

# **An Obliging Handheld Auditory Helper for Travellers**

**Mayuresh Agrawal**  
Dept. of Information  
Technology,

K.J. Somaiya College of  
Engineering,  
Mumbai, India

**Smruti Berad**  
Dept. of Information  
Technology,

K.J. Somaiya College of  
Engineering,  
Mumbai, India

**Twara Gandhi**  
Dept. of Information  
Technology,

K.J. Somaiya College of  
Engineering,  
Mumbai, India

**Deepti Patole**  
Dept. of Information  
Technology,

K.J. Somaiya College of  
Engineering,  
Mumbai, India

## **ABSTRACT**

Despite the various options available today to a person to aid with reading disabilities and language barriers, there is still an absence of a system which provides a solution to both these problems together in a portable and easy to use form. There is a need of a system which can help auditory learners, travellers in a foreign country, dyslexic individuals etc. to overcome the hurdles they face. The system proposed in this paper puts forth the idea of an Android application which recognizes and dictates text from an image all the while employing translation if needed. The application will let it's user scan or upload an image, obtain the text from the image for translation to another language, have the text spoken out loud or save the content in the form of a document.

The paper further explains each aspect of the idea in detail and focuses on portraying a clear description of Optical Character Recognition, Text to Speech, Machine Translation and Document Saving which will help paint a clear picture for the reader.

## **Keywords**

Android; Optical Character Recognition; Text to Speech; Machine Translation; Natural Language Processing; Artificial Intelligence.

## **1. INTRODUCTION**

The main idea behind the system proposed in this paper is to create a mobile application which will let the user scan or upload an image of his/her liking. The application will then let the user recognize the text from the image and will have functionalities using which the user can hear the recognized text out loud in the language of choice, save the text into a document and even translate the text into a desired language. Thus, the main modules in the application consist of Optical Character Recognition [1], Text to Speech [2], Machine Translation [3] and Document Saving. Most of these topics fall under the umbrella of Artificial Intelligence [4] and require providing a knowledge base which can help the application to cater to the needs of the user and provide results.

To generate the knowledge bases which are in reality the pillars on which the application rests, Application Program Interfaces [5] can be used. These Interfaces provides one with the interaction of the software components by providing tools, protocols and routines to build the application as per the requirement. The Application Program Interfaces are tied to the components of the Android phone on which the application is to be used and would need access to the camera, microphone, media gallery etc.

A mobile application is picked as the medium over which, the idea proposed can be implemented due to the widespread mobile usage all over the world today. Smartphones are owned by majority of the population of the world today and are frequently

used for every day convenience. Apart from the usual purpose of communication, one can now uses a mobile to share pictures and videos, navigate, find cooking recipes; keep track of their daily activities and so on. According to a study conducted by comScore [6], out of the digital time devoted by persons across the United States, which is amongst the countries with the most digital use [7], mobile represents 65% of the digital media time whereas desktops account for only 35% of the time. Hence mobiles are more convenient to use and can be used on the go. Moreover, this application differs mainly from Google Translate [8] and other applications providing similar features over the fact that the text to be recognized need not be highlighted every time, the text can be saved into an editable document and the application is completely free.

The proposed idea, as the name suggests, is used as an auditory helper as it enables the user to hear the content rather than to read it. It can be used by a wide array of users. It would be helpful to students who prefer auditory learning. They can simply scan their textbooks for the content and have it read out to them which would be similar to a teacher reading the content. This will help in a better understanding of the study material. Users who find themselves in a foreign country without much knowledge of the native language used can make use of the application to scan board signs and have it translated to a language they understand. Similarly if the user cannot read or write and only knows a specific language, he/she can always scan the text and hear it in the language they know. If the application is implemented on a larger scale like on a Go Pro camera, it can essentially make a blind user independent by helping to make sense of the surroundings. The uses of the system are multiple and can be quite advantageous to people all over the world.

## **2. SECTIONS**

For many years now, subjects like Optical Character Recognition, Text to Speech and Machine Translation have been studied in detail and used for various purposes. There are many pioneers in this field who have found a way to implement these technologies in our lives for our use and convenience.

