YANGON TECHNOLOGICAL UNIVERSITY

DEPARTMENT OF COMPUTER ENGINEERING
AND
INFORMATION TECHNOLOGY

Development of Children Gender Classification System Using Speech

(Pre-Defence Seminar)

Supervised By
Dr. Su Su Maung

Presented By
Ma Khin Aye Chan
M.E CEIT -2

November 6, 2019

* Outlines

- Objectives
- Introduction
- Application Areas
- Problem Statement
- System Design
- Dataset Preparation
- Methodology
- Results and Discussion
- Conclusion

* Objectives

- To study the algorithm of feature extraction (MFCC) and machine learning algorithms for classification
- To understand the speech recognition operations in details
- To implement a gender classifier that can automatically predict the gender of the speaker

* Introduction

- Gender classification is to determine a person's gender, e.g., male or female, based on his or her biometric cues.
- There are a number of biometrics which may be used to classify gender such as the face, eyes, fingerprint and hand shape, speech etc.
- This system analyzes speech signals to predict the gender of the speaker.

* Application Areas

- Automatic Speech Recognition
- Human-Computer Interaction
- Multimedia Information Retrieval
- Commercial Development
- Demographic Research
- Mobile Applications and Video Games

* Problem Statement

- Gender identification of children is difficult than adults, it is confusing to identify whether the speaking child is male or female.
- Due to underdeveloped vocal tract and thin vocal folds in both male and female child, there is no significant difference in their acoustic-phonetic properties.

