

Design and Development of a Personal Scheduler Based on Continuous Speech for Smart Phones

R. Vishnupriya^{#1*}, T. Devi^{#2}

Department of Computer Applications,
Bharathiar University, Coimbatore, Tamilnadu, India
1 vispriyaraju@gmail.com
2 tdevi5@gmail.com

Abstract - A scheduler is one of the most essential applications for all people like various business and IT professionals. Usability of smartphones is exponentially on the rise as it replaces multiple and most essential gadgets that make the life of a human person much easier. One of such important tool is to provide hands free access like speech input with smart phone is much on demand. Speech recognition is one best possible way that converts speech to text which is very convenient mechanism during hectic schedules. Hence, implementing speech recognition that schedules activities of a person through smart phone is a great boon for busy people using smart phones. This research aims at designing and developing of a personal mobile scheduler application based on "Continuous Speech" for smart phones.

Keywords - Speech Recognition, Mobile Application, Personal Scheduler.

I. INTRODUCTION

Personal Scheduler based on Continuous Speech is a mobile application, developed on android platform for scheduling daily appointments. Most of the people face difficulties in remembering their appointments and in the course of time lose their important commitments. To overcome these difficulties diaries, memo pads, computer based scheduling programs, organizers and Personal Digital Assistants (PDAs) are used to note the appointments. All these devices need handwriting or a keyboard/pointing device as the primary mode of input, which is not always efficient, in particular when the user does not have time to note down anything. Smart Phone schedulers usually use handwriting recognition (Graffiti) or a keyboard based interface for input. While a keyboard based interface is good to use, it could become a difficult task when the user is left with very little time. Similarly, Graffiti can be equally difficult to use in such similar situations. Speech is the easiest, convenient and most common way for people to communicate. Speech is also the fastest way to communicate than typing on a keypad and more expressive than clicking on a menu item. For these reasons, speech applications are gaining importance, especially for users with low literacy or little lettering knowledge, such as those in the developing regions [11].

Devices which support speech are ubiquitous in the market. Some of these devices include wireless phones with support for voice enabled dialing and voice commands, PDAs with voice driven command interfaces, etc [2]. But, these devices are not effective and need to be programmed or trained before they can be used efficiently. Most of them work by relating a recorded voice command to the particular system command. On the other hand, mobile internet devices, internet tablets, smart phones and cell phones have widely proliferated the market [4]. Thus mobile phones with voice recorder will be a better choice for this kind of application. Current research focuses on using speech as a mode of input which is very convenient and useful for a quick note of things and eliminates the need to carry a PDA in addition to a phone.

Apple developed an intelligent personal assistant and knowledge navigator which is called "SIRI" that works as an application for iOS platform. SIRI allows one to send messages, schedule meetings, find places, browse information and make phone calls without requirement of any training [1]. Google Now is an intelligent personal assistant which is available for Android mobile phones within the Google Search application. It allows the users to send messages, schedule meetings, find places, browse information and make phone calls. Popular Science an American magazine presented an award of "Innovation of the Year 2012" for Google now tool [9,5]. SIRI has limited accuracy for normal conversation language and even though Google Now has high accuracy it does not take the sentence to scheduler instead it takes the given speech to the search engine since it is present inside Google search. The proposed scheduler anticipated a scheduling aid through smart-phones using speech as an input which can recognize all type of phrases which are used frequently for scheduling tasks.

II. SPEECH RECOGNITION

Speech Recognition or Automatic Speech Recognition (ASR) or computer speech recognition is the process of converting speech signals to a sequence of words, by means of an algorithm implemented as a computer program [3].

A. Architecture of Speech Recognition

The architecture of typical speech recognition system is shown in Figure 1. Using frequency response analysis, the input speech signal is processed to extract feature values which are input to speech recognition engine. The input feature values are compared by the speech recognition engine with acoustic and language models which are trained using previously accumulated data, determining a list of the most likely morphemes as output. The acoustic model expresses the correspondence between speech feature values and phonemes while the language model expresses the likelihood that a morpheme would precede or follow a given morpheme.

