International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 03 | Mar-2016 www.irjet.net p-ISSN: 2395-0072

A Cutting Edge Consistent Scheme for Blind People Interaction Using Android Smartphone

P.Mathivanan[1], N.Sudharsanan[2], M.Kambavanan[2], R.Bharath[2]

[1]*Assisant Professor, Department of Information Technology, Manakula Vinayagar Institute of Technology, mathi64it@gmail.com*
[2]*Students, Department of Information Technology, Manakula Vinayagar Institute of Technology, Pondicherry*
[2] *sudhar.crusher@gmail.com, kambavanan@gmail.com, Bharathyan@gmail.com*

Abstract - *Voice SMS application is developed in this paper to allow the user to capture and transform audio messages into a text message. The user can transfer the messages from the Dialer or a predefined contact from Contacts. Speech recognition is performed using the Internet, making a connection to Google's server. The application is set to input the messages in a fairly global language – English. The App is developed using tools from the Android SDK. In this article we have given the basic features of speech recognition and the algorithm we have used. Speech recognition for Voice SMS is mainly based on Hidden Markov Model (HMM). It is considered as the most efficient approach to speech recognition. I.*

Key Words: Voice SMS; audio message; contacts; Android SDK; Markov; approach;

1. INTRODUCTION

These Mobile phones have become an integral part of our everyday life, causing higher demands for content that can be used on them. Smart phones offer customer enhanced methods to interact with their phones but the most natural way of interaction remains speech.

Market for smart mobile phones provides a number of applications with speech recognition implementation. Google's Voice Actions and recently 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. 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 4S. The Siri's advantage is that it can act on a wide variety of phrases and requests and can understand and learn from natural language, whereas Google's Voice Actions can be operated only by using very specific voice commands.

In this work we have developed an application for sending SMS messages which uses Google's speech recognition engine. The main goal of application Voice SMS is to allow user to input spoken information and send voice message as desired text message. The user is able to manipulate text message fast and easy without using keyboard, reducing spent time and effort. In this case speech recognition provides alternative to standard use of Keyboard for text input, creating another dimension in modern communications.

2. ANDROID

2.1 Software environment

Android is a software environment for mobile devices that includes an operating system, middleware and key applications [1]. In 2005 Google took over company Android Inc., and two years later, in collaboration with the group the Open Handset Alliance, presented Android operating system (OS).

2.2 Main features of Android

The Main features of Android operating system are:

- Enables free download of development environment for application development.
- Free use and adaptation of operating system to manufacturers of mobile devices.
- Equality of basic core applications and additional applications in access to resources.
- Optimized use of memory and automatic control of applications which are being executed.
- Quick and easy development of applications using development tools and rich database of software libraries.
- High quality of audiovisual content, it is possible to use vector graphics, and most audio and video formats.
- Ability to test applications on most computing platforms, including Windows, Linux…

The Android operating system (OS) architecture is divided into 5 layers (fig. 1.). The application layer of Android OS is visible to end user, and consists of user applications. The application layer includes basic applications which come with the operating system and applications which user subsequently takes. All applications are written in the Java programming language.

Framework is extensible set of software components used by all applications in the operating system. The next layer represents the libraries, written in the C and C + + programming languages, and OS accesses them via framework.

Dalvik Virtual Machine (DVM) forms the main part of the executive system environment. Virtual machine is used to start the core libraries written in the Java

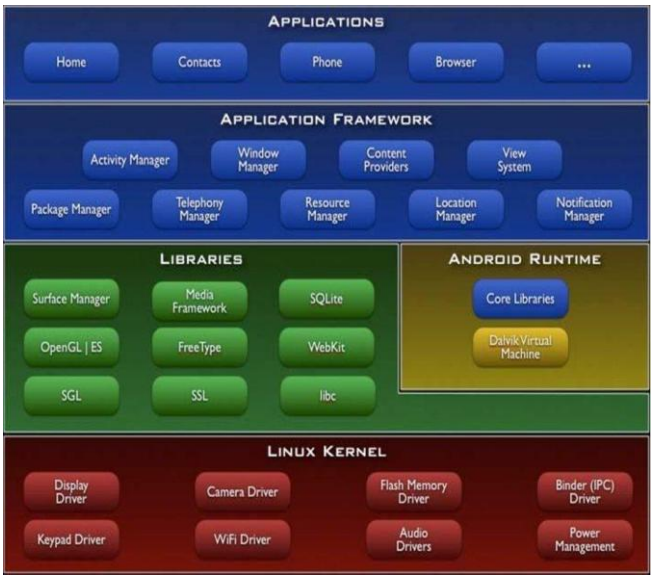

Figure 1. Structure of Android operating system [1]

Programming language. Unlike Java's virtual machine, which is based on the stack, DVM bases on registry structure and it is intended for mobile devices. The last architecture layer of Android operating system is kernel based on Linux OS, which serves as a hardware abstraction layer. The main reasons for its use are memory management and processes, security model, network system and the constant development of systems.

