*

While speech recognition technology has improved considerably in recent years, voice user interfaces still suffer from parsing or transcription errors in which a user’s speech is not interpreted correctly. These errors tend to be especially prevalent when the speech content uses technical vocabulary (e.g. medical terminology) or unconventional spellings such as musical artist or song names.

2. Methodology

A. *Python Installation*

Python is a popular high-level programming language used for general purpose programming.

- Open command prompt in your system and then run these commands on command prompt to install pythons.
\$ sudo apt-get update
\$ sudo apt-get install python3.6
- Install pip and import packages from python using get-pip command. Pip is a python based package manager that keeps all the python modules up-to-date.
\$ python -m pip install
\$ python -m get-pip.py
- Import files by using Json library using “import json” command. Json library is comprised of all the latest features that can be used in python.

B. *Google API*

[6] Google APIs is a set of application programming interfaces (APIs) developed by Google which allow communication with Google Services and their integration to other services. Examples of these include Search, Gmail, Translate or Google Maps. The APIs provide functionality like analytics, Usage of some of the APIs requires authentication and authorization using the OAuth 2.0 protocol. OAuth 2.0 is a simple protocol. To start, it is necessary to obtain credentials from the Developers Console. Then the client app can request an access token from the Google Authorization Server, and uses that token for authorization when accessing a Google API service. There are client libraries provided by google in various languages which allow developers to use Google APIs from within their code, including Java, JavaScript, .NET, Objective-C, PHP and Python.

C. *Maps.html*

Google Maps is a web mapping service developed by Google. It offers satellite imagery, aerial photography, street maps, 360° panoramic views of streets (Street View), real-time traffic conditions (Google Traffic), and route planning for

traveling by foot, car, bicycle and air (in beta), or public transportation. The project overall deals with the chosen domain of transportation mainly focusing on the journey of booking a cab to the dropping station. For this to be achieved, the back end needs a map to draw results from. We basically on getting start and end locations entered via voice by our customer which will directly link to the Google maps database. Choosing Google maps was the most efficient idea due to the fact that it mostly recognizes maximum number of languages and also consists of a huge database of locations almost accurate and most relevant. After the user gives us the two points: source and destination, the maps.html file links to the Google map and finds the most relevant latitudes and longitudes of the same and returns an optimized path leading us to reach the shortest and hassle free path.

1) *Google Latitude*

Google Latitude was a feature from Google that lets users share their physical locations with other people. This service was based on Google Maps, specifically on mobile devices. There was an iGoogle widget for Desktops and Laptops as well.

2) *Features*

- *Place Details:* Provide names, addresses, and other rich details like ratings, reviews, or contact information for over 150 million places.
- *Current Place:* Identify a place based on real-time signals like time of day or user location.
- *Find Place:* Turn a phone number, address, or name into a place.
- *Autocomplete:* Automatically return location suggestions while users type.
- *Geocoding:* Convert addresses to geographic coordinates, or the reverse.
- *Geolocation:* Return the precise location of a device based on Wi-Fi or cell towers.
- *Time Zone:* Return a time zone for any location.

[3] Google Earth is a computer program that renders a 3D representation of Earth based on satellite imagery. The program maps the Earth by superimposing satellite images, aerial photography, and GIS data onto a 3D globe, allowing users to see cities and landscapes from various angles. Users can explore the globe by entering addresses and coordinates, or by using a keyboard or mouse.

3. Working

The project can be divided into two important parts; one being the speech recognition part and the other being the actual machine learning part. The final product is conveniently an amalgamation of these two majors. The first part is speech recognition so the input is given via a microphone in speech format. It then undergoes pre-processing. In preprocessing phase, all the unwanted background noise has been removed using python low pass filter. Then the output of pre-processing phase goes as an input to conversion phase. In this phase, the

speech is converted into text using a Google API (Application Program Interface). The output after this conversion is then given as input to machine learning classifier.

Following it in the second part, the user provides the start and the end location and the interface will search the provided locations in reference with the google maps. The interface is linked with google maps so that the user can provide any source or destination location all across the globe and the interface will be able to detect it. All the data required like GPS coordinates, the latitude and longitude points are stored in Firebase database which is a database required to store specific details related to maps. The database files will be mapped with the google maps and this provided the cabs the intended location where they want to move and hence the final part as said will be a merger of these two parts.

An application will be prepared through which the user will provide the speech which is recognized by the system and the actions will be performed like to book a cab, provide the source and destination through voice and the google maps will help the cabs to locate the exact place provided by the user. The data stored in database will be mapped with google maps and will show the cabs the required direction to reach the final destination. In this way, a physically challenged individual will be able to move from one place to another.

The figures given below give us the rightful idea of those two parts. as a service or access to user data.

4. Conclusion

A voice enabled application is a new attempt to make the life of the suffering a little more easier. Here other than the regular websites that uses a Google voice searching system, we offer a completely independent solution to enjoy the easy access to a vehicle to not only all the people who are fit enough but also to those who aren't. The people with disabilities are mainly suffering while moving from one place to another and there are many other domains too but transportation is one among those barriers which restrict them to move from one place to another. Helping them in one or the other way not only provides them happiness but also helps them to uplift in the society and therefore taken a step forward to help these people.

References

- [1] M. Lucente, "Conversational Interfaces for Applications", Comm. ACM, vol. 43, no. 9, pp. 59-61, 2000.
- [2] N. Anerousis, and E. Panagos, "Making Voice Knowledge Pervasive", IEEE Pervasive Computing, vol. 1, no. 2, pp. 42-48, 2002.
- [3] W. Chou, L. Li, F. Liu, "Web Service Enablement of Communication Services", Proceedings of IEE International Conference on Web Services (ICWS'05), vol. 2, pp. 393-400.
- [4] L. Li, W. Chou, "Two Way Web Service: From Interface Design to Interface Verification", Proceedings of IEEE International Conference on Web Services (ICWS'05), vol. 2, pp. 525-532, July 2005.
- [5] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, J. Tygar, "Adversarial machine learning", Proceedings of the 4th ACM workshop on Security and artificial intelligence, pp. 43-58, 2011.
- [6] https://en.wikipedia.org/wiki/Google_APIs

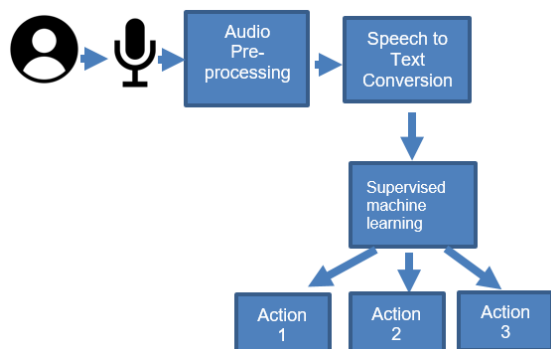


Fig. 1. Part I

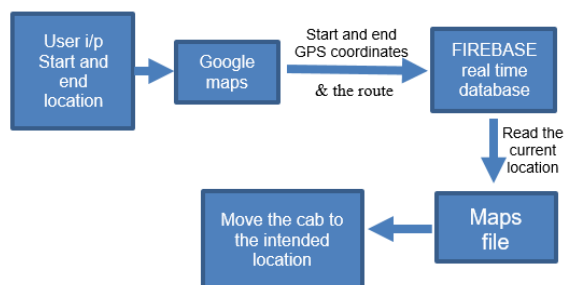


Fig. 2. Part II