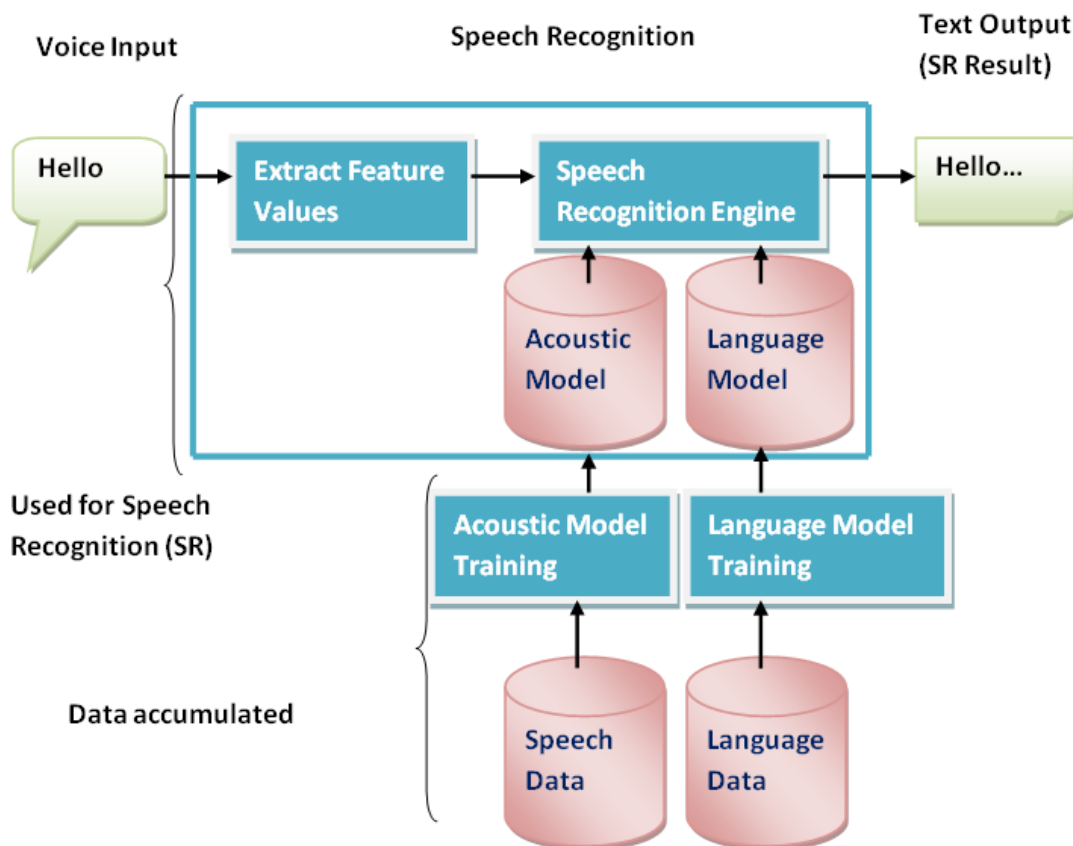


Fig. 1 Architecture of speech recognition [8]

The accuracy of speech recognition depends on the condition of how the acoustic and language models are trained close to the actual input environment [7]. It is necessary to reflect the features of the actual user when training the acoustic model and also it is important to include a large vocabulary in order to recognize a wide range of utterances when training the language model. Thus, training with a large text data set is required to build a language model.