To enhance the idea of the system presented in this paper, some literature was surveyed written by people who proposed ideas to implement these technologies into projects. These literature materials gave us some valuable insight into the elements of our idea and helped us understand it in depth so that we could easily implement them into our application as well. Some of the materials which were referenced are:

### **2.1 Autonomous OCR Dictating System for Blind People**

In the paper an idea of customizable gear for blind users is proposed. The gear consists of glasses fitted with HD camera and a Bluetooth headset. The headset guides the user to hold up

a piece of text correctly within the recognizable limit using input from the glasses. The image is then calibrated and processed such that optimal results can be achieved and read out to the people with disabilities. [9]

## 2.2 A Text to Speech Environment: Applications in Education

This piece of literature describes the working of Text to Speech in detail. It explains how the input is first normalized as it contains different punctuations and abbreviations. Sounds are associated to the words and are enhanced to give a smooth output which is the text converted in a speech format. Applications in the field of education are mentioned to emphasize the importance of the concept. [10]

## 2.3 An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism

The main aim of the authors is to provide a comprehensive study of language translation. They provide an intensive description of the different approaches to translations and the systems available today which can be used for this purpose. They go on to propose a system which will take a sentence as an input, break it into pieces to identify the verbs and nouns and different parts of speech and will then go ahead to translate the sentence into another language by finding the counterparts of the identified words. [11]

## 3. PROPOSED SYSTEM

The following modules have been included in the system proposed in this paper so as to provide the user with a variety of functionalities which will help him/her achieve the desired use:

1. Optical Character Recognition
2. Text to Speech
3. Machine Translation
4. Document Saving

### 3.1 Optical Character Recognition

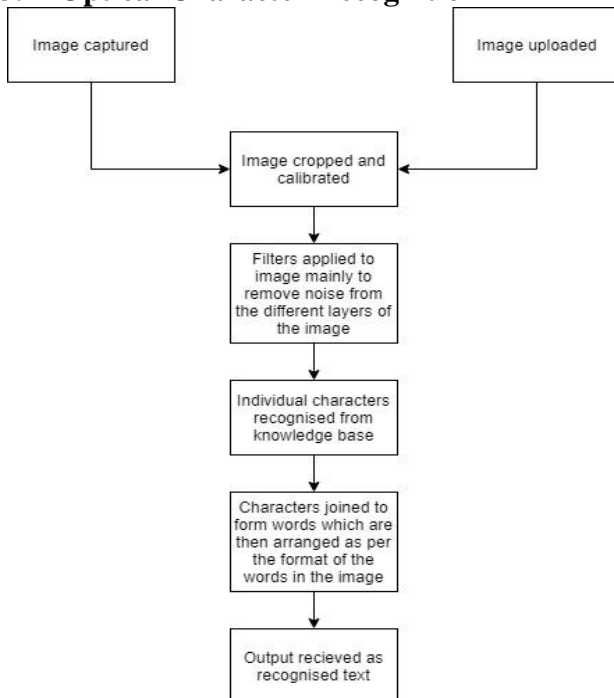


Fig. 1. Block Diagram demonstrating Optical Character Recognition

In the Optical Character Recognition module the application will

be responsible to recognize the text present in the image. What Optical Character Recognition does is it will essentially recognize all the characters in the words from the image, identify which letters they are and give us an output of all the letters which form the sentence resulting in the constitution of the text. Thus the module can identify text from a wide array of written materials like newspapers, road signs, handwritten notes etc. Optical Character Recognition falls under the topic of Artificial Intelligence as it is work done by a machine to understand the characters presented to it from a wide knowledge base of characters it contains.

The module first obtains the image from either a camera or the photo gallery of a phone which require the enabling of permissions and then formats the image. Formatting images make the words clearer. Filters are applied to the image so that character recognition becomes easier for the application. Thus if a picture is poorly lit or shakily taken, processing it can make its quality better and in turn give us more efficient results. Multiple passes of formatting could be implemented which lead to more accurate results. The key to correct character recognition lies in the knowledge base. The more varieties of characters the application is capable of recognizing, the easier it is to achieve accuracy. There are many engines available freely to anyone who wishes to implement Optical Character Recognition amongst which Tesseract [12] is widely used.

