Personal Assistant with Voice Recognition Intelligence

Dhanashri Sajjan[1], Nehal Waghmare[2], Pinkal Dhokariya[3], Vaishnavi Dhage[4], Mr. A. J. Patankar[5]
[1, 2, 3, 4, 5] Department of Information Technology
[1, 2, 3, 4, 5] D. Y. Patil College of Engineering, Akurdi, Pune

Abstract- *Nowadays the Mobile Technology is being very famous for the User Experience, because it is very easy to access the applications and services from anywhere of your Geo-location. Android, Apple, Windows, Blackberry, etc. are various famous and commonly used Mobile Operating Systems. All the Operating Systems provides plenty of applications and services for users .The Most famous application of iPhone is "SIRI" which helps the end user to communicate end user to mobile with voice and it also responds to the voice commands of the user. Same kind of application is also developed by the Google that is "Google Voice Search" which is used for in Android Phones. But this Application mostly works with Internet Connections. But our Proposed System has capability to work with and without Internet Connectivity. It's named as Personal Assistant with Voice Recognition Intelligence, which takes the user input in form of voice or text and process it and returns the output in various forms like action to be performed or the search result is dictated to the end user. The main motivation of the system is to allow any kind of native user to use smart phones. In addition, this proposed system can change the way of interactions between end user and the mobile devices. The system is being designed in such a way that all the services provided by the mobile devices are accessible by the end user on the user's voice commands.*

Keywords- Acoustic Model, Voice Recognition Intelligence, Recognizer intent, Offline application.

I. INTRODUCTION

The past research, Local search specializes in serving geographically constrained search queries on a structured database of local business listings. Most text-based local search engines expect a user to submit a search query in two text fields: the "Search Term" and the "Location Term" fields. Most voice-enabled local search systems employ a two-turn dialog strategy. In the first turn, the system solicits from the user a Location Term followed by a Search Term. Although the two-field interface has been widely accepted, it has several limitations for voice-enabled mobile search. To overcome this further introduce a voice search system that allows users to specify search requests in a single natural language utterance. The output of ASR is then parsed by a query parser into three fields: Location Term, Search Term, and Filler. And near-term future work will be evaluating Search performance.

Both ASR and Query parser make errors, however these errors are not equally important in terms of their impact on search. After this work, a set of techniques for improving the performance of automated voice search services intended for mobile users accessing these services over a range of portable devices. Voice search is implemented as a two stage search procedure where string candidates generated by an automatic speech recognition (ASR) system are re-scored in order to identify the best matching entry from a potentially very large application specific database. The work in this paper deals specifically with user utterances (saying) that contain spoken letter sequences corresponding to spelled instances of search terms.

As more data becomes available for a given speech recognition task, the natural way to improve recognition accuracy is to train larger models. Then, the development and advances in automatic speech recognition for the AT&T Speak4it R voice search application. With Speak4it as real-life example, we show the effectiveness of acoustic model (AM) and language model (LM) estimation (adaptation and training) on relatively small amounts of application field-data. And introduce algorithmic improvements concerning the use of sentence length in LM, of non-contextual features in AM decision trees, and of the Teager energy in the acoustic front-end.

Language Model: A Katz back off 3-gram LM was built using 10 million anonymous business search queries from yellowpages.com. These are typed, not spoken, queries consisting of two fields: search term (e.g. business name), and location. To approximate the text of spoken queries for LM estimation, these fields were combined in one sentence using them ost frequent carrier phrases in the development data, weighted by relative frequency. Misspellings and inconsistent business name tokenization in the user-typed web queries were corrected by 62k substitution rules. After text normalization, the vocabulary contains 240k words.

Acoustic Model: The "general" telephone acoustic model of the AT&T recognizer is a discriminatively trained tri-phone hidden Markov model (HMM), with states

consisting of Gaussian mixtures, estimated on 2k hours of telephone audio data. The acoustic frames have 60 dimensions, defined by discriminative projections of MFCC vectors.

To better match the AM to the iPhone recordings, we adapted the AM on 120 hours of wireless speech data from a directory assistance (DA) application, by the maximum a posteriori (MAP) method. The MAP prior was set to optimize the iPhone pilot data recognition accuracy. This prior value weights the DA data for half of the total count.

