ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Designing a Teaching Assistant System with the help of Speech Recognition

Shelly Gupta, Utkarsh Telang, Tania Dawra, Ketan Tewari, Prof. R.G. Masand

Dept. of Computer Engineering, Vishwakarma Institution of Information Technology, Pune, Maharashtra, India

ABSTRACT: In today's world, there exists a basic need to simplify human machine interaction as well to make the most of technologies to not only enhance the way machines learn but also how we learn. In classrooms, delivery of lectures with sound as a medium has become a clichéd process which has its own pros and cons. Due to certain constraints (like noise and dialect) students find it difficult to grasp the concepts which are being taught. Thus comes into picture the need for designing a teaching assistant system which takes voice as an input . Speech recognition is the ability of a machine or software to identify words and phrases in spoken language and convert them to a machine-readable format .This paper provides an insight into existing algorithms and methods used for Speech Recognition. Our focus is on developing a framework using Hidden Markov model to obtain readable notes (documents) from speech. These documents will be broadcast to student's personal devices over a Wireless Network. Thus students don't have to worry if they miss any important point during the lecture because it will be directly broadcast to the students present in the lecture in the form of text file.

KEYWORDS: Speech Recognition (SR), Hidden Markov Model (HMM), Broadcasting.

I. INTRODUCTION

Speech recognition (SR) allows user to interact with various devices using Natural Language. Speech provides an easy, fast way of manipulation of text, an alternative to traditional (keyboard entry) and newer (biometric, visual) methods for text input. Thus, there exists a dire need for speech recognition systems.

Lot of research has been going in this area and as a result of which many commercial applications have hit the market in the recent years. These include Dragon NaturallySpeaking and IBM ViaVoice. They promise to provide high accuracy but it cannot be ascertained. Also efforts by the Liberated Learning Consortium (LLC) have established that commercially available software is unfit for real time transcription of data.

Our system aims to convert the instructor's lecture into text format. It differentiates itself from traditional SR softwares by providing delivery of notes in the form of text documents to student's personal device. The purpose of making this system is twofold. Firstly, it maximizes students learning and understanding as the lecture notes are readily available in text form so that students don't miss any important point. They can refer the lecture notes along with lecture slides and audio lecture to recreate the lecture and learn at their own pace. These notes also help people with disabilities to learn. The second important factor is that teachers can use these notes to evaluate their teaching methodology.

Apart from serving these purposes, speech recognition modules from our system can also serve as part of other integrated applications which include voice user interfaces such as voice dialing, call routing, domestic appliance control, preparation of structured documents, media interviews, judiciary documentation. They can be used for commercial as well as personal purposes.

II. RELATED WORK

II.A.1 HTK:

HTK is the most advanced and widely used system for modeling non-stationary data using HMM models. It is free for educational or academicals purposes and can be downloaded after a registration. System was specially designed to cope with the task of automatic speech recognition; however it can be applied to other areas as well. Its current version is 3.3, but there is already an alpha version of 3.4. The main advantages of the HTK are: it is a complex system that covers all development phases of a recognition system, system is regularly updated to catch up with the latest

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

advances in the field of recognition, and it is well documented both theoretically and practically. In the data (transcription and lexicon) preparation process HTK provides 2 flexible and quite useful tools represented by HDMan (dictionary processing) and HLed (transcription file management). Dictionary can contain both speech models as well as models of background. HDMan enables to use and merge several different dictionaries. In the decoding process HVite tool is used, which calculates the best path and its probability across concatenated models (no full probability is calculated). It can return multiple hypotheses where more tokens per state are allowed. It has several utilizations like: recognition, models (speech unit) time alignment that can be used in speech syntheses, selection of alternative pronunciation if there are any in the dictionary and it supports some kind of real time speech recognition where the input is a direct audio. HTK system is more complex, very flexible, provides up to date functionality, it is regularly updated and well documented. Introduction of ATK greatly enabled its real time application.