III. Dataflow of CSBPS

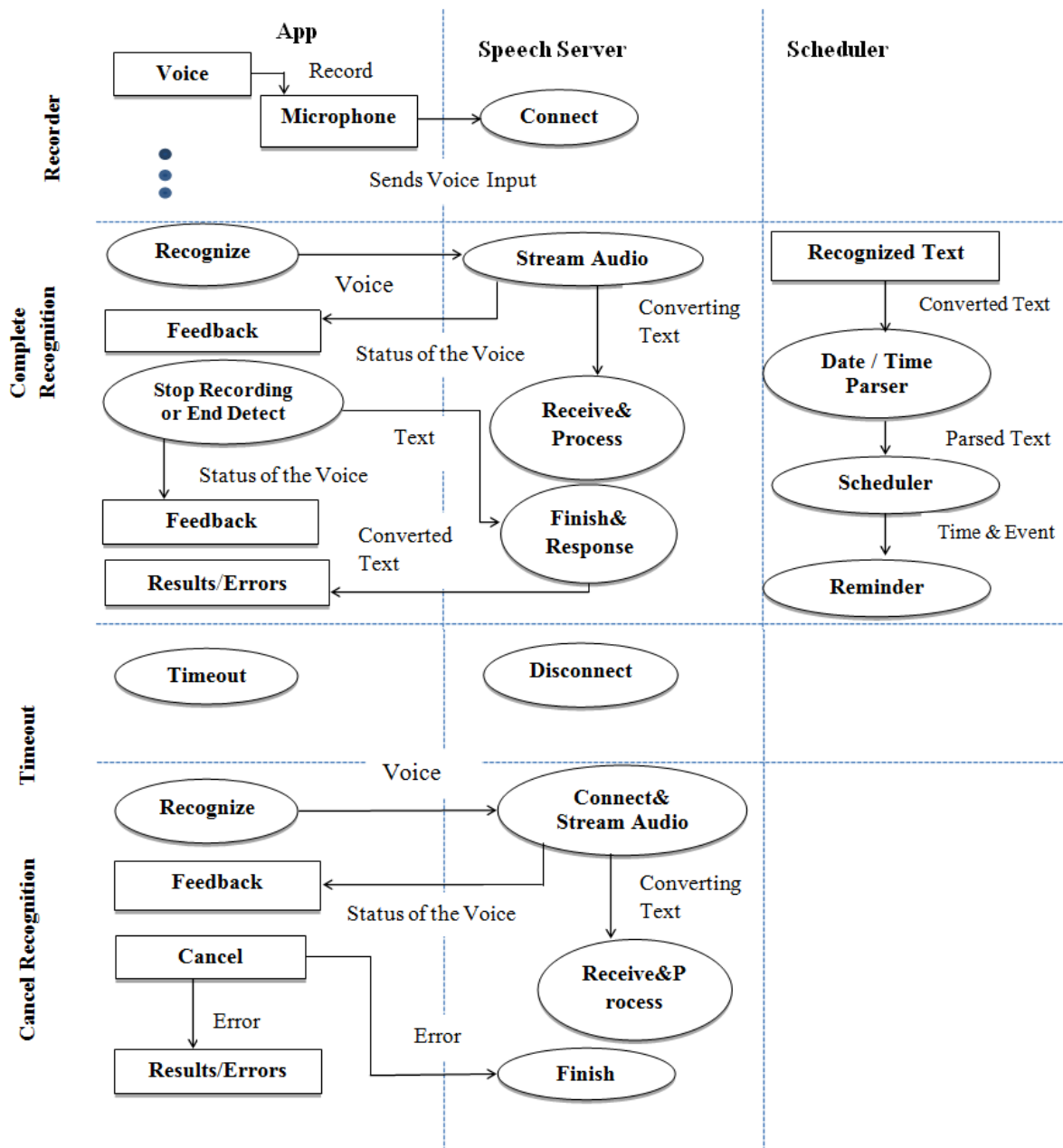


Fig. 2 Dataflow of Personal Scheduler based on Continuous Speech

Figure 2 depicts the dataflow diagram of Personal Scheduler based on Continuous Speech for smart phones. Users input in the form of voice will be recorded by a microphone which is then connected to a speech server. Streamed speech data will be given to the speech server which converts it to phonetics and then to text. The scheduler automatically detects the ending of the section and user can also stop the recording. If the scheduler detects an ending then the speech server will return the recognized sentence to the scheduler. During recognition period the server can detect many types of errors such as audio recording error, client side errors, insufficient permissions, recognition result matched, recognition service busy, server sends error status and no speech input. If the scheduler fails to detect the ending of the sentence or the user didn't stop the recorder, then scheduler will show timeout error message to the user after 5 minutes. If the user cancels the recording in the middle, partial voice data will be given to the server and later the result will be deleted without any response to the user. The following process will take place while the correct recognized word is given to the scheduler ① The recognized sentence will be parsed as date and event, ② The event will be scheduled in the scheduler and ③ The scheduled events will be reminded to the user on every day morning 6'o clock and one hour before the appointment.

IV. RESULTS

An experiment has been conducted on human speech in proposed scheduler as shown in Fig 3. The experiment is conducted with continuous speech that is converted to text which is given to the scheduler returning a text reminder to the user. The proposed scheduler uses normal human conversational language for scheduling day to day appointments and reminds about the appointment in the morning of the appointment date and one hour before the appointment time. Fig 3 shows the output in which toggle button is used to start and stop the speech recordings from the user. The button shows the status of the recording that is whether it is in recording mode or not. If the user presses the toggle button then the recording will start and the speech will be converted to text.

