**Speech Recognition and Summarization using Machine Learning**

**By**

**Arghajeet Christopher Gomes [ 12200118062 ]**

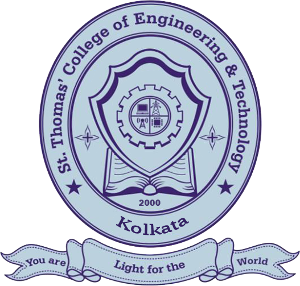
**Sumedha Samanta [ 12200218007 ]**

**Shyam Sundar Jha [ 12200118023 ]**

**Tamal Tripathy [ 12200118012 ]**

**Under the esteemed guidance of**

**Dr. Ghazaala Yasmin**



A project synopsis submitted for the partial fulfillment of

Bachelor of Technology in the

**Department of Computer Science and Engineering**

**St. Thomas’ College of Engineering and Technology**

**Affiliated to**

**Maulana Abul Kalam Azad University of Technology, West Bengal**

**St. Thomas’ College of Engineering and Technology**

**Department of Computer Science and Engineering**

**Declaration**

We declare that this written submission represents our ideas in our own words, and we have adequately cited and referenced the original sources. We also declare that we have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Arghajeet Shyam Sumedha Tamal**

**Christopher Sundar Samanta Tripathy**

**Gomes Jha**

**[ 12200118062 ] [ 12200118023 ] [ 12200218007 ] [ 12200118012 ]**

**St. Thomas’ College of Engineering and Technology**

**Table of Contents**

**Page No.(s)**

1. Preamble

1.1 Vision and Mission i

1.2 Program Outcome (PO) and Program Specific Outcome (PSO) i - ii

1.3 PO and PSO mapping with justification ii - iii

2. Abstract/Summary

2.1. Introduction

* + 1. Problem Statement
    2. Objective
    3. Literature Survey
    4. Brief Discussion on Problem
    5. Organization/ Planning
  1. Concepts and problem analysis
  2. Conclusion

1. Reference / Bibliography
2. **Preamble**
   1. **Vision and Mission**

Vision of the Institute:

* To evolve as an industry oriented, research-based Institution for creative solutions in various engineering domains, with an ultimate objective of meeting technological challenges faced by the Nation and the Society.

Mission of the Institute:

* To enhance the quality of engineering education and delivery through accessible, comprehensive and research-oriented teaching-learning-assessment processes in the state-of-art environment.
* To create opportunities for students and faculty members to acquire professional knowledge and develop managerial, entrepreneurial, and social attitudes with highly ethical and moral values.
* To satisfy the ever-changing needs of the nation with respect to evolution and absorption of sustainable and environment friendly technologies for effective creation of knowledge-based society in the global era.
  1. **Program Outcome (PO) and Program Specific Outcome (PSO)**

POs:

* PO1: **Engineering Knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals and engineering specialization to the solution of engineering problems.
* PO2: **Problem Analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering science.
* PO3: **Design & Development of Solutions:** Design solutions for complex engineering problems and design system components, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
* PO4: **Conduct Investigations of Complex Problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
* PO5: **Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
* PO6: **The Engineer and Society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues, and the consequent responsibilities relevant to the professional engineering practice.
* PO7: **Environment and Sustainability:** Understand the impact of professional engineering solutions in social and environmental context and demonstrate the knowledge of and need for sustainable development.
* PO8: **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norm of engineering practice.
* PO9: **Individual and Teamwork:** Function effectively as an individual and as a member or leader in diverse teams and in multi-disciplinary settings.
* PO10: **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
* PO11: **Project Management and Finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team to manage projects and in multi-disciplinary environments.
* PO12: **Life-long Learning:** Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PSOs of Computer Science and Engineering:

* PSO1: **Programming skills:** Apply fundamental knowledge and programming aptitude to identify, design and solve real life problems.
* PSO2: **Professional skills:** Students shall understand, analyze, and develop software solutions to meet the requirements of industry and society.
* PSO3: **Competency:** Students will be competent for competitive examinations for employment, higher studies, and research.
  1. **PO and PSO mapping with justification**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PO 1 | PO  2 | PO  3 | PO  4 | PO  5 | PO  6 | PO  7 | PO  8 | PO  9 | PO  10 | PO  11 | PO  12 | PSO  1 | PSO  2 | PSO  3 |
| **Audio Summarization** | 2 | 3 | 2 | 2 | 3 | - | - | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

* **PO1-** Throughout this project we are applying the concepts of Physics and engineering in terms of data analysis and advanced machine learning algorithms.
* **PO2**- In this project we are analysing and solving the problem of summarization of audio.
* **PO3**- A real world in-demand problem has been taken up and the solution for the same Is being developed.
* **PO4**- In this project we are using research-based knowledge and analysis of audio files to develop a viable solution.
* **PO5**- In this project we are using advanced algorithms of Machine Learning and analysis in Python.
* **PO8**- Professional ethics are being followed as following the ethical principles and committing to professional ethics are upheld as sacrosanct.
* **PO9**- This is a team project, and we are contributing to it individually as well as collectively.
* **PO10-** As a team we always discuss and put forward each other’s opinions and advance accordingly.
* **PO11**- The knowledge and understanding of management has been applied in our work.
* **PO12**- This project gives us the ability to engage ourselves in independent and lifelong learning and is keeping us updated with the latest technologies.
* **PS01**- Fundamental knowledge and programming aptitude has been applied at every step.
* **PSO2**- Requirements have been analysed and prototype is being developed to meet industry standards.
* **PSO3**- This project gives us the ability to engage ourselves in independent and lifelong learning and is keeping us updated with the latest technologies and thus, will definitely help us improve ourselves in the industry.

