# Android Based Smart Speech Recognition Application to Perform Various Tasks

Lakshya Aggarwal
Department of Computer Engineering
Maharashtra Institute of Technology, Pune
*Email: lakshya.aggarwal.007@gmail.com*

Mayank Kaushik
Department of Computer Engineering
Maharashtra Institute of Technology, Pune
*Email: mkmayankaushik@gmail.com*

***Abstract*—** A smart speech recognition application is theorized in this paper. User can control a variety of applications on an android based platform, which include native applications as well as user installed applications with voice commands. These include - calling, texting, switching on and off sensors (Wi-Fi, GPS, Bluetooth),setting alarms. The application provides online as well as offline services. The application also applies machine learning concepts to identify usage patterns and create an environment which anticipates user requirements. The tasks being performed repetitively are automated. Services of activity recognition, recognizing nearby friends using Bluetooth are performed. The importance of the project is that it provides visually challenged people as well as the general population an alternate and a very easy way to control applications on android smart phones.

***Keywords*-** *Activity Recognition, Android, Automated, Bluetooth, GPS, Machine Learning, Smart Speech Recognition, Visually Impaired, Wi-Fi.*

*****

## I. INTRODUCTION

Speech recognition (SR) is a field of computer science that deals with recognition and conversion of spoken language into text by computational devices. It incorporates research and methodologies from various other fields of computer science such as NLP, DSP.

Some SR systems use training wherein an individual speaker reads text or isolated vocabulary into the system. The system records this and analyzes that person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker independent" systems while systems that use training are called "speaker dependent".

Speech recognition has huge list of applications. One of it's application is user input for mobile applications. Google's Voice Actions and Iphone's Siri are applications that enable control of a mobile phone using voice, such as calling, businesses and contacts, sending texts and email, listening to music, browsing the web, and completing common tasks[1].

Both Siri and Voice Actions require an active connection to a network in order to process requests and most of Android phones can run on a 4G network which is faster than the 3G network that the iPhone runs on. There is also an issue of availability, Voice Actions are available on all Android devices above Android 2.2, but Siri is available only for owners of the iPhone[1].

In this work we have theorized the creation of an application that provides voice commands to call a person from contacts, send SMS messages, set alarms, turn on and off Bluetooth,Wi-Fi and GPS using Google speech recognition engine. The main goal of this application is to provide visually challenged persons with easy access to features of a smartphone. This will also set a precedent towards making smart phones completely touch free.

## II. LIERATURE SURVEY

In the Google API to be used in our application a large vocabulary speech recognition system is described that is accurate, has low latency, and yet has a small enough memory and computational footprint to run faster than real-time on an Android Smartphone. It employ a quantized Long Short-Term Memory (LSTM) acoustic model trained with connectionist temporal classification (CTC) to directly predict phoneme targets, and further reduce its memory footprint using an SVD-based com-pression scheme. Additionally, it minimizes our memory footprint by using a single language model for dic-tation and voice command domains, constructed using Bayesian interpolation. Finally, in order to properly han-dle device-specific information, such as proper names and other context-dependent information, it injects vocabulary items into the decoder graph and bias the language model on-the-fly. This system achieves 13.5percent word error rate on an open- ended dictation task, running with a median speed that is seven times faster than real-time.[3]

Voice SMS is an application that allows a user to record and convert spoken messages into SMS text message. User can send messages to the entered phone number or the number of contact from the phonebook. Speech recognition is done via the Internet,connecting to Google's server. The application is adapted to input messages in English. Used tools are Android SDK and the installation is done on mobile phone with Android op- erating system. In this article basic features of the speech recognition and used algorithm have been explained. Speech recognition for Voice SMS uses a technique based on hidden

Markov models (HMM - Hidden Markov Model). It is currently the most successful and most flexible approach to speech recognition.[1]