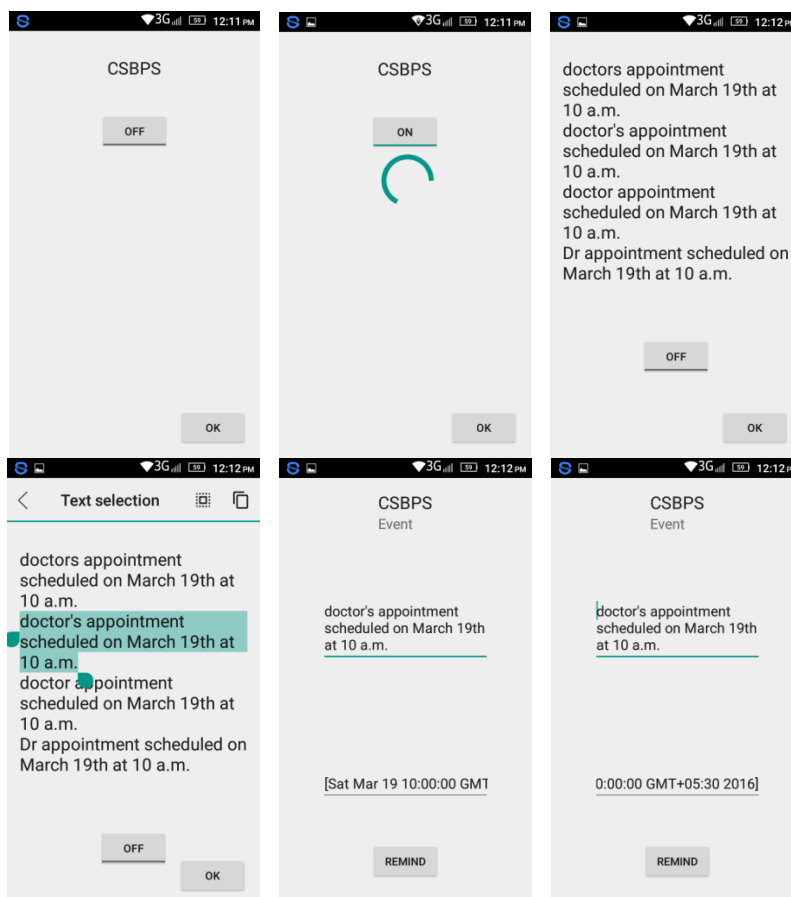


Fig. 3 Personal Scheduler based on Continuous Speech

For example, if the user speaks a sentence, “*Doctor’s appointment is on March 19th 2016 at 10 am*”, CSBPS converts the speech into text and the text will be parsed into strings, date and time. The output of the CSBPS will be *String: Doctor’s appointment, Date: March 19th 2016 and Time: 10 am*. The parsed string, date and time will be given to the scheduler in which the string will be set as appointment, the date and time will be set as schedule date and time respectively. If there is any mispronunciation then the application will display all the possible sentences and makes the user to select one from the possibilities. After selecting the correct sentence, user can click the ‘OK’ button to see the parsed sentence. Then the scheduled appointment will be reminded to the user in the morning of the appointment date and one hour before the appointment time.

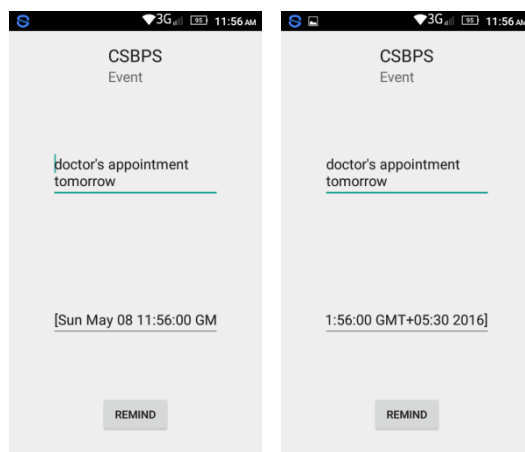


Fig. 4 Smart Capability - Recognizing Tomorrow's Date

In addition to recognizing the speech, the proposed scheduler has some smart capability to recognize the dates. Fig 4 shows the output of recognizing tomorrow's date for the sentence "Doctor's appointment tomorrow". The speech is recorded on May 7th at 11.56 am and the date and time recognized by CSBPS is May 08 2016 Sunday and 11:56:00 GMT+5.30 respectively.