The obstacles which the module can face are mainly due the availability of a plethora of text material in the world. The user can make use of the application to obtain text from any of the hundreds of fonts available today or newspapers which have small fonts or even handwritten material which is problematic since every person has a unique handwriting which is not always legible. Picture quality is also a contributor towards optimal results and hence phone camera quality or the picture quality can have a huge impact on the output.

### 3.2 Text to Speech

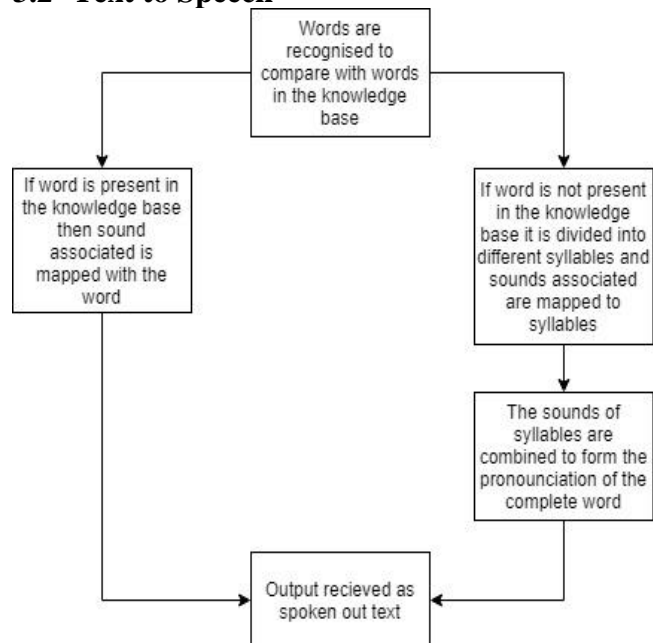


Fig. 2. Block Diagram demonstrating Text to Speech

Using the Text to Speech module an auditory dimension is added into the application which gives the user an option to listen to the result obtained from the image provided to the application. Thus a spoken sound is created out of the text which

has been previously recognized. The module contains a knowledge base in which the basic pronunciations are recorded. The sounds of certain letters are maintained as well as the sounds of specific words. These sounds are compiled for many different languages with the colloquial intonation. When the module comes across a certain word it will first try to identify it. If it recognizes it and has a sound associated to it, the word is spoken out. If for some reason the word is not recognized or its sound not known the software then bisects the word into various different syllables known to it then speaks the word out as a combination of these syllables in the accent matching the one from the language selected. The language option also enables the software to understand the different words belonging to various languages.

Text to Speech is a part of Artificial Intelligence since a machine is responsible for the input of words, understanding them and making a decision to emit a sound matching these words. Many types of software extend the functionality of Text to Speech like Android, Microsoft etc.

The module can run into issues sometimes. If the output from the Optical Character Recognition module is not accurate it sometimes only recognizes some of the characters and arranges it as letters instead of a complete word. In such cases the letters will be read out separately and would not make sense. A common error would be mispronouncing the word, pronouncing it different from the chosen dialect or pronouncing the word in a way which is not recognizable to the user.

### 3.3 Machine Translation

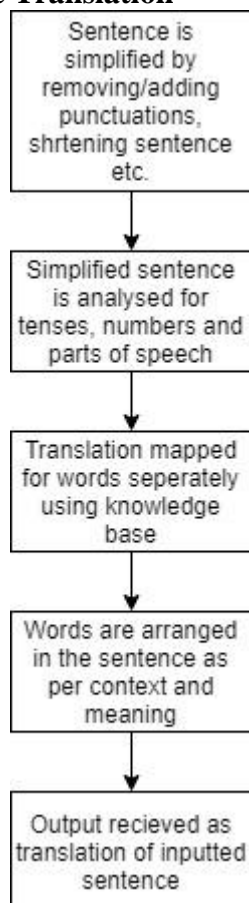


Fig. 3. Block Diagram demonstrating Machine Translation

The Machine Translation module can be used by the user to change the content the user has obtained after recognition to a language of choice. This option has been especially implemented for native speakers or people who are more comfortable with a particular language. The feature also enables people all over the world to use the application.