Falling is one of the most common accidents with potentially irreversible consequences, especially consid-ering special groups, such as the elderly or disabled. One approach to solve this issue would be an early detection of the falling event. Towards reaching the goal of early fall detection, this paper works on distinguishing and monitoring some basic human activities such as walking and running. Since we plan to implement the system mostly for seniors and the disabled, simplicity of the usage becomes very important. This paper successfully implements an algorithm that would not require the acceleration sensor to be fixed in a specific position (the smart phone itself in our application), whereas most of the previous research dictates the sensor to be fixed in a certain direction. This algorithm reviews data from the accelerometer to determine if a user has taken a step or not and keeps track of the total amount of steps. After testing, the algorithm was more accurate than a commercial pedometer in terms of comparing outputs to the actual number of steps taken by the user.[5]

Smart devices such as Apples iPod nano (5th gen-eration), Nike+, and existing smartphone applications can provide the functions of a pedometer using the accelerometer. To achieve a high accuracy the devices must be worn on specific on-body locations such as on an armband or in footwear. Generally people carry smart devices such as smartphones in different positions, thus making it impractical to use these devices due to the reduced accuracy. Using the embedded smart-phone accelerometer in a low-power mode this paper presents an algorithm named Energy-efficient Real-time Smartphone Pedometer (ERSP), which accurately and energy-efficiently infers the real-time human step count within 2 seconds using the smartphone accelerometer. This method involves extracting 5 features (4 novel and 1 derived) from the smartphone 3D accelerometer without the need for noise filtering or specific smartphone on-body placement and orientation; ERSP classification ac-curacy is approximately 94 percent when validated using data collected from 17 volunteers.[6]

Machine learning: Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a ”Field of study that gives computers the ability to learn without being explicitly programmed”. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions. Alpaydin defines Machine Learning as the capability of the computer program to acquire or develop new knowledge or skills from existing or non-existing examples for the sake of optimizing performance criterion.[7]

Googles Android has become the leading platform in smart phone market. Android smart phones have also integrated various sensors including gyroscope, orienta-tion and accelerometer, so it is quite suitable to design mobile sensor applications. Pedometer is a common aux-iliary device used for maintaining health and fitness. In this paper, an intelligent pedometer is developed using Android. The user’s walking motion was detected via android sensor and pedometer application then analyzes the signal, calculates the walking distance and calories burned, and provides real time feedback to user via Bluetooth. The system provides three action modes: time-based mode, distance-based mode and count-based mode. All the tracking data are saved in SQLite database, and automatic threshold detection is used to improve the accuracy.[8]

## III. ANDROID

Android is a software environment for mobile devices that includes an operating system, middleware and key applica-tions[1]. It provides a rich application framework that allows you to build innovative apps and games for mobile devices in a Java language environment[2].

Android apps are written in the Java programming language. The Android SDK tools compile your code along with any data and resource files into an APK, an Android package, which is an archive file with an .apk suffix. One APK file contains all the contents of an Android app and is the file that Android-powered devices use to install the app[2].

App components are the essential building blocks of an Android app. Each component is an entry point through which the system or a user can enter your app. Some components depend on others.

There are four different types of app components:

1) Activities
2) Services
3) Content providers
4) Broadcast receivers

Each type serves a distinct purpose and has a distinct lifecycle that defines how the component is created and destroyed[2].

Three of the four component types activities, services, and broadcast receivers are activated by an asynchronous message called intent. Intents bind individual components to each other at runtime. You can think of them as the messengers that request an action from other components, whether the component belongs to your app or another[2].

 ISSN: 2321-8169
83 – 86

A. Android Speech Recognition

The speech recognition for the application will be done by using the class provided in android SDk, class Speech Recognizer. This class provides access to the speech recognition service. This service allows access to the speech recognizer.The class is instantiated by calling create Speech Recognizer(Context). This class's methods are invoked only from the main application thread. This API is not intended to be used for continuous recognition, which consumes a significant amount of battery and bandwidth. This API requires to have RECORD AUDIO permission to use this class[2].

The class RecognizerIntent is used for getting the Speech-to-Text conversion result through intents. This text is then used for further processing.

