Aaryan Jha

Senior Project Proposal

*Speech Emotion Recognition System*

Section: CSC498 A

May 13, 2020

Table of Contents

[Original Proposal 3](#_Toc134896510)

[Project Summary 3](#_Toc134896511)

[Significance 4](#_Toc134896512)

[Required Tools & Availability 4](#_Toc134896513)

[Demonstration Plans 5](#_Toc134896514)

[Qualifications 6](#_Toc134896515)

[Project Specifications 7](#_Toc134896516)

[Functional Specifications 7](#_Toc134896517)

[User Interface Specification 7](#_Toc134896518)

[Technical Details 8](#_Toc134896519)

[Timeline 9](#_Toc134896520)

[Future Enhancements 11](#_Toc134896521)

[Deviations from Original Proposal & Justifications 11](#_Toc134896522)

[Project Code 12](#_Toc134896523)

[Conclusion and Lessons Learned 12](#_Toc134896524)

[References 13](#_Toc134896525)

# 

# Original Proposal

# Project Summary

Technology is frequently advancing. The progress that is being made in the Artificial Intelligence (AI) field has made life simpler and more efficient for every individual. The Speech Emotion Recognition project gives insight into AI capabilities. It is a subsection of Artificial Intelligence and Natural Language Processing that focuses on the analysis of emotional states of people expressed through speech. The aim of the project is to recognize and successfully categorize different emotions and extract the emotion expressed.

The project involves developing different algorithms and constructing AI models that can identify emotions from speech signals. The speech data can be in the form of text transcripts, audio recordings, or video recordings with verbal speech. The system then analyzes the patterns, prosody and other linguistic and acoustic features of the speech and determine the emotional state of the speaker. The system tells the users the following result from the speech recognition such as happy, sad, angry, or continuous emotional score in the form of a probability distribution.

The completion of the project requires knowledge of constructing a model to train and test audio/voice data along with AI, NLP and other speech processing techniques. Expertise in developing and evaluating machine language models is highly crucial in this project.

# Significance

The Speech Emotion Recognition project is a project that will successfully predict human emotions. Since the project involves Artificial Intelligence, Natural Language Processing, Machine Learning, and other various aspects of computer science it significantly helps improve your overall knowledge of Computer Science.

The project is a challenging problem in AI with its requirement of developing advanced algorithms and machine language learning models to accurately recognize human emotions. Along with that, the NLP techniques involved in this project to analyze speech signals and determine emotional state is highly significant into understanding the functionality of AI. The completion of the project will allow enhance human-computer interaction by allowing computers to recognize and respond to emotions which could be used in the further development of an Alexa or Siri type system.

Overall, the successful development of this system has the potential to revolutionize the way we interact with computers and bring us closer to creating truly intelligent and empathetic machines.

# Required Tools & Availability

To complete a Speech Emotion Recognition project, the required tools are as follows:

* Software:
  + Python programming language to implement the machine learning algorithms and models
  + A deep learning framework like TensorFlow and Pytorch to train and evaluate the models
  + A speech processing library like librosa or Praat to extract speech features from audio recordings
  + An NLP library to process text data
  + Data visualization tools to plot the results
  + Installation of Anaconda and further installation of Spyder can help fulfill all the above requirements.
* Hardware:
  + A computer with a CPU or GPU suitable for deep learning.
  + A large dataset of speech signals with annotated emotional labels.

The required tools and hardware to complete the speech emotion recognition project are widely available, making it possible for any individuals to develop and implement these systems in their project.

# Demonstration Plans

To demonstrate the Speech Emotion Recognition project during any of the checkpoint meetings, these items will be required:

* A laptop to run the demonstration
* Projector to display the presentation prepared and results of the project.
* All the cables to connect a Mac into a projector, extension cords, type C converter etc.
* As backup, all the files will be also sent to the professor if Mac fails to perform, the professor could use DePauw’s windows computer to display the presentation.

By following these steps, the presenter can provide a comprehensive and interactive demonstration of the Speech Emotion Recognition project, highlighting its strengths and limitations, and showcasing its potential for real-world applications in all the checkpoint meetings throughout the semester.

