UNIVERSITI TEKNOLOGI MARA
CAWANGAN PULAU PINANG

VOICED-UNVOICED SEGMENTATION USING TIME-DOMAIN FEATURES FOR IMRPOVED SPEECH RECOGNITION SYSTEM

AHMAD AKMAL BIN MOHD ROSLEY

BACHELOR OF ENGINEERING (HONS.)
ELECTRICAL AND ELECTRONIC ENGINEERING

February 2025

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Ahmad Akmal bin Mohd Rosley
Student I.D. No.	:	202276
Programme	:	Bachelor of Engineering (Hons.) Electrical and Electronic Engineering – (CEEE200)
Faculty	:	Electrical Engineering Studies
Thesis	:	Voiced-Unvoiced Segmentation using Time-domain Features for Improved Speech Recognition System
Signature of Student	:	.....
Date	:	February 2025

ABSTRACT

This thesis explores a method for voiced-unvoiced segmentation using time-domain features to enhance speech recognition systems. Voiced unvoiced segmentation is the process of identifying both voiced and unvoiced parts of a spoken stream. Voiced-unvoiced segmentation in speech recognition often struggles with balancing accuracy and computational efficiency, with simpler methods failing in complex conditions and frequency-domain approaches being computationally demanding. V-UV segmentation was designed to increase accuracy and decrease computation time in data processing. The study utilizes a database of audio recordings from 51 female speakers uttering three isolated English words. Preprocessing methods, such as normalization and pre-emphasis, are used to prepare audio data for analysis. V-UV segmentation utilizes ZCR and STE to separate voiced and unvoiced frames from the speech signal where only voiced frames are the focus of the feature extraction procedure, which uses LPC to extract coefficients that reflect the spectral envelope of the spoken signals. Following segmentation, the findings show considerable frame reduction percentages for each word, "Aluminium" showed a frame reduction of 24.95%, "Better" showed a frame reduction of 31.95%, and "Communication" showed a frame reduction of 40.58%. Moderate accuracy was demonstrated by the DA classifier utilizing Mahalanobis distance, where the highest overall accuracy recorded was 73.33% for LPC order 19, offering the greatest speech signal representation for classification in the proposed system.

ACKNOWLEDGEMENT

I would like to convey my heartfelt appreciation to my supervisor, Dr. Yusnita binti Mohd Ali, and my co-supervisor, Ts. Anith Nuraini binti Abd Rashid, for providing me with valuable advice, support, and encouragement during my Final Year Project. Despite having taken part in a different elective, Dr. Yusnita's passion and expertise in the subject matter equipped me with the information and abilities required to successfully finish this final year project.

I would also like to express my heartfelt gratitude to my parents for their constant support, patience, and encouragement during my academic path. Their confidence in my ability has been a constant source of inspiration.

Finally, I want to express my gratitude to Universiti Teknologi MARA (UiTM) Permatang Pauh for providing the necessary resources for the completion of my final year project.

TABLE OF CONTENTS