## IV. Application Functionalities

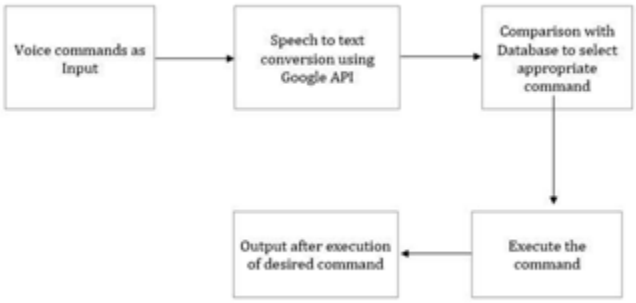


Fig. 1. Speech recognition block diagram

The given "Speech recognition" module is performed in five steps. The initial block is to take input from user in the form of voice commands. Then this is passed to the second block which is the API that converts speech into text. The output of Conversion Block is the command in textual form which is given as an input to the next block. Here the input is compared to the predefined list of commands from the database and the appropriate one is chosen. This command is then executed in the next block. In the last Output Block the command given by the user is saved as a log entry in the Database which will be later used by the Machine Learning module.


Fig. 2. Machine Learning block diagram

The "Machine Learning" module provides multiple functionalities. The first and foremost is the automatic execution of very frequent commands given by the user with users permission. The module uses the command log generated by the Speech Recognition module to search for patterns , rules and learn frequently used commands. It will then ask user permission to execute them automatically. Example is if a user sets an alarm for 6:00 AM frequently via voice commands, then the module recognizes this as a frequent task and will now ask the user to set the alarm next time automatically.


Fig. 3. Activity recognition block diagram

Another functinonality is to count the steps taken by user within a specific period of time using Activity Recognition. This time period is defined by user input . It takes start voice command from the user and stops with the users stop voice command. This requires sensor input from Accelerometer. The flow of control is depicted in fig. This module also notifies the user of friends in the vicinity. This is achieved by searching for available paired bluetooth devices in users surroundings. This will require telling the system about the paired device. This information will be saved in the application database.

## V. Conclusion

This paper theorizes the development of an Android plat-form based smart voice recognition system to operate multiple android apps with simple voice commands. This technology will be implemented in a user friendly and compact device. Since the application will operate in both online and offline modes, it has a decided advantage over other such applications. This project has the capability of modern smart speech recognition software to increase independence for persons with disabilities. Major purpose of this work is to provide a system so that the visually disabled population can easily control many functions of a smart phone via voice. The system is very useful for the general population as well. Users can command a mobile device to do something via voice such as calling, texting, setting alarms. These commands are then immediately executed. The application is also using machine learning concepts to execute frequent voice commands automatically. It is also keeping track of steps taken by user and paired Bluetooth devices in vicinity.

## ACKNOWLEDGEMENT

We would like to thank Varun Kulkarni and Pooja Kamble, students of MIT, Pune for working hard and supporting us to write this paper. Special thanks to Mayank Kaushik for his brilliant work in creating the illustrations used in this paper. We would also like to thank Prof. R.A. Agrawal for her support and guidance.

## REFERENCES

[1] Sanja Primorac and Mladen Russo, Android application for sending SMS messages with speech recognition interface, 35th International Convention MIPRO, 2012.

[2] ”Android Developers”, http://developer.android.com

[3] Google Inc., Personalized speech recognition on mobile devices , Pro- ceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016

[4] Melis Oner, Jeffry A. Pulcifer-Stump, Patrick Seeling and Tolga Kaya, Towards the Run and Walk Activity Classification through Step Detection, An Android Application , Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2012

[5] Thomas Olutoyin Oshin, ERSP: An Energy-efficient Real-time Smart-phone Pedometer, IEEE International Conference on Systems, Man, and Cybernetics, 2013

[6] Dr. Raj Kamal, A. Chaudhary and S. Kolhe Machine Learning Techniques for Mobile Intelligent Systems: A Study , Ninth International Conference on Wireless and Optical Communications Networks, 2012

[7] Wu, Shyi-Shiou and Hsin-Yi Wu, The Design of an Intelligent Pedometer using Android, Second International Conference on Innovations in Bio-inspired Computing and Applications, 2011J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.