# Qualifications

To complete a Speech Emotion Recognition project, the experience and knowledge I have are as follows:

* **Courses:**

I have completed coursework for Data Mining, Data Science and I am taking Artificial Intelligence right now to enhance my knowledge in this field. Furthermore, outside online courses such as Google Data Analytics Certification, Machine Learning A-Z, The Data Science Course 2023 will also aid in the understanding and completion of the project

* **Knowledge of Programming Languages and Tools:**

The experience of using the programming language, Python, will aid immensely in the completion of the project. Since I have done most of my projects in the past in Python, I have a deep understanding of all the functionality of Python and the libraries it offers to users.

* **Knowledge of Machine Learning and Deep Learning Algorithms:**

Experience in making models for Housing Price Prediction Project, Mask Detection Project, I have a good understanding of linear and logistic regression, decision trees, and neural networks. Since I have already applied different modeling techniques in my previous projects, application of those techniques will help me complete this project

The foundation of machine learning and deep learning along with the growing interest in the same field will aid me in the completion of my Speech Emotion Recognition project.

# Project Specifications

## Functional Specifications

The functional specifications for the Speech Emotion Recognition project are as follows:

* The program will first accept files and process the signals received from the files that is to be used for emotion recognition.
* The signals will then be pre-processed to remove any noise or sound to smoothen the data and improve the quality of the signals to attain higher accuracy.
* After the data has been cleaned, the program will use the signals and extract prosody, pitch, rhythm etc. that are indicative of emotions to be used in the models.
* The system will then train different machine learning models on the dataset obtained from the previous step using models like the Convolutional Neural Network, Long Short-Term Memory Network and Support Vector Machines.
* After the data has been trained, it will classify the different emotions from the trained data and showcase the results.
* Further evaluation will be done using different models and will be compared between one another using accuracy, precision, etc.

## User Interface Specification

The User Interface of this project will be simple and easy to use and understand. Since the result of this project will be based on a prepared report, the User Interface will only have these following components:

* Users will be allowed to enter audio files into the system; existing audio files or freshly created audio files.
* Users will have the option to choose a specific model they would want to train the data or the option of train it with all the models and compare the results.
* The results from running the code will be presented in table and graph form for users to interpret the results according to their own understanding and derive their own conclusion.
* It will be a single page display where people can put their files, choose what model they want to use and see the results of the program.

# Technical Details

The Speech Emotion Recognition Project will be required to detect human emotions from the dataset that is provided to the system. To make the system work properly, it needs to be designed in a specific pattern. The design of the program are as follows:

* **Entry of Data and Pre-Processing of Data**

The audio files are required to be inputted into the system for the program to begin its task. Once the files are properly accepted within the system, the data needs to be cleaned before it can be used which includes filtering any missing audio recordings and detecting outside noise and filtering them out. Once that is completed the audio files can be used to extract the pitch, energy, and prosody of the sound produced by the files that can be used to train the data.

* **Model Selection and Training**

Once the files are cleaned and ready to be trained, appropriate model needs to be selected to train the data. Dense Neural Network is one of the models that will be implemented. It is highly effective in a variety of speech processing tasks. Another model is the Support Vector Machine. It is useful for this project due to its ability to handle high dimensional data and classifying data points into multiple categories. It can separate data into different classes which is well suited for this project. Other models will also be selected after proper research is done to find the best options. After the selection of the models, the data will be trained according to the models required by splitting it into training, testing, and validation sets. This will help the system test out its experiments, make sure the data is not overfitting and give out results after the completion of the experiments.

* **Evaluating and Showcase of Results**

The completed of the experiments will give the system results. These results will be accuracy, precision, etc. in different forms to showcase the results of the experiment. This will allow the program to understand and compare models in the system and decide which model will be best fit to get the most efficient results. The resulted data can be visualized and displayed in the form of tables and graphs to further comparison. Additionally, a report will be made on all the results of the project to provide a comprehensive understanding of the performance and limitations of the projects and demonstrate its practical application and further future development that can be made.

# Timeline

The timeline for this project is as follows:

**Checkpoint 1 (March 6th):**

* I will obtain the different dataset I would like to use for this project
* I will pre-process and extract the data and have it ready for model implementation.
* Have a report and a presentation of my work to explain what I have done and what are my plans for the next step.