* 1. **Introduction**
     1. **Problem Statement**

Introducing a supervised learning-based model for recognizing speech and incorporating the recognized speech for summarization through unsupervised learning.

* + 1. **Objective**

The main objective of the proposed model is to understand and extract the informative features from the speech data and recognizing the speakers from the given speech. Furthermore, the recognized speech will be summarized by extracting the informative contents of the speech by incorporating relevant features in an unsupervised way. The summarized file will contain considerable length and the core sentiments as well as information of the original audio file. This concept can be applied to a wide variety of applications which includes news, meeting, podcast, sports-cast summary.

The motivation to pursue such a concept is to help people get a gist of the vast content while enabling them to save time and conserve the limited storage resources.

* + 1. **Literature Survey**

Furui et al. [1] has introduced speech-to-text and speech-to-speech automatic summarization based on important sentence extraction and words concatenation based on linguistic and grammatical likelihood, thereby generating summary of the original speech.

C. Hori and S. Furui [2] proposed a new automatic speech summarization method based on a “summarization score” which is a measure of word significance, confidence, linguistic likelihood, and word concatenation probability (given by Stochastic Dependency Context Free Grammar, i.e., SDCFG).

Murray, Gabriel, et al. [3] analyzed the importance of features like speaker activity, turn-taking and discourse cues in an automatic speech summarization system used for summarizing meeting recordings.

Penn, Gerald, and Xiaodan Zhu [4] criticized evaluation standards in terms of: baselines chosen, comparison of the results with what “summaries” should be, and the efficiency of adding speech-related features to automatic speech recognition systems.

Maskey, Sameer, and Julia Hirschberg [5] introduced Broadcast News Summarization System using a combination of lexical, acoustic/prosodic, structural and discourse features for predicting the news segments to be included in the summary.

Povey, Daniel, et al. [6] designed Kaldi which is a speech recognition system based on finite-state transducers and scripts for building complete recognition systems using modelling of arbitrary phonetic-context sizes, acoustic modelling with subspace and standard Gaussian mixture models.

Swamy, Suma, and K. V. Ramakrishnan [7] developed a speech recognition system using Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and Hidden Markov Model (HMM) for making recognition of speakers and speech faster, while making the system more efficient and accurate.

Halle, Morris, and Kenneth Stevens [8] proposed a model which transforms the speech into a sequence of phonemes, carried out through an active or feedback process. Here, patterns are generated internally in the analyzer according to an adaptable sequence of instructions until an optimum match is found.

Abdel-Hamid, Ossama, et al [9] has shown that the efficiency of speech recognition systems using hybrid deep neural network (DNN)- hidden Markov model (HMM) can be further enhanced by using convolutional neural networks (CNNs) which with its local connectivity, weight sharing, and pooling exhibit some degree of invariance to small shifts of speech features along the frequency axis, thereby reducing the error rate by 6 – 10%.

* + 1. **Brief Discussion on Problem**

**Diagram

Description automatically generated**

* 1. **Concepts and problem analysis**

**2.3 Conclusion**

[mention imp of speech summarization] The proposed methodology has introduced an unsupervised based notion for selecting important and informative frames by incorporating relevant features. The novelty of the problem may be extended by recognizing speakers in a mixture of multiple speakers and then to summarize it accordingly.

1. **Reference / Bibliography**

[1]. Furui, Sadaoki, et al. "Speech-to-text and speech-to-speech summarization of spontaneous speech." IEEE Transactions on Speech and Audio Processing 12.4 (2004): 401-408.

[2]. C. Hori and S. Furui, "A new approach to automatic speech summarization," in IEEE Transactions on Multimedia, vol. 5, no. 3, pp. 368-378, Sept. 2003, doi: 10.1109/TMM.2003.813274.

[3]. Murray, Gabriel, et al. "Incorporating speaker and discourse features into speech summarization." Proceedings of the Human Language Technology Conference of the NAACL, Main Conference. 2006.

[4]. Penn, Gerald, and Xiaodan Zhu. "A critical reassessment of evaluation baselines for speech summarization." Proceedings of ACL-08: HLT. 2008

[5]. Maskey, Sameer, and Julia Hirschberg. "Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization." Ninth European Conference on Speech Communication and Technology. 2005.

[6]. Povey, Daniel, et al. "The Kaldi speech recognition toolkit." IEEE 2011 workshop on automatic speech recognition and understanding. No. CONF. IEEE Signal Processing Society, 2011.

[7]. Swamy, Suma, and K. V. Ramakrishnan. "An efficient speech recognition system." Computer Science & Engineering 3.4 (2013): 21.

[8]. Halle, Morris, and Kenneth Stevens. "Speech recognition: A model and a program for research." IRE transactions on information theory 8.2 (1962): 155-159.

[9]. Abdel-Hamid, Ossama, et al. "Convolutional neural networks for speech recognition." IEEE/ACM Transactions on audio, speech, and language processing 22.10 (2014): 1533-1545.