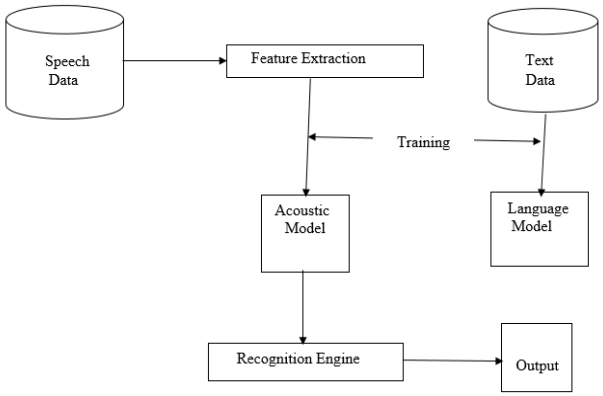

Figure 1. System Architecture for proposed system

II. LITERATURE SURVEY

Speech recognition has a long history with several waves of major innovations. Speech recognition for dictation, search, and voice commands has become a standard feature on smartphones and wearable devices. Design of a compact large vocabulary speech recognition system that can run efficiently on mobile devices, accurately and with low latency. [1] This is achieved by using a CTC-based LSTM acoustic model which predicts context-independent phones and is compressed to a tenth of its original size using a combination of SVD-based compression and quantization. Quantized deep neural networks (DNNs) and on-the-fly language model rescoring to achieve real-time performance on modern smartphones. The ASR and Search components perform speech recognition and search tasks. In addition to ASR and Search, we also integrate a query parsing module between ASR and Search for a number of reasons. [3] Set of techniques for improving the performance of automated voice search services intended for mobile users accessing these services over a range of portable devices. Voice search is implemented as a two stage search procedure where string candidates generated by an automatic speech recognition (ASR) system are re-scored in order to identify the best matching entry from a potentially very large application specific database. Study provides a good example of how additional domain specific knowledge sources can be used with a domain independent ASR system to facilitate voice access to online search indices. As more data becomes available for a given speech recognition task, the natural way to improve recognition accuracy is to train larger acoustic models. There are a non-parametric empirical model that exploits abundant training data to directly learn pronunciation variation.

Interpolating the empirical model with a parametric model yields the best performance, with a relative improvement of 5.2% in WER over the baseline. [2] There are a number of ways in which this work could be extended. First, closer integration Personal Assistant With Voice Recognition Intelligence with acoustic model training is likely to yield sharper distributions and a tighter fit to the data. Second, estimating word-pronunciation co-occurrence counts in semi-supervised fashion (e.g. through word recognition instead of forced alignment) would broaden its applicability to a wide range of speech genres and tasks.

Usually many devices support voice recognizer for helping the end user to communicate with voice commands. But all of them work with internet connectivity even for performing simple operations of Mobile Devices. Basic functions like connecting a call, sending a text message or opening any installed application, etc. doesn't require any internet connectivity.

Although Google has proven that offline voice recognizer works with 7 times faster speed than online voice recognizer.

Google Voice Search and iPhone's SIRI also works with the internet connectivity even for performing basic functions of their respective Mobile Devices. Although jellybean will provide good functionality and stability to user but it's not exploiting the current extent of hardware. The latest version of android is Android 'M' (Marshmallow) which is Android 6.0 which can run very well on the offline voice search engine. Also Apple Company works on such research topic and soon will implement in the latest Apple Phones Releases.

This all voice search applications needs to physically run the application which supports the voice recognition system. Hence this applications are not user friendly for Native Users especially for Blind Persons. Blind persons can use Laptop or any other computer devices by their senses.

Hence we needed to design and developed a basic mobile application which has offline (without Internet Connectivity) voice recognition feature which can also

perform basic operations mentioned above without internet connectivity.

III. PROPOSED WORK

PARI has various branches of the services, but the main feature of PARI is Voice Recognition Engine which has an ability to work without internet connection i.e. Offline Voice Recognition.

This Voice Recognizer works offline and performs various operations as per the user commands and requirements. This is the first activity that's opens whenever we starts PARI and it directly can be opened by pressing Power Button of the mobile devices.