II.A.2 JULIUS:

"Julius" is a high-performance, two-pass large vocabulary continuous speech recognition (LVCSR) decoder software for speech-related researchers and developers. Based on word N-gram and context-dependent HMM, it can perform almost real-time decoding on most current PCs in 60k word dictation task. Major search techniques are fully incorporated such as tree lexicon, N-gram factoring, cross-word context dependency handling, enveloped beam search, Gaussian pruning, Gaussian selection, etc. Besides search efficiency, it is also modularized carefully to be independent from model structures, and various HMM types are supported such as shared-state triphones and tied-mixture models, with any number of mixtures, states, or phones. Standard formats are adopted to cope with other free modeling toolkit such as HTK, CMU-Cam SLM toolkit, etc.

The main platform is Linux and other UNIX workstations, and also works on Windows. Most recent version is developed on Linux and Windows (cygwin / mingw), and also has Microsoft SAPI version. Julius is distributed with open license together with source codes.

II.A.3 BROADCASTING:

Broadcast is a mechanism for disseminating identical information from one source to many receivers. It is widely used in many applications ranging from satellite communications to wireless mobile ad hoc networks. There are a variety of Broadcast Schemes defined depending on factors such as speed, reliability, power management adaptability and redundancy. Some of these Schemes are:

II.A.3.1 ROBUST BROADCAST ALGORITHM

A simple protocol that bolsters the reliability of broadcasting in a wireless network. Broadcast generally use Flooding algorithms. RBP focuses on increasing the reliability of each node performing a flood operation. A perfect flood achieves 100% reliability with each node transmitting the broadcast packet exactly once.

$$RCM=(F*BytesPerFlood)\ (Nodes * PktSize)$$

F – Flooding operation frequency of a node
RCM – Normalized Cost.

II.A.3.2 BROADCASTING USING NETWORK CODING

Network coding approaches to wireless network have shown promising schemes for reducing the energy and bandwidth. In case of packet loss, the source node maintains a list of all the receiver nodes where packet need to be retransmitted. These lists from all source nodes are then XOR-ed together using network coding. This helps in retransmitting packets to only those node which are present in the resultant lists. Retransmission operation is performed after a fixed interval of time.

II.A.3.3 SHORT SEEK FIRST SCHEDULING SCHEME

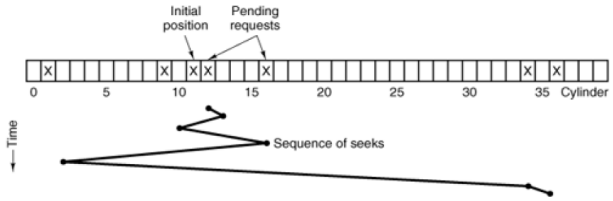
Shortest seek first algorithm scans the request queue for the request that is nearest the head and serves that request first. This algorithm minimizes the total seeking that the head must perform. This algorithm can allow requests to starve. If new requests keep coming in that are near the current position of the head at a sufficient rate, the disk head will never move near enough to other requests to service them.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016



Channel Input Output Relationship is given by:

$$yin = hin \cdot xn + win \text{ for } 1 \leq I \leq k$$

The shortest seek first algorithm determines which request is closest to the current position of the head, and then services that request next.

III. PROPOSED ALGORITHM

A. FUNDAMENTAL EQUATIONS FOR SPEECH RECOGNITION

We will discuss two approaches in this section for speech recognition: Stochastic approach and Template based approach.

III.A.1 STOCHASTIC APPROACH

Assume L be a language composed of set of sentences that the assistant system has to recognize. Let D denote dictionary. An observed sequence X is given for an unknown sentence W, output of the sentence $\hat{W}$. Assume the for each sentence $W = w_1,\ldots,w_g \in L$. Let Pr(W) be the probability for uttering W.

The speech recognition would pick up sentence $\hat{W}$ such that

$$Pr(\hat{W}) = \max_{w} \{Pr(W|X)\}$$

$$Pr(W|X) = \frac{Pr(X|W)Pr(W)}{Pr(X)} \quad \text{[Bayes' formula]}$$

III.A.2 TEMPLATE BASED APPROACH

In the template-based approach to speech recognition, one first builds a collection of reference templates, each itself a sequence of feature vectors that represents a unit (usually a whole word) of speech to be recognized. Then, the feature vector corresponding to the current utterance is compared with each reference vector in turn, via some distance measure. Various distance measures (e.g., log spectral distance, cepstral distance, weighted cepstral distance, and likelihood distortions) have been the subject of research and application.