Machine Translation is an old branch of Artificial Intelligence as it requires intelligence shown by the machine to convert any given text into a proper one in the needed language and is used in numerous ways like bilingual communication, translation aid etc.[13] Translation is a very complicated subject due to the wide array of vocabulary every language contains. The knowledge base for translation essentially consists of all the words in a given language. These words are then mapped to words of similar meaning in the other languages such that when an English user needs to say “Hello” to someone in French the knowledge base will help find the consequently mapped word for it and return “Bonjour” as a result. The mapping is not only for specific words with similar meanings but also expressions and phrases which are commonly used in languages. For our application, the knowledge base will receive content from the Character Recognition module and will give an output of words with a similar meaning to convey the message from the scanned/uploaded text in a language the user is comfortable with.

Working of the module consists of various steps [14]. The Translation module will first need text as input. The obtained text is formatted into a simpler version by removing or adding punctuations, shortening the sentence etc. This is done so that the meaning of the sentence is clear to the translator. The text is then analyzed for tenses, numbers and parts of speech to determine the basic gist of the text. Once an understanding is established words from the target language which match the ones from the text are generated and are then combined keeping in mind that the basic meaning of the translated text is similar to the one inputted and the translated text is correct grammatically. There are many Application Program Interfaces which can be employed to use like the ones offered by Google or Bing. Some of these are free whereas the more advanced ones require payment. [15]

Languages are not easy to map. There is a huge volume of words in even the simplest language and these words can form many different combinations with each other to give us many different meanings. Keeping a track of all these words in all of these languages can become a cumbersome task. Words can be missed out or sometimes the words can be translated to a different meaning if the knowledge base is not very advanced. Additionally, it is not possible to have translations in every single language, especially the lesser known native ones. India is a country of more than 1800 different languages. [16] Having knowledge bases for each and every language could prove to be tedious. We could have translations to Hindi as it is a language which is spoken by a larger part of the population, but translations in a language like Dogri could not be included for practical purposes.

### 3.4 Document Saving

A Document Saving functionality is also included in the Android application. When the person needs to save the content which has been recognized or have a copy of it to export and make changes in later, the text can be saved in the form of a document. The document will be in a Portable Document Format and would be saved on the application and will have the option to save it into the phone as well. All the text which has been recognized would be arranged in a similar manner as the picture

it was obtained from. After arranging it as per the requirement, the text will be saved into a document where the user can make changes if he/she requires. The user only needs to give the application permission to save the document into the phone.

The accuracy of the document will be based upon the accuracy from the Character Recognition module. A problem the user might face could be improper arrangement of the text in the document. Sometimes due to poor picture quality the application might not be able to analyze the image properly enough to understand its format. Thus the document would not be a perfect replica of the image in such cases.

#### 4. RESULTS AND DISCUSSION

The system proposed in this paper was implemented as an Android application and a variety of different inputs were tested across the different modules of the application.

For the Optical Character Recognition module different written materials such as newspaper articles, handwritten notes, websites, textbooks etc. were checked for recognition accuracy. As demonstrated in Table-I, the accuracy observed from the test cases was 84% on using the Tesseract Application Program Interface. 100 images were tested to achieve the level of accuracy amongst which were images of newspaper clippings, webpages, handwritten notes etc.

$$\text{Accuracy of the Optical Character Recognition Module} = \frac{\text{Text present in picture} - \text{Recognised text in OCR module}}{\text{Text present in picture}} * 100$$

The Text to Speech module was tested for the clarity of speech. The clearness of speech and proper pronunciation of the word was evaluated and the accent for each language was inspected for accuracy. Texts which were easy to pronounce along with the ones which were a little harder were supplied to the module and an accuracy of 82% was observed on using the Text to Speech Application Program Interface provided by Android. [17] The text recognized from the 100 images in the Optical Character Recognition module was consequently used to be read out to check for correctness.

$$\text{Accuracy of the Text to Speech Module} = \frac{\text{Recognised text in OCR module} - \text{Text read out by TTS module}}{\text{Recognised text in OCR module}} * 100$$

Lastly, the Machine Translation module was checked to ensure that the correct translation was achieved between two languages chosen by the user. The translation obtained from the Google Translate Application Program Interface was checked for the correct words translated along with the contextual meaning and fit of the word in the sentence and the accuracy of the output was observed to be 87%. To calculate the accuracy of the module, the text obtained from the 100 images entered into the Optical Character Recognition module were all written in English. The translation from English to Hindi was evaluated.