There are four basic components used in construction of applications: activity, intent, service and the content provider. An activity is the main element of every application and simplified description defines it as a window that users see on their mobile device. The application can have one or more activities. Main activity is the one that is used as startup. The transition between the activities is carried out in a way that launched activity calls a new activity. Each activity as a separate component is implemented with inheritance of Activity class. During the execution of applications, activities are added to the stack; currently running activity is on the top of the stack. Intent is a message used to run the activities, services, or recipient's multicast. An intent can contain the name of the components you need to run, the action which is necessary to execute, the address of stored data needed to run the component, and component type. A service is a component that runs in the background to perform long-running operations or to perform work for remote processes. One service can link multiple applications and service is executed until a connection with all applications is done. A content provider manages a shared set of application data. Data can be stored in the file system, a SQLite database, on the web, or any other persistent storage location which application can access [1]. Through the content provider, other applications can query or even modify the data (if the content provider allows it).

3. SPEECH RECOGNITION

Speech recognition for application Voice SMS is done on Google server, using the HMM algorithm. HMM algorithm is briefly described in this part. Process involves the conversion of acoustic speech into a set of words and is performed by software component. Accuracy of speech recognition systems differ in vocabulary size and confusability, speaker dependence vs. independence, modality of speech (isolated, discontinuous, or continuous speech, read or spontaneous speech), task and language constraints [2].

Speech recognition system can be divided into several blocks: feature extraction, acoustic models database which is built based on the training data, dictionary, language model and the speech recognition algorithm. Analog speech signal must first be sampled on time and amplitude axes, or digitized. Samples of speech signal are analyzed in even intervals. This period is usually 20 ms because signal in this interval is considered stationary.

Speech feature extraction involves the formation of equally spaced discrete vectors of speech characteristics. Feature vectors from training database are used to estimate the parameters of acoustic models. Acoustic model describes properties of the basic elements that can be recognized. The basic element can be a phoneme for continuous speech or word for isolated words recognition. Dictionary is used to connect acoustic models with vocabulary words. Language model reduces the number of acceptable word combinations based on the rules of language and statistical information from different texts.

Speech recognition systems, based on hidden Markov models are today most widely applied in modern technologies. They use the word or phoneme as a unit for modeling. The model output is hidden probabilistic functions of state and can't be deterministically specified. State sequence through model is not exactly known. Speech recognition systems generally assume that the speech signal is a realization of some message encoded as a sequence of one or more symbols [3]. To effect the reverse operation of recognizing the underlying symbol sequence given a spoken utterance, the continuous speech waveform is first converted to a sequence of equally spaced discrete parameter vectors. Vectors of speech characteristics consist mostly of MFC (Mel Frequency Cepstral) coefficients [4], standardized by the European Telecommunications Standards Institute for speech recognition. The European Telecommunications Standards Institute in the early 2000s defined a standardized MFCC algorithm to be used in mobile phones [5].

Standard MFC coefficients are constructed in a few simple steps. A short-time Fourier analysis of the speech signal using a finite-duration window (typically 20ms) is performed and the power spectrum is computed. Then, variable bandwidth triangular filters are placed along the perceptually motivated Mel frequency scale and filter-bank energies are calculated from the power spectrum. Magnitude compression is employed using the logarithmic function. Finally, auditory spectrum thus obtained is decor related using the DCT and first (typically 13) coefficients represent the MFCCs.

4. MAIN PARTS OF THE PROJECT

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1 Voice Recognition Activity class

The Voice Recognition Activity is startup activity defined as launcher in AndroidManifest.xml file.

REQUEST_CODE is static integer variable, declared on the beginning of activity and used to confirm response when engine for speech recognition is started. REQUEST_CODE has positive value. Results of recognition are saved in variable declared as List View type.