B. HIDDEN MARKOV MODEL

HMM is an algorithm which uses state transition to produce output. HMM converts digital signals of audio as an input and match the input signal with the model for next signal.

Let Σ be an alphabet of M symbols. A hidden Markov model is a quintuple

$\lambda = (N,M,A,B,\pi)$, where

—N is the number of states, denoted by the integers 1, . . . , N. In the magic example, N = 3, and the states correspond to which hat (red, blue, or yellow) the magician is about to use.

—M is the number of symbols that each state can output or recognize. M = 3 in the magic example, as each symbol corresponds to an animal (hare, guinea pig, or parrot) that can be pulled out of a hat.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

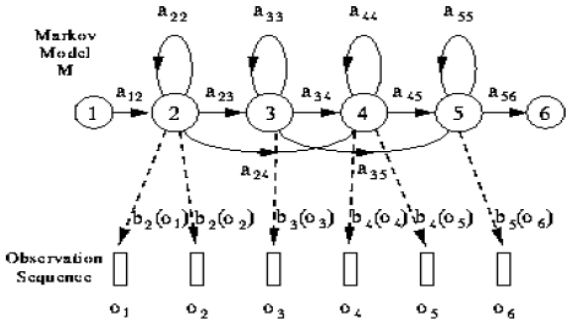
—A is an N ×N state transition matrix such that a_{ij} is the probability of moving from state i to state j, $1 \leq i, j \leq N$.

—B is an observation probability distribution such that $b_j(\sigma)$ is the probability of recognizing or generating the symbol σ when P in state j.

—π is the initial state probability distribution such that π_i is the probability of being in state i at time 1.

The term "hidden" comes from the fact that the states of the Markov model are not observable. In fact, the number of states, output symbols, as well as the remaining parameters of the hidden Markov model are estimated by observing the phenomenon that the unknown Markov chain describes.

An example of HMM state transition model is given in the figure.



C. MARKOV SOURCES

We now define Markov sources (MS), following the notation of Bahl, Jelinek, and Mercer [1983]. Let V be a set of states, E be a set of transitions between states, and

$\check{E} = \Sigma \cup \{\varphi\}$

be an alphabet, where φ denotes the null symbol. We assume that two elements of V, sI and sF, are distinguished as the initial and final state, respectively.

An **information source** is a sequence of random variables ranging over a finite alphabet Γ, having a stationary distribution. A Markov information source is then a (stationary) Markov chain *M*, together with a function

$f : S \rightarrow \Gamma$

that maps states *S* in the Markov chain to letters in the alphabet Γ.

D. VITERBI ALGORITHM

The acoustic signal is treated as the observed sequence of events, and a string of text is considered to be the "hidden cause" of the acoustic signal. The Viterbi algorithm finds the most likely string of text given the acoustic signal.

E. A* ALGORITHM

Computer algorithm that is widely used in path finding and graph traversal, the process of plotting an efficiently traversable path between points, called nodes. Noted for its performance and accuracy.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