* System Design

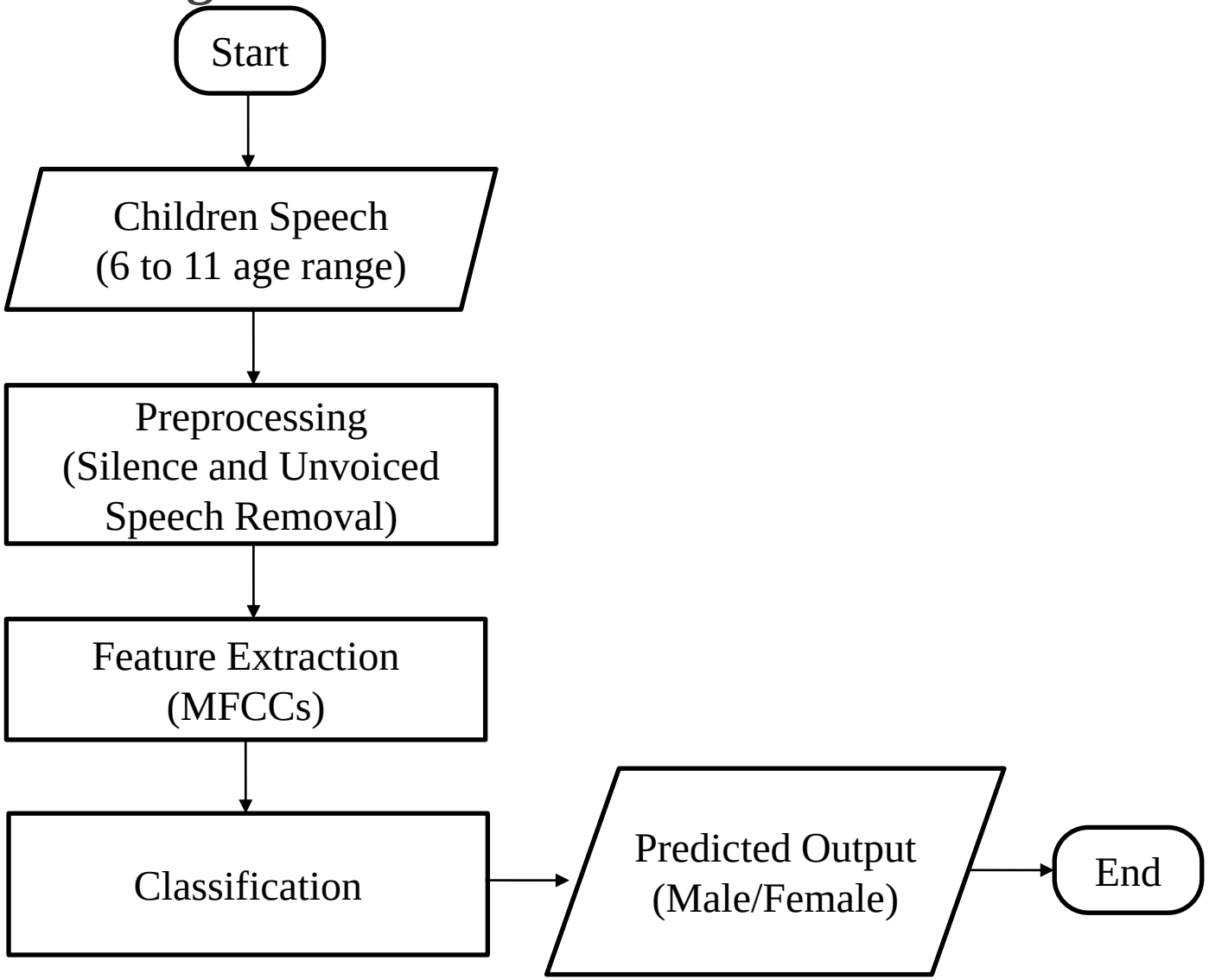

Figure 1: System Flow Diagram

* Dataset Preparation

- Firstly, I prepared five Myanmar sentences:
  1. မင်္ဂလာပါ
  2. ကျေးဇူးတင်ပါတယ်
  3. နာမည်ဘယ်လိုခေါ်လဲ
  4. ဘယ်သွားမလို့လဲ
  5. နေကောင်းလား

File Type	.wav format
Duration	2 or 3 second
Numbers of Channel	Mono (1 Channel)
Sampling Frequency	44.1 kHz
Number of Bits	16 bits

Table I. Recording Specifications

* Dataset Preparation (Cont'd)

- Dataset used in the system includes total of 1100 audio records.
- The children range in age from 6 years to 11 years.
- The female records contain 566 samples where male records have 534 samples.

❖ **Silence and Unvoiced Speech Removal**

- The speech recordings of children consists of many silence and unvoiced regions.
- Leading/trailing silence in the audio may not contain much information and thus not useful for the classification.
- Hence, removing this silence is done in preprocessing step.

* Methodology

❖ Feature Extraction using MFCC Algorithm

- Mel-frequency Cepstral Coefficients (MFCCs)
- MFCC technique takes frequency domain as its standard base and thus it approximates the human system response more closely than any other system.
- It is based on the short term analysis, and thus from each frame of speech signal a MFCC vector is computed.

* Methodology (Cont'd)