**Checkpoint 2 (March 20th):**

* I will select CNN model as my first model to test out my data and evaluate my results.
* I will create different forms of depth, layers, etc. to see the difference each style of CNN will bring to the results.
* I will have a report of the first model I choose, how I trained and tested my data and combine it with how I pre-processed and extracted my data and have a presentation to show it in class.

**Checkpoint 3 (April 17th):**

* I will add Long Short-Term model to successfully train the data and showcase the results in different data visualization forms such as Heat Maps, Line Charts, Bar Charts, etc.
* The report will be further updated with recent changes and a presentation will be prepared to share all the progress I have made in the project.

**Checkpoint 4 (May 1st):**

* I will evaluate the performance of the system and include larger dataset, implement the Support Vector Machine model in the project and compare all the results from the desired models I implemented.
* I will make the system user friendly by allowing users to decide what files they want to use to train and test their models.
* I will make the system allow users to choose if they want to use a specific model for training and want to compare different models’ efficiency to have their own conclusion.
* I will report the results and performance of the system and provide a conclusion and future directions for further improvement.
* A final presentation will be prepared to conclude my project and briefly explain every step I have taken throughout the project.

# Future Enhancements

This project has an endless number of future enhancements that can be made to it. The project could be enhanced to provide real time emotion recognition and allowing immediate analysis of the emotions expressed. Facial expression recognition system can also be implemented to improve the accuracy of the experiments. This project will only understand the language “English” for the current experiment. Expanding the range of the system’s ability of understand multiple languages could be another future enhancement to this project. Finally, this system could be developed into having a real time conversation with the user and providing responses according to the emotion of the user.

# Deviations from Original Proposal & Justifications

Initially, the project was designed with the end results focus on testing the accuracies of various models and drawing conclusions based on the results. The intention was to evaluate the performance of different emotion recognition models and analyze their effectiveness in accurately identifying emotions in speech.

However, as the project progressed in the due course, additional features was added into the system to enhance its usability and practicality. A Graphical User Interface was included which allowed the users to record their own audio or input audio files for real – time emotion recognition. This addition was the different from the original proposal which allows a better, efficient user-friendly experience.

Project Code

The code can be found in this link:

<https://drive.google.com/drive/folders/1-FHYg_nliRVcsRqyE6mzDqPJN5lD3O8D?usp=share_link>

# Conclusion and Lessons Learned

The experience of creating the Speech Emotion Recognition System project was both educational and fulfilling. I learned a great deal about audio processing, machine learning algorithms, and user interface design. The experience I gained while implementing and fine-tuning various machine learning models was invaluable to me.

Some aspects of the project went exceptionally well. Preparing the data for the models and attaining satisfactory accuracy levels in detecting emotions from speech samples allowed the system to meet all expectations. Visualizing the data to interpret how the models decided which emotion to predict also provided insight into how the models functioned in the background.

However, there were some challenges along the way. Having no prior experience in creating a graphical user interface that would accept a learning file and implement predicted data to evaluate real-time results, initially, putting the idea into action was difficult. Breaking the process down into multiple steps and handling each section of the GUI separately helped in the completion of the project.

These challenges contributed to the sense of accomplishment I felt upon completing the project. The Speech Emotion Recognition System has the potential to analyze and understand human emotions in various contexts. This system can be further improved to respond to humans based on their emotions. With limitless possibilities for growth, the journey of creating the project has been a rewarding experience that I plan to continue even further.

# References

Nasim, Sheena. *Speech Emotion Recognition Using Deep Learning*. 2 Feb. 2021, blog.dataiku.com/speech-emotion-recognition-deep-learning.

ProjectPro. “Speech Emotion Recognition Project Using Machine Learning.” *ProjectPro*, 14 June 2022, [www.projectpro.io/article/speech-emotion-recognition-project-using-machine-learning/573](http://www.projectpro.io/article/speech-emotion-recognition-project-using-machine-learning/573).

Hmida, Firas. “How to Build Your Own Speech Emotion Recognition?” *Vivoka*, 3 Nov. 2022, vivoka.com/how-to-speech-emotion-recognition.