IV. SIMULATION RESULTS

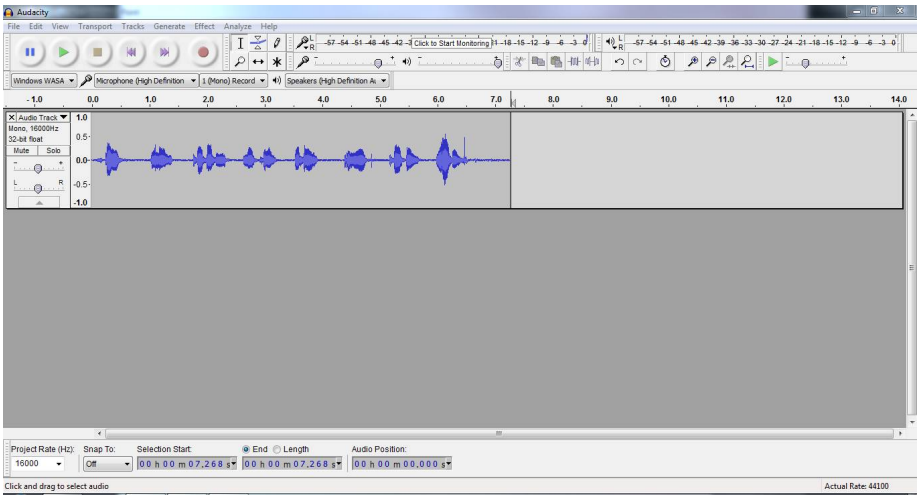


Figure 1. Audacity

The Figure 1. shows the step 1 of the project. In this step , software Audacity is loaded, and a teacher will record his lecture in the Audacity and will save the wav file.

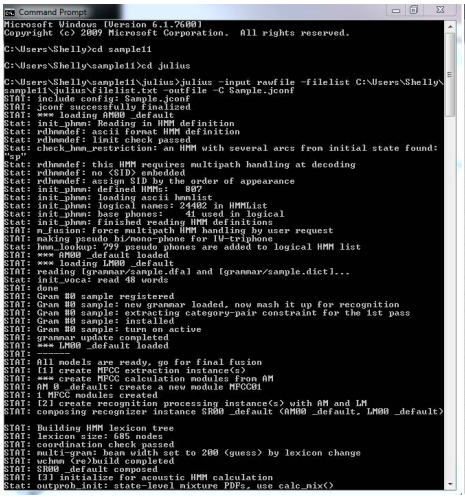


Figure 2. Julius - I

Figure 2. Julius – II

In Figure 2. Julius (I and II), is the simulation of the working of the julius, where the audio file recorded using Audacity is then used as an input to the julius to convert the audio lecture into text, which will then sent to the next step for broadcasting the .txt document to the students currently present in the lecture, through a mobile application.

V. CONCLUSION AND FUTURE WORK

Hence we conclude that, after studying speech recognition and networking, we can combine both the domains to develop a product which can be useful to the students and teachers as well. Since quality education done the right way (using this system) will not only increase the probability of good results but also lead to high retention of concepts and knowledge. Being highly user friendly it will encourage all types of students to use it, hence indirectly increasing the lecture's audience.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

The disadvantage of such a system is that it will require to be trained for each individual to give accurate results. Although HMM models are quite successful, still there are some limitations in this model.

Further research will be required to fully understand the impact of this system on academic performance of the students. Also how this system can be more beneficial to students with disabilities could be considered as an area of improvement. We also need to contemplate on how upcoming technologies like artificial intelligence can help in making such a system perfect for everyday use.

REFERENCES

1. R. G. Adam L. Buchsbaum, Algorithmic Aspects in Speech Recognition: An Introduction.
2. G. Gaikwad and Yannawar, A Review on Speech Recognition Technique. Online International Journal of Computer Application, 2010.
3. N. K. Ranu Dixit, Speech Recognition Using Stochastic Approach: A Review. Chandigarh Engineering College, Landran, Mohali Punjab, India: International Journal of Innovative Research in Science, Engineering and Technology, 2013.
4. T. N. Dong Nguyen, Thinh Nguyen,Wireless Broadcasting Using Network Coding. Oregon State University Corvallis, OR 97331, USA: IEEE Press Wiley.
5. T. N. Dong Nguyen, Thinh Nguyen, Wireless Broadcasting Using Network Coding. Oregon State University Corvallis, OR 97331, USA: IEEE Press Wiley.
6. G. Tiwari, Text Prompted Remote Speaker Authentication: Joint Speech and Speaker Recognition/Verification System.
7. S. R. F. Jelinek, B. Merialdo and M. S. I, A Dynamic Language Mode for Speech Recognition. Thomas J.Watson Research Center, Yorktown Heights, NY 10598:BM Research Division, 1992.
8. J. Kacur, HTK vs. SPHINX for SPEECH Recognition.
9. K. R. Suma Swamy, AN EFFICIENT SPEECH RECOGNITION SYSTEM.