Indian Journal of Science and Technology, *Vol 9(37), DOI: 10.17485/ijst/2016/v9i37/102102, October 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Speech Recognition based Adaptive Examination Application for Visually Impaired Students

Viswanath Subramanian Ramesh*, P. Saranya and Ruchira Kapoor

SRM University, Kattankulathur, Kanchipuram - 603 203, Tamil Nadu, India; saaiviswanath@yahoo.com, saranya.p@ktr.srmuniv.ac.in, ruchirakapoor15@gmail.in

Abstract

Objectives: To develop an application in order to make the examination process easier for the visually impaired and also to reduce the conventional examination setup overheads. **Methods:** The conversion of text to speech and vice versa has been carried out by existing modules in the android platform. Adaptive examination is carried out based on performance analysis of the candidate by assessing the cognitive ability (reaction time) which is implemented through the concept of time stamping. **Findings:** The equation formulated for making the examination adaptive, returned authentic and error proof values which could be used for setting the difficulty of the consequent sections of the question paper, by comparison with predetermined ideal performance value. Also it was evident that the setup process was relatively simpler than a pen-paper examination setup. The speech rate of the audio module had to be personalized according to the candidate's requirement. By incorporating the changes and handling the runtime glitches, it was found that application made the examination process hassle-free for visually impaired students. **Improvements/Applications:** Objective analysis of cognitive ability has been made possible through the newly formulated equation. The result obtained can be applied in multiple domains.

Keywords: Android, Automatic Speech Recognition (ASR), Adaptive Examination, HMM, Speech Recognition, Time Stamping

1. Introduction

Adaptive examinations are one of the most popular types of examinations these days. This type of examination contains section of questions that have to be answered by a student. Based on the performance in one section, the difficulty level of the next section is determined. For example, if the student answers the number of questions correctly above the threshold value, the next section will contain tougher questions with higher weightage of marks. Similarly, if the student answers the number of questions correctly below the threshold value, the next set of questions will be easier and carry lesser weightage of marks. Based on this the scoring of a student can be done.

The basic problem faced by visually impaired students is that these types of adaptive examinations require to either be converted into braille or a scribe is needed. And this means that they are limited to non-computerized examinations. The introduction of speech recognition based application enables a visually impaired student to enjoy the benefits of a computerized exam.

The algorithm attempted at attaining the final score of a student includes analysis of cognitive ability of the student with help of their reaction time. This method includes timestamping of each activity. Timestamping helps keep tract of the time being taken by each student and deduce whether the time taken is more than or less than the average time taken by a student. With the help of the number of correct questions and the time taken we can determine whether a student has performed above average or below average in a section. Based on this performance the adaptive nature of an examination can be explored.

**Author for correspondence*

2. Automatic Speech Recognition

Speech Recognition is also referred to as Automatic Speech Recognition (ASR)[1,2] or simply Speech-to-text (STT). ASR Figure 1 is the ability of a computer to convert the words spoken by a person into text. In this process first a person has to speak some words into the microphone which converts the acoustical energy into electrical energy. Once the sound waves are mapped as an audio signal, they are then converted into digital signals for the computer to understand them. With the help of the acoustic model, language model and lexicon the audio signal is converted into a word string which has been trained.

Speech Recognition requires continuous training of data. The data can be trained using the most used model, Hidden Markov Model (HMM). The language can also be selected for instance English, French, German etc. With help of the language model being used, the lexicon model and the acoustic model, one can decode the audio signal into a word string.

2.1 Hidden Markov Model

Hidden Markov Model Figure 2 is the most commonly used model for speech recognition. In this model, the output is a series of syllables or symbols. The system being modelled is considered to be a Markov Process with hidden states. HMM[3,4] finds its application in the field of cryptanalysis, machine translation, gene prediction, speech recognition to name a few. HMM determines the confidence probability of possible output of each stage and selects the one with higher confidence probability.

How HMM works in speech recognition is as follows:

- The user gives an input in the form of a sound wave. This can consist of a letter, word or even a sentence.
- Considering that a word has been taken as an input, it is split into syllables.
- These syllables are then analyzed with the help of the trained data sets.
- Each possible outcome is given a confidence measure. The one with higher confidence measure is selected.
- After this procedure is performed on the entire word, we get the final product with highest confidence measure.

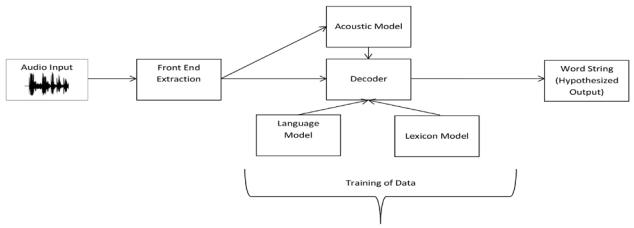

Figure1. Block diagram for an Automatic Speech Recognition System.

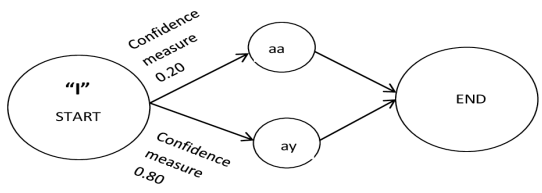

Figure2. Hidden Markov Model-Speech Recognition.

3. Integrating Speech Recognition with Android

The software environment used in most mobile devices is android[5] and it comprises of the operating system, key applications and middle ware. With the help of certain modules such as TTS (Text-to-speech) and STT (Speech-to-text) API we can integrate speech recognition in an android application. This feature extends the use of an android application for visually impaired people.