Offline Voice Recognizer-

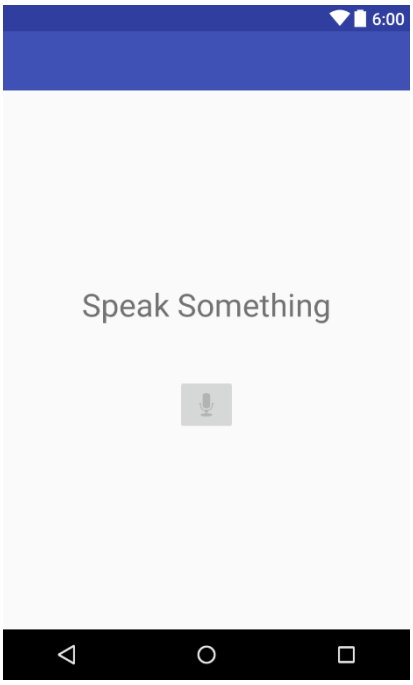

Figure 2. Voice Recognizer Activity

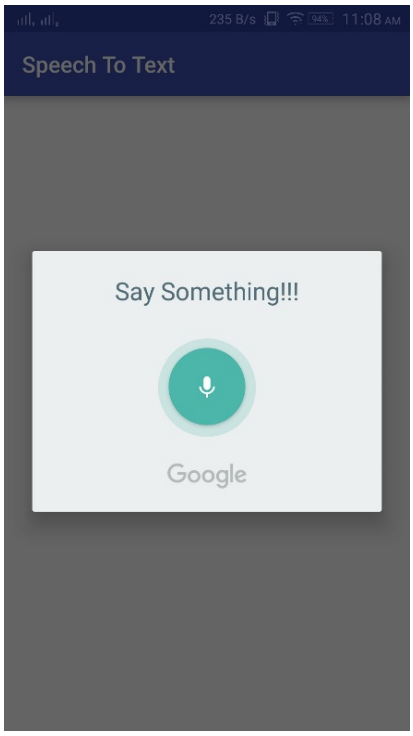

Figure 3. Voice Recognizer Activity

This feature is specially design for Blind Persons who wish to use the Android Smart Phones but are unable to connect this technology. Also Native user who barely knows to unlock their smart phones can easily open this application and using voice commands in their local languages as per need.

PARI responds to basic commands like, Open Applications, Close Applications, Connect Calls To respective person/contacts, send text SMS to respective person or contact, Capture Photos from camera (Front Camera/ Back Camera), Add/Delete/Update Contacts, Run any media file, Start various services like Hotspot, Wi-Fi, Bluetooth, and various Services from the respective Notification Panel. All this can be performed on the voice commands of the end user without internet connectivity.

Operations such as Browsing or Searching for any topic, using Applications that need internet connections example "Send message to ABC, Hi I'm PQR sending message to you. Using WhatsApp". All these basic operations are performed by the voice commands of the end user.

Google did quietly enable offline recognition in that Search update, but there is no API or additional parameters available within the Speech Recognizer Class. The functionality is available with no additional coding, however the user's device will need to be configured correctly for it to begin working and this is where the problem. Also, Google

have restricted certain Jelly Bean devices from using the offline recognition due to hardware constraints. Which devices this applies to is not documented, in fact, nothing is documented, so configuring the capabilities for the user has proved to be a matter of trial and error.

Steps to Start Offline Voice Recognizer in Latest Android Smart Phones-

1. Make sure the default Android Voice Recognizer is set to Google.
2. Uninstall any offline recognition files you already have installed from the Google Voice Search Settings.
3. Go to your Android Application Settings and see if you can uninstall the updates for the Google Search and Google Voice Search applications.
4. If you can't do the above, go to the Play Store see if you have the option there.
5. Reboot (if you achieved 2, 3 or 4)
6. Update Google Search and Google Voice Search from the Play Store (if you achieved 3 or 4 or if an update is available anyway).
7. Reboot (if you achieved 6)
8. Install English UK offline language files.
9. Reboot.
10. Use Utter with a connection.
11. Switch to aero plane mode and give it a try.
12. Once it is working, the offline recognition of other languages, such as English US should start working too.

Learning Capability-

Language is not a Barrier for PARI to understand the user voice commands for performing respective operations. Hence PARI uses its Intelligence to store the voice commands detected by the end user with respect to the Default Commands stored in PARI.