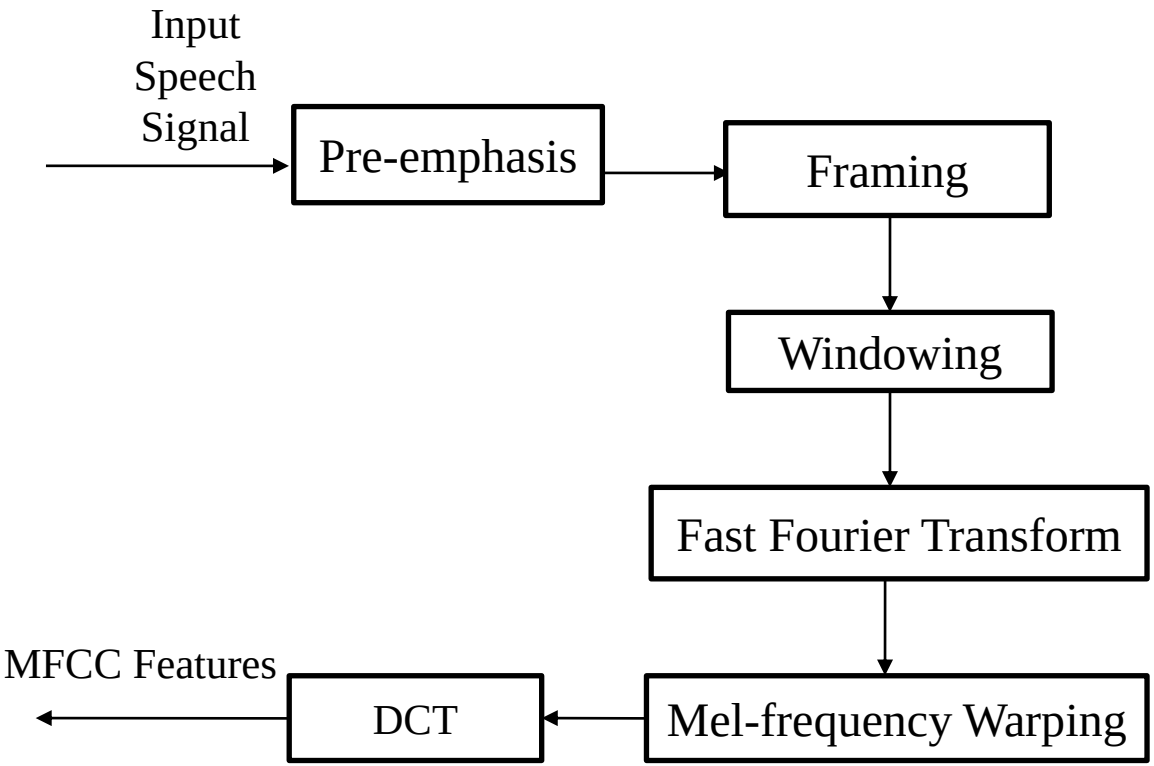

Figure 2: Steps for Computing MFCCs

* Methodology (Cont'd)

- Extracted features points which were collected in the CSV file can be seen in the figure 3.

mfcc1	mfcc2	mfcc3	mfcc4	mfcc5	mfcc6	mfcc7	mfcc8	mfcc9	mfcc10	mfcc11	mfcc12	label
-4.015884	-5.337801	3.880235	-2.612753	-30.765772	-31.996782	-15.814099	-3.350799	-3.706910	-17.836277	-13.396485	-10.769270	male
-0.398436	-22.386601	-17.671561	1.157095	-25.274026	-13.027446	-30.996784	9.600207	1.963798	-13.853052	-30.468489	-15.506199	male
-1.668531	-4.332610	-13.569998	-10.448085	-34.465060	-11.890284	-23.942041	-33.508446	13.719242	18.943690	-9.748084	-11.789779	male
-2.241353	-8.969100	7.169721	18.020628	-5.844882	-24.045389	-23.155617	-3.639931	-29.574649	-11.762388	-7.103109	-15.055187	female
-5.199849	7.509791	-29.952706	9.367084	-20.967097	8.949095	-19.338754	-26.104784	-4.497418	6.279948	-13.067591	-7.391434	female
-2.333304	-11.949313	5.424600	4.928143	-26.588636	-20.469231	-20.621021	-12.143301	-27.850742	-15.173062	4.316390	-17.864485	female

Figure 3

* Methodology (Cont'd)

❖ Classification

- Classification is establishing a mathematical model that separates into male and female based on the features of children's speech.
- Classification model is built on the training set and check the accuracy of the model by using it on the testing set.
- In this system, machine learning classification algorithms are compared using MFCC feature dataset.
- Train and test set accuracies are observed for five classification algorithms.

* Methodology (Cont'd)

❖ Classifiers

- Random Forest (RF)
- Artificial Neural Network (ANN)
- Logistic Regression (LR)
- Support Vector Machine (SVM)
- Gaussian Naive Bayes (GNB)

* Results and Discussion

- In this system voice dataset contains 1100 audio records (550 for each gender).
- Two testing is done to estimate the performance of the model: k-fold cross validation method and simple train test split.

❖ k-Fold Cross-Validation

- 10 fold cross validation are used to train the models.
- The accuracies are shown in the Table I.

* Results and Discussion(Cont'd)

Classifiers	RF	ANN	SVM	LR	GNB
10-fold Cross Validation Score	80%	77%	76%	81%	72%
	80%	76%	75%	70%	77%
	88%	75%	74%	70%	75%
	85%	80%	79%	76%	80%
	79%	77%	75%	72%	72%
	85%	76%	74%	82%	73%
	75%	75%	72%	80%	72%
	81%	74%	80%	77%	70%
	85%	78%	80%	75%	79%
	89%	72%	81%	78%	70%
Average	83%	76%	77%	76%	74%

Table II. Cross Validation Score

* Results and Discussion (Cont'd)

❖ Simple Train Test Split

- 90% of dataset is used as training dataset and 10% is used for testing.
- Correct and incorrect predictions can be seen in following bar charts.