Firstly, a data set has to be trained which comprises of all the possible speech inputs. Once we have the trained data, with the help of a microphone we can take the sound waves as an input. With the help of all the above mentioned processes and models the input is compared with the trained data set and the output is the word string. It is obviously easier to interpret words in comparison to entire phrases or paragraphs. Hence, developing an adaptive examination portal with multiple choice questions is easier to implement as the input are just single words.

3.1 Text-to-Speech

TTS[6] is a Google API that is used in making of the application. With the help of this the textual data can be converted into sound waves. This is needed for reading out the instructions and questions to the student. There are many existing languages that can be converted from text to speech such as English, French, German, Spanish etc. Nowadays, even various accents are available along with a choice of a female voice or a male voice. According

to where the application is being used, the language and accent can be chosen.

3.2 Speech-to-Text

STT is another name for automatic speech recognition. This API is used for the more difficult part of converting the audio word into a textual string. Large sets of data need to be trained so that the system can recognize the words being spoken. The string returned by the API can be used in conditional statements to perform various actions. Pronunciation of various words can be different with different people even though the language is the same. The data set has to be trained accordingly. For example, the letter 'a' can be pronounced as 'aye', 'hey' and so on.

3.3 Analysis of Cognitive Ability

The adaptive nature of examination can be achieved by Knowledge Reaction curve. We need to consider two things while assessment of a section 1) the number of correct answers by a student in that section, and 2) the time taken to answer those questions. The performance measure is determined individually for each student.

We consider 'n' as the number of questions. The variable 'answer' can have two values of 0 and 1. 0 meaning the answer is incorrect and 1 meaning it is correct. Reaction time is the time elapsed between the question being read out and the student answering the question. The *ideal reaction time* will be the time taken by the teacher who has set the examination questions to answer the question since they will know the answer before-hand.

$$\text{Total\_mark} = \sum \text{mark}(i) \quad \text{(i)}$$

The equation shows that the total mark of the section is a summation of all the individual marks of each question.

$$\text{Total\_RT} = \sum \text{Reaction\_Time}(i) \quad \text{(ii)}$$

Similarly, the total reaction time is a summation of all the individual reaction time. Let, θ be the *threshold value.* Then the equation of θ is as follows:

$$\theta = \text{Total\_mark} / (\text{Total\_RT} * n) \quad \text{(iii)}$$

The equation for the *ideal threshold value* will be calculated as follows:

$$\theta_{ideal} \propto 1/RT_{ideal} \quad \text{(iv)}$$

If the percentage difference between the threshold value and the ideal threshold value is lesser than 50 percent then the student's performance is considered to be good, therefore the next set of questions will be harder but with more weightage assigned to each question. Similarly, if the percentage difference is greater than 50 percent, it signifies a poor performance, and hence resulting in an easier set of questions but with lesser weightage to each question. The difference in weightage ensures fair standards for the test takers. Cognitive ability analysis is done mainly by regression analysis[7]. Linear regression modelling is often used for the analysis, such as

$$x_{i=} \alpha_0 + \alpha_1 a_i + \alpha_2 b_i + \alpha_3 c_i + \varepsilon_i \quad \text{(v)}$$

4. Future Enhancement

The algorithm proposed for analysis of cognitive ability can be altered to help in analyzing the memory power of differently abled children. With the help of the algorithm we can make the children take a test regularly and monitor their response time and growth accordingly. The knowledge reaction curve can help determine a child's quickness, in answering a question from memory. The same data can be used to train the child's memory power.

As of now the proposed model works for single word inputs. In future a system can be made for descriptive examinations with phrases and paragraphs as inputs. This can completely revolutionize the education sector we have nowadays. For this to be accomplished we need much more sophisticated tools[8,9] for speech recognition.

5. Conclusion

In conclusion, with the help of Text-to-speech API, Speech-to-text API and the algorithm for analysis of cognitive ability we can develop a platform for visually disabled students to take examinations without any hassle. The data has to be trained for the various inputs taken from the user so that it can recognize every input taken and save it in a variable accordingly. After the answer is saved the scoring is done as per the adaptive nature of the test. This application can help increase the efficiency of the existing examination process of visually impaired students by a great extent. Further efficiency can be increased by developing this model for descriptive examinations.

6. References

1. Preeti S, Parneet K. Automatic Speech Recognition: A Review. International Journal of Engineering Trends and Technology. 2013; 4(2):132–6.

2. Ghai W, Singh N. Literature review on automatic speech recognition. International Journal of Computer Applications. 2012 Mar; 41(2):42–50.
3. Renita R, Rosita R, Mani S. A Novel speech recognition system using hidden markov model. Indian Journal of Science and Technology. 2015 Nov; 8(32):1–5.
4. Kim D, Kim B. Speech recognition using hidden markov models in embedded platform. Indian Journal of Science and Technology. 2015 Dec; 8(34):1–4.
5. Raghavendhar R, Mahender E. Speech to text conversion using android platform. International Journal of Engineering Research and Applications (IJERA). 2013 Jan-Feb; 3(1):253–8.
6. Rupali DD, Ingole PV. Text detection and recognition with speech output for visually challenged person: a review. International Journal of Engineering Research and Applications (IJERA). 2015 Mar; 5(3):84–7.
7. Brian J, Steuerle Lynnev S, Lowell TJ. The use of cognitive ability measures as explanatory variables in regression analysis. IZA Journal of Labor Economics. 2012 Oct; 1(4):1–19.
8. Atiqua Nuzhat N, Safaet H. Speech to text conversion in real-time. International Journal of Innovation and Scientific Research. 2015 Aug; 17(2):271–7.
9. Suma S, Ramakrishnan KV. An efficient speech recognition system. Computer Science and Engineering: An International Journal (CSEIJ). 2013 Aug; 3(4):21–7.