Fig 5 shows the output of recognizing upcoming Friday's date for the sentence "Doctor's appointment on Friday". The speech is recorded on May 7th at 11 hours 58 minutes and 23 seconds in the morning and the date and time recognized by the proposed scheduler is May 13 2016 Friday and 11:58:23 GMT+5.30 respectively.

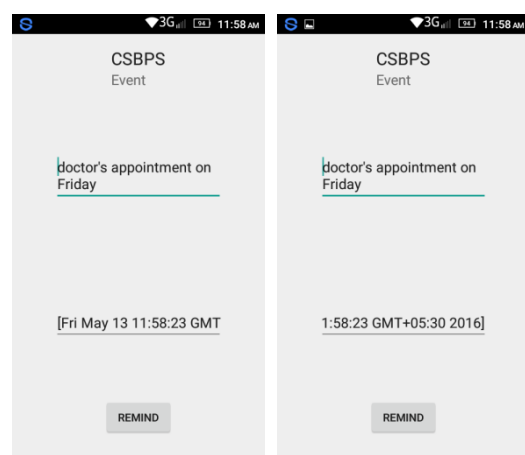


Fig. 5 Smart Capability - Recognizing Upcoming Friday's Date

The scheduler is tested with 100 sentences with 1066 words in which 1024 words have been recognized yielding a result of 96% accuracy. Among the 100 sentences, 80 sentences are related to appointment while 20 are not related to appointment, hence the performances of the scheduler in deducting sentences which are not related to appointment are also analyzed with confusion matrix. Fig 6 shows the percentage of the recognized and non-recognized words.

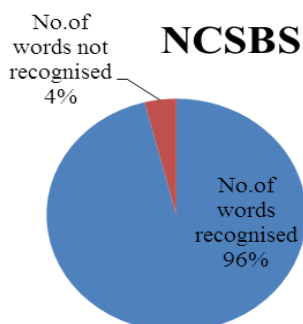


Fig. 6 Percentage of the Spoken Words Prediction

V. DISCUSSION

Table 1. Recognition Percentage of Personal Scheduler based on Continues Speech and Comparison of Recognition Result with SIRI

Test Set	Number of Input Sentences	Number of Sentences Recognised Correctly	Number of Sentences Not Recognised Correctly	Recognition Percentage of proposed scheduler	Recognition Percentage of SIRI
1	10	07	3	97.00	78.000
2	20	17	3	97.95	71.300
3	40	31	9	97.30	70.775

Table 1 summarizes the comparisons between SIRI and the proposed scheduler for normal human conversational language. The developed scheduler has been tested with three input datasets consisting of 10, 20 and 30 sentences respectively. SIRI yields 78, 71.3 and 70.775 percentage of recognition for 10, 20 and 40 sentences dataset respectively. The proposed scheduler yields 97, 97.95 and 97.3 percentage of recognition for 10, 20 and 40 sentences dataset respectively. Thus the proposed scheduler shows improvement of 19, 26.65 and 26.525 percentage of recognition than SIRI for the three datasets.

The confusion matrix determines the performance of the proposed scheduler. This matrix describes all possible outcomes of a prediction results in table structure. The possible outcomes of a two class predictions can be represented as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The sentences related to appointment and not related to appointment are correctly classified as True Positive and True Negative respectively. True Positives (TP) are predicting that the given sentences as appointments which are actually related to appointments. True negatives (TN) are predicting that the sentences as not related to appointment which are actually not related to appointment. False positives (FP) are predicting that the sentences as related to appointment, but the sentences are actually not related to appointment. False Negatives (FN) are predicting that the sentences are not related to appointment, but the sentences are actually related to appointment.

Table 2. Confusion Matrix

N = 100		Predicted	
		Negative	Positive
Actual	Negative	a (TN = 14)	b (FP = 6)
	Positive	c (FN = 0)	d (TP = 80)

In the confusion matrix given in Table 2, 'a' is the number of correct predictions that an instance is negative, 'b' is the number of incorrect predictions that an instance is positive, 'c' is the number of incorrect predictions that an instance negative, 'd' is the number of correct predictions that an instance is positive [6].