* Results and Discussion (Cont'd)

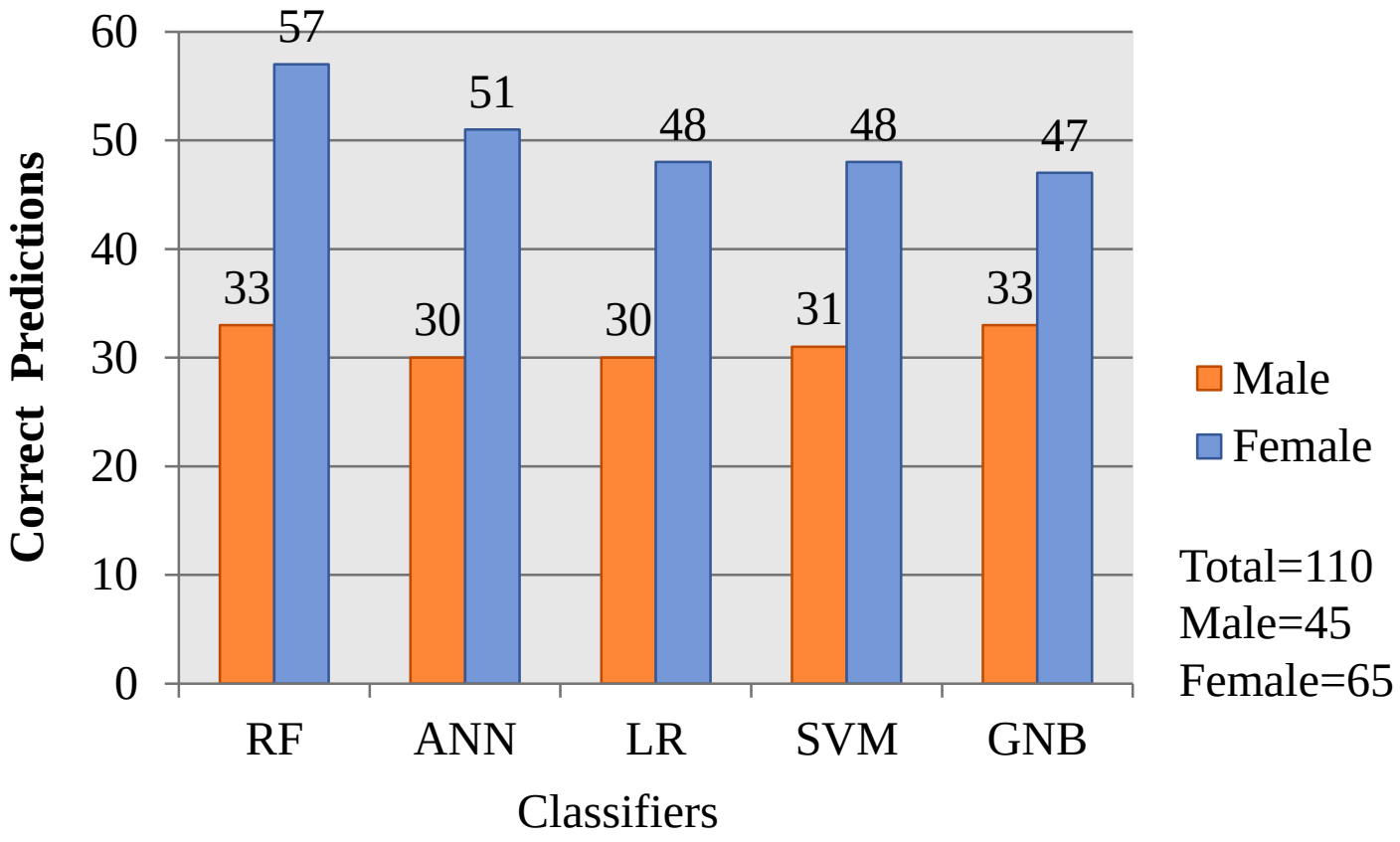

Figure 4:Bar Chart for the Number of Correct Predictions for Each Classifier

* Results and Discussion (Cont'd)