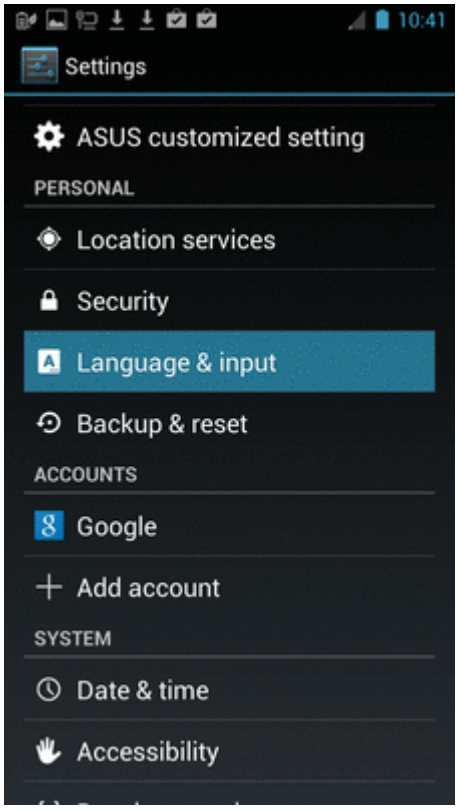

Figure 4. Selecting Various Languages

Android provides Built-in applications for phone calls, in some occasions we may need to make a phone call through our application. This could easily be done by using implicit Intent with appropriate actions. Also, we can use Phone State Listener and Telephony Manager classes, in order to monitor the changes in some telephony states on the device.

Now you can enter a desired mobile number and a text message to be sent on that number. Finally click on Send SMS button to send your SMS. Make sure your GSM/CDMA connection is working fine to deliver your SMS to its recipient.

You can take a number of SMS separated by comma and then inside your program you will have to parse them into an array string and finally you can use a loop to send message to all the given numbers. That's how you can write your own SMS client. Next section will show you how to use existing SMS client to send SMS.

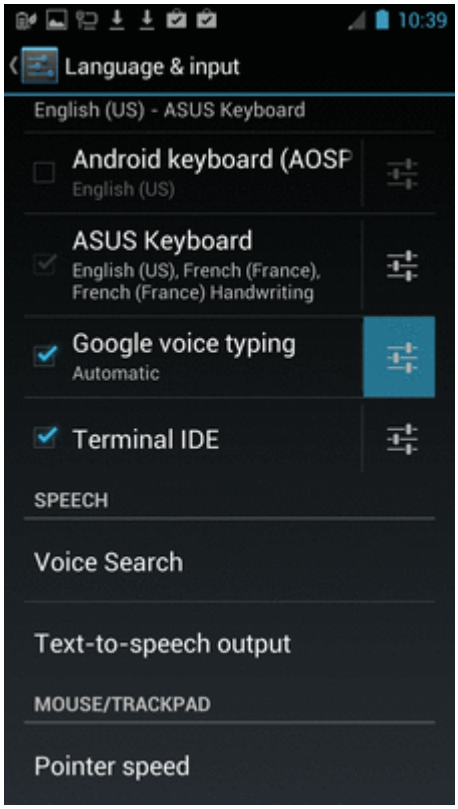

Figure 5. Selecting Various Languages

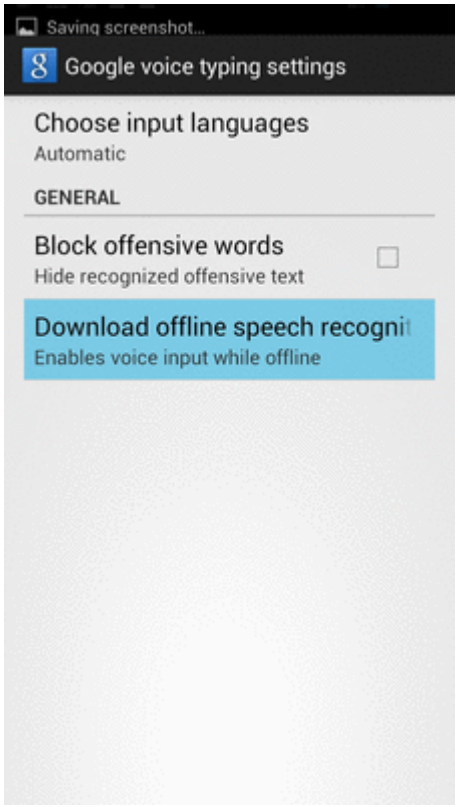

Figure 6. Selecting Various Languages

Figure 7. Selecting Various Languages

Extra Features:

1. Sync up your storage your computer has local storage, as does your phone -- two separate, unconnected virtual vats of space. But with a clever cloud-embracing app and a few minutes of configuration, the devices' drives can act as if they're one.