The recall and precision can be calculated using equations shown below:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

By applying the formulae to the developed scheduler, the recall and precision are 100% and 93.02% respectively.

Sensitivity and specificity are statistical measures of the performance of a binary classification called classification function. Sensitivity measures the proportion of actual positives and specificity measures the proportion of negatives. The sensitivity and specificity are is calculated as shown below [10]:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

By applying the above formulae to the developed scheduler, the sensitivity and specificity are computed as 100% and 70% respectively.

VI. CONCLUSION

This research work is concerned with design and development of a personal mobile scheduler application based on "Continuous Speech" wherein inputs are based on normal human conversational language for scheduling meetings, reminds user about the scheduled events and to improve the effectiveness of Speech Recognition (SR). A comparative study of the existing tools reveals that SIRI has limited accuracy for normal conversational language. Even though Google Now has high accuracy it does not take the sentence to scheduler instead it takes the given speech to the search engine since it is embedded inside Google search. The proposed

scheduler is developed using normal human speech and specially for scheduling the appointments and reminds about the appointment to the user in the morning of the appointment date and one hour before the appointment time. The concepts developed are tested through a mobile application yielding good performance which is developed using android and eclipse platform.

REFERENCE

- [1] Ankit Asthana, RGS Asthana, 2012, "Android 4.0 and Windows 8 – A Review", Beacon IEEE, Vol no. 31.
- [2] Anuj Kumar, Anuj Tewari, Seth Horrigan, Matthew Kam, Florian Metze and John Canny, 2011, "Rethinking Speech Recognition on Mobile Devices", Proceedings of Intelligent User Interfaces for Developing Regions, Stanford University, Palo Alto, California, USA.
- [3] Anusuya M.A, Katti S.K., 2009, "Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security, Vol. 6, No. 3.
- [4] Devi Thirupathi, Ashwin K. Whitchruch, Vishnupriya. R, 2013, "Speech Driven Personal Scheduling System", International Journal of Advanced Research, Vol. 1, No. 7.
- [5] Google Now wins 'Innovation of the Year', runs laps around 'outdated' Siri at <http://bgr.com/2012/11/15/google-now-wins-popular-science-award/>, 14 October 2015.
- [6] Kohavi and Provost, Confusion Matrix at [http://faculty.smu.edu/tfomby/eco5385/lecture/Confusion%20 Matrix.pdf](http://faculty.smu.edu/tfomby/eco5385/lecture/Confusion%20Matrix.pdf), 27 July 2015.
- [7] Rupal Patel, 1998, "Control of Prosodic Parameters in Dysarthria", Acoustic Analyses of Dysarthric Speech.
- [8] Shinya Iizuka et al., "Speech recognition and technology and applications for improving terminal functionality and service utility", 2012, NTT DOCOMO Technical Journal, Vol.13, No.4.
- [9] Vishnupriya R., Devi T., 2014, "Speech Recognition Tools for Mobile Phone – A Comparative Study", International Conference on Intelligent Computing Applications (ICICA), archived in IEEE Digital xplora, pp. 426 – 430.
- [10] Wikipedia, Sensitivity and specificity at http://en.wikipedia.org/wiki/Sensitivity_and_specificity, March 7, 2016
- [11] Zheng-Hua Tan and Borge Lindberg, 2010, "Speech Recognition on Mobile Devices", Publisher: Springer Berlin Heidelberg, ISBN: 978-3-642-12348-1.

Author Profile



Mrs. R. Vishnupriya received her Bachelor of Information Technology from P.S.G.R Krishnammal College for Women, Coimbatore in 2010 and Master of Computer Science in P.S.G College of Arts and Science, Coimbatore. She is currently doing Ph.D. Computer Science in Department of Computer Applications, Bharathiar University, Coimbatore. Her area of interest is speech recognition. She published some papers in International Journals.



Dr. T. Devi received Master of Computer Applications from P.S.G. College of Technology, Coimbatore in 1987 and Ph.D. from the University of Warwick, United Kingdom in 1998. She is presently heading the Department of Computer Applications, Bharathiar University, Coimbatore. Prior to joining Bharathiar University, she was an Associate Professor in Indian Institute of Foreign Trade, New Delhi. Her current research includes Software Engineering, Product Introduction, Technical Process Management and Concurrent Engineering. She has contributed more than 140 papers in various National / International conference / journals.