$$\text{Accuracy of the Machine Translation Module} = \frac{\text{Recognised text in OCR module} - \text{Text translated by translation module}}{\text{Recognised text in OCR module}} * 100$$

**Table 1. Simulation Results**

No	System Module	Accuracy
1.	Optical Character Recognition	84%
2.	Text to Speech	82%
3.	Machine Translation	87%

#### 5. CONCLUSION

To conclude, this paper talks about an idea which is aimed at helping auditory learners, people with reading disabilities or people who need language translation due to the ability to understand only a certain language. The system will be implemented in the form of an Android application for convenient usability by a person and will contain modules which offer Optical Character Recognition, Text to Speech, Machine Translation and Document Saving as functionalities.

The idea proposed in this paper could be expanded further to suit a wider target audience. It can be used to implement on a device with sensors or portable camera like a GoPro such that it would make a blind person independent of any human support and help. It could also employ other interactive functionalities like speech to text which could make the application user friendly and could suit the needs of a different target audience than the one considered for this project.

#### 6. ACKNOWLEDGMENTS

We would like to express our gratitude to our professor Deepti Patole for helping us formulate and enhance the idea of our application, suggest changes to make our work easier and for guiding us throughout the time period of the system implementation.

#### 7. REFERENCES

- [1] Sarika Pansare, Dhanshree Joshi, "A Survey on Optical Character Recognition Techniques," in International Journal of Science and Research (IJSR)
- [2] M.H. O'Malley, "Text-to-speech conversion technology," in Computer ( Volume: 23, Issue: 8, Aug. 1990 ) Machine Translation Whitepaper. (2016, January). Retrieved from [https://en-gb.thebigword.com/custom-uploads/20160129083810\\_MachineTranslationWhitepaperJan2016.compressed.pdf](https://en-gb.thebigword.com/custom-uploads/20160129083810_MachineTranslationWhitepaperJan2016.compressed.pdf)
- [3] Bertram Raphael, "Artificial intelligence," in Computer ( Volume: 6, Issue: 5, May 1973 )
- [4] What is an API? Your guide to the internet business (r)evolution. Retrieved from <https://www.3scale.net/wp-content/uploads/2012/06/What-is-an-API-1.0.pdf>
- [5] All digital growth now coming from mobile usage — comScore. (2016, April 3). Retrieved from <https://marketingland.com/digital-growth-now-coming-mobile-usage-comscore-171505>
- [6] Mobile Marketing Statistics compilation. (2018, January 30). Retrieved from <https://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/>
- [7] Google Translate Help. Retrieved from <https://support.google.com/translate/answer/6142483?co=GENIE.Platform%3DAndroid&hl=en>
- [8] Christos Liambas and Miltiadis Saratzidis, "Autonomous OCR dictating system for blind people," in IEEE 2016 Global Humanitarian Technology Conference (GHTC).
- [9] Gerard K. Rambally, "A text to speech environment: applications in education," in IEEE Proceedings of Southeastcon '91.
- [10] Jayashree Nair, Amrutha Krishnan K and Deetha R, "An efficient english to hindi machine translation system using hybrid mechanism," in 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI),

Sept. 21-24, 2016, Jaipur, India.

- [11] Ray Smith, "An Overview of the Tesseract OCR Engine," in Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
- [12] John Hutchins, "Uses and applications of machine translation", Presentation, Westminster University. <http://www.hutchinsweb.me.uk/Westminster-2009.pdf>
- [13] Machine Translation Process. (2010, November 1). Retrieved from <http://language.worldofcomputing.net/category/machine-translation>
- [14] Which Translation API Should I Choose?. Retrieved from <https://www.smartling.com/translation-software/translation-api/>
- [15] M. Hanumanthappa, M. Narayana Swamy, "A Detailed Study on Indian Languages Text Mining," in International Journal of Computer Science and Mobile Computing, Vol.3 Issue.11, November- 2014, pg. 54-60
- [16] Text to Speech. Retrieved from <https://developer.android.com/reference/android/speech/tts/TextToSpeech.html>