The Android app FolderSync allows you to sync up storage between your desktop PC and Android smartphone. The secret resides in FolderSync, an Android utility that costs a mere (There's a free version, but it's peppered with ads and offers limited functionality.) FolderSync works with a ton of cloud storage providers, including Amazon, Box, Dropbox,

Google Drive, and OneDrive. As long as the provider you select offers a companion program on the PC side (all of the ones I mentioned do), you'll be good to go. To get everything up and running, first install the app on your phone and follow the prompts to connect it to the cloud storage service of your choice. Set up "pairs" for any folders you want to keep linked with your computer -- a folder containing your documents or downloads, for instance -- and create new folders in your cloud storage to match.

Be sure to set the pair to use two-way sync. You can either opt to sync instantly, if you want everything to be kept up-to-date at all times, or you can go for a more battery-friendly setup like syncing once a day if you don't mind a little bit of latency.

Now install the desktop app for whatever cloud service you're using. Open the app, find the paired folder you created, and get it ready for use. If you paired a Documents folder, for example, you might place a shortcut on your desktop and dump all of your existing documents into it.

Then treat that folder as if it were a regular local resource. Anything you do in it will be saved onto your hard drive and synced into the cloud, where it'll automatically find its way onto your Android device. Any changes on your phone will make their way back to your computer in the same manner.

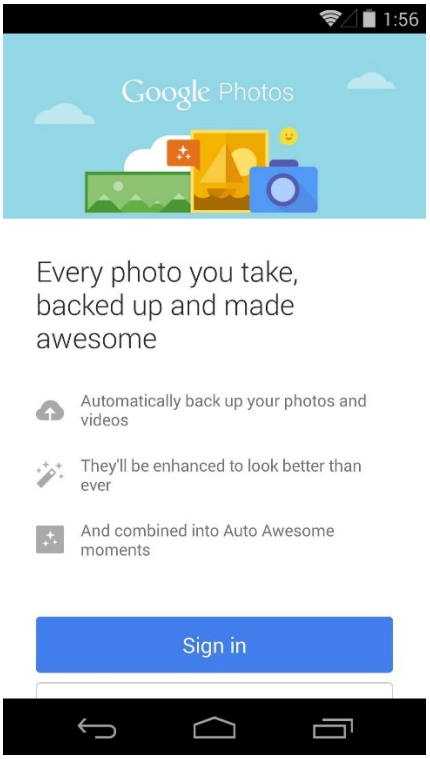

Figure 8. Sync up storage

2. Find and secure your phone can't find your phone in its usual place between couch cushions? Or, worse, get home from a day of travel only to realize you lost your device somewhere along the way? No need to panic: Your computer can tell you exactly where your mobile buddy is. It can even remotely lock it down and erase it if need be.

All you have to do is get on your computer (or any computer, really), navigate togoogle.com in the browser, and type "Find my phone" into the search box. (You'll have to sign into Google first -- which you'll probably want to do in an incognito window if it isn't your own PC.) Within seconds, Google will give you a detailed map showing your smartphone's last logged location. Clicking on it will bring up the full Android Device Manager interface, where you'll find options to ring lock, or fully erase your phone right then and there.

IV. FUTURE SCOPE

PARI available on all Android devices on version 4.0 and above. To make platform Independent and make Available in various mobile Operating Systems. It include Various Language Narrator's Voice i.e. in different languages narrator's voice is available for dictation. Also make use of this application in Networking Devices.

V. CONCLUSION

PARI is designed to help Native and especially for blind persons which works on their voice Commands. PARI also has the capability of recognizing the voice commands without internet connection.

VI. ACKNOWLEDGMENT

We take this opportunity to thank Mr. A. J. Patankar, the Head of the Department (Information Technology) who was also our project guide for his valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are thankful to all the staff members of the Department of Information Technology of D. Y. Patil College of Engineering, Akurdi for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the Institute for providing the required facilities, internet access and important books.

REFERENCES

[1] "Accurate and compact large vocabulary speech recognition on mobile devices," in INTERSPEECH. 2013, pp. 662–665, ISCA.

[2] "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in ICML, 2006, pp. 369–376.

[3] "Lattice-based optimization of sequence classification

criteria for neural-network acoustic modeling".

[4] "Language modeling for What-with Where on GOOG-411," Proc. of INTERSPEECH, 2009.

[5] "DFirect construction of compact context-dependency transducers from data," Proc. of INTERSPEECH, pp. 218–221, 2010.

[6] "Automatic question generation for decision tree based state tying," Proc. of ICASSP, 1998.

[7] "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010.