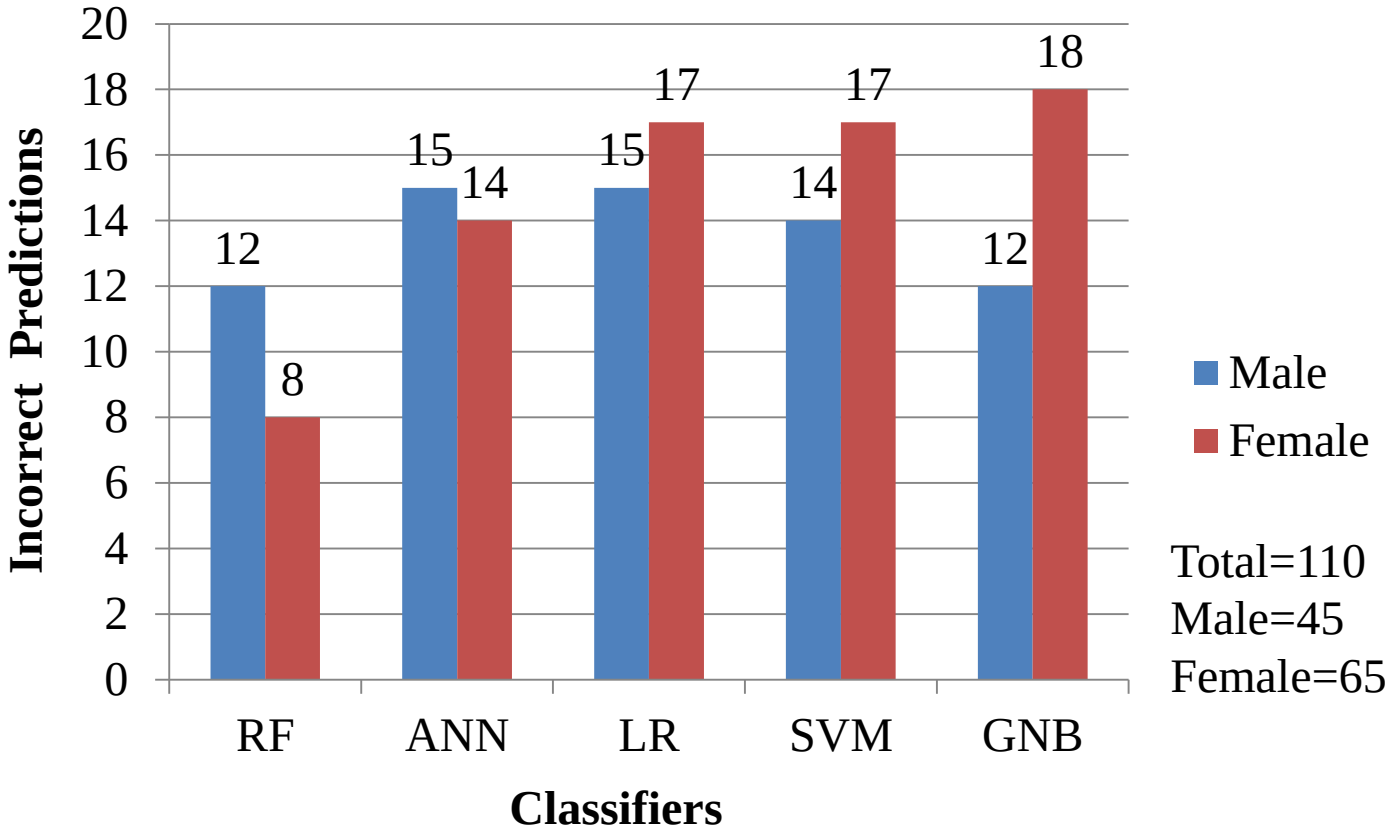

Figure 5:Bar Chart for the Number of Incorrect Predictions for Each Classifier

* Results and Discussion (Cont'd)

Classifiers	Training Accuracy	Testing Accuracy		Precision	Recall	F1-score	Support
RF	99%	83%	Female	81%	88%	84%	65
			Male	85%	77%	81%	45
ANN	92%	78%	Female	76%	86%	81%	65
			Male	81%	69%	74%	45
LR	78%	76%	Female	77%	74%	75%	65
			Male	76%	79%	77%	45
SVM	79%	77%	Female	72%	82%	77%	65
			Male	83%	73%	78%	45
GNB	75%	74%	Female	74%	75%	74%	65
			Male	74%	72%	73%	45

Table III. Performance Measures of Each Classifier

* Results and Discussion (Cont'd)

- From the above results: cross validation scores and testing set accuracies, Random Forest performs better compared with other machine learning algorithms to classify the gender of a child using MFCC features of voice.

* Conclusion

- The gender classification system implemented by applying Python programming language and the experimental results has been analyzed.
- The analysis of the results shows that the performance of the proposed system is good, as the average accuracy of RF classifier is 83% .
- Therefore, it can be extended to the another researchers and can also be tested by using different features and other classification techniques.

Thank you for your attention!!!