Method onCreate is called when activity is initiated. This is where the most initialization goes: setContentView (R.layout.voice_recognition) is used to inflate the user interface defined in res > layout > voice_recognition.xml, and findViewById (int) to programmatically interact with widgets in the user interface. In this method there is also a check whether mobile phone, on which application is installed, has speech recognition possibility. PackageManager is class for retrieving various kinds of information related to the application packages that are currently installed on the device. Function getPackageManager () returns PackageManager instance to including technological, economical, legal, and social factors. There is a need for simple, systematic and logical method or mathematical tool to guide decision makers (service requesters) in considering a number of selection criteria and their interrelations. Thus, the Multi Criteria Decision Making (MCDM) theory is a natural fit to the problem of QoS based web service selection. It provides the essential principles to interpret the user' judgment and find global package information. Using this class, we can detect if the phone has a possibility for speech recognition. If a mobile device doesn't have one of many Google's applications which integrate speech recognition, further work of this application Voice SMS will be disabled and message on the screen will be "Recognizer not present". Recognition process is done trough one of Google's speech recognition applications. If recognition activity is present user can start the speech recognition by pressing on the button and thus launching startActivityForResult (Intent intent, int requestCode). The application uses startActivityForResult () to broadcast an intent that requests voice recognition, including an extra parameter that specifies one of two language models. Intent is defined with intent.putExtra (RecognizerIntent.

EXTRA_LANGUAGE_MODEL, RecognizerIntent.LANGUAGE_MODEL_FREE_FORM). The voice recognition application that handles the intent processes the voice input, then passes the recognized string back to Voice SMS application by calling the onActivityResult() callback. In this method we manage result of activity startActivityForResult. Received result Data is stored in ListView variable and shown on the screen.

Interface set OnItemClickListener registers a callback to be invoked when an item on ListView widget has been clicked. Method OnItemClick enables control over statement that user selected from the list. Method contains definition of new Intent and user is transferred into interface for writing SMS. Selected text is transferred from VoiceRecognitionActivity to SMS activity with putExtra (String name, String data) label. The first parameter is a string "key" and the second is the selected text.

4.2 SMS class

Headings In onCreate method label setContentView (R.layout.sms) opens new user interface. Command final String message1 = getIntent().getExtras().GetString ("Key") retrieves the data associated with the string "key" which in this case is selected text from recognition,

4.3 Graphical User Interface

VMS is designed to use simple Graphical User Interface (GUI), as illustrated by figure 1, with four interfaces:

Main menu; Find contact; Message dictation; and Noise level setting.

The GUI of Indo VMS is designed to use Indonesian language since it is for Smartphone users in Indonesia. In the main menu, two microphone symbols are used. The first one located on the right above is to find contact on the phone book. The second one, on the bottom middle, is to dictate a message sentence. This GUI also provides a keyboard to find contact and to write a message. So, a user could use either speech or keyboard to interact with the Indo VMS. The GUI for find contact is deliberately designed to produce some contacts as illustrated. The user could select a desired contact from the list. This design has two reasons, i.e. the contact names are usually similar for some people and it is quite hard for a user to remember a contact name exactly. As illustrated, a contact name “Alifya” has some similar names. The GUI for message dictation provides a wide space for many message sentences and two buttons, the one on the left is to clear a last sentence and another one on the right is to send the message. This GUI gives a full control for user to either add or delete message sentences.

The GUI for setting the noise level is designed simply to set the percentage of the noise level as the threshold of signal energy rate. The default of noise level is set to 3. It means that signal segments below 9% (3 x 3%) of the signal energy rate will be treated as noise and the Pockets phinx does not process them. The maximum threshold is defined to 60% (20 x 3%).

3. CONCLUSIONS

With the development of software and hardware capabilities of mobile devices, there is an increased need for device-specific content, what resulted in market changes. Speech recognition technology is of particular interest due to the direct support of communications between human and computers. The integration of the speech recognition in mobile devices enables easier and faster use of smart phones. Use of such systems opens countless possibilities. Besides to facilitating the manipulation for current users of mobile devices, the market is opened to a whole new range of people who might use these technologies. The elderly, and persons with various disabilities, would have the opportunity to participate in the technological present and feel the benefits of smart phones.

REFERENCES

[1] “Android developers”,http://developer.android.com

[2] J. Tebelskis, Speech Recognition using Neural Networks, Pittsburgh: School of Computer Science, Carnegie Mellon University, 1995.

[3] S. J. Young et al., "The HTK Book", Cambridge University Engineering Department, 2006.

[4] K.Davis, and Mermelstein, P., "Comparison of parametric representation for monosyllabic Word recognition in continuously spoken sentences", IEEE Trans. Acoust., Speech, Signal Process. vol. 28, no. 4, pp. 357–366, 1980.

[5] ETSI, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms.", Technical standard ES 201 108, v1.1.3, 2003.