

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

**Speech Recognition Algorithm of Major Nigerian Languages (Yoruba, Hausa, Ibo) Using K-NN**

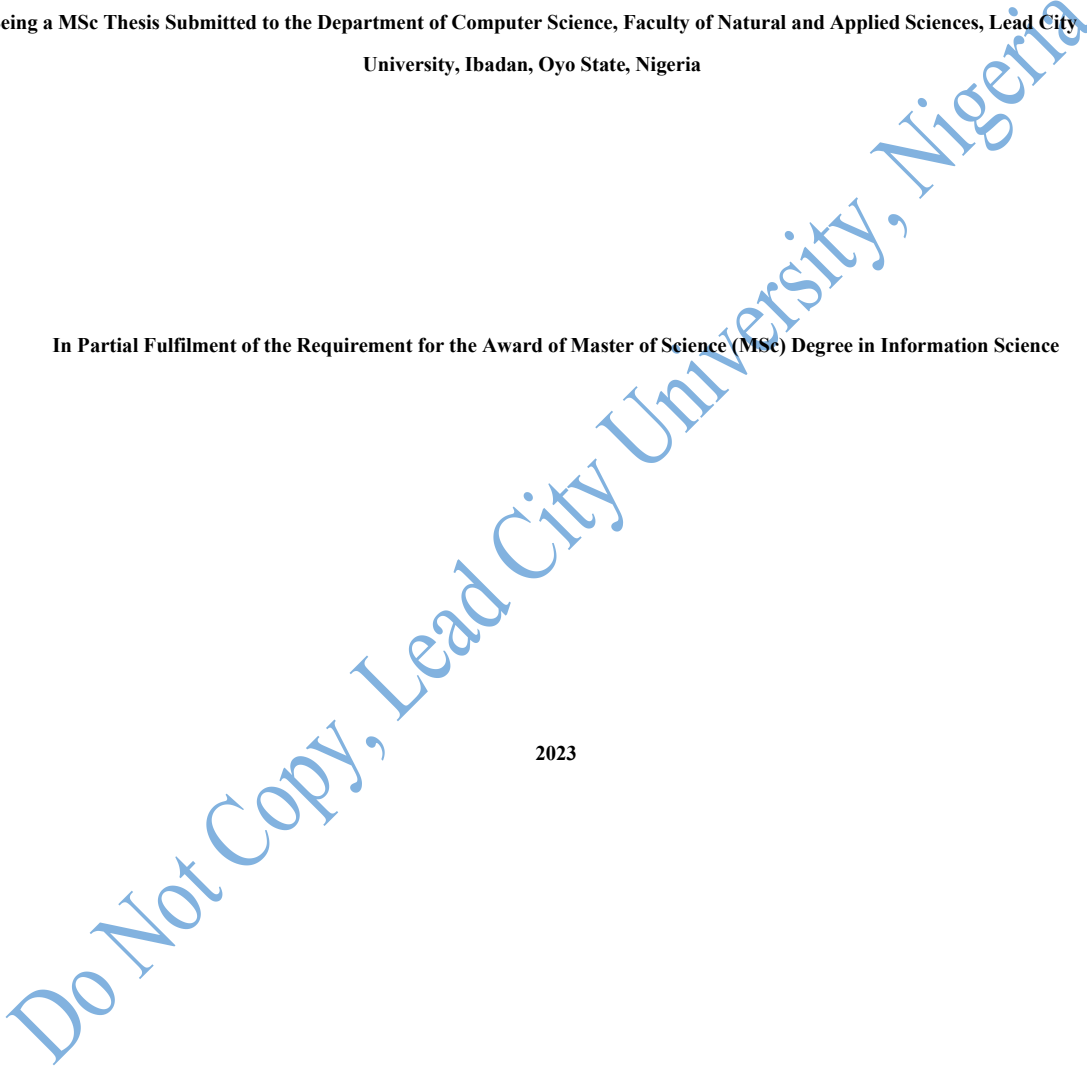
**Taiwo Mauyon KUPONU**

**LCU/PG/002632**

**Being a MSc Thesis Submitted to the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, Nigeria**

**In Partial Fulfilment of the Requirement for the Award of Master of Science (MSc) Degree in Information Science**

**2023**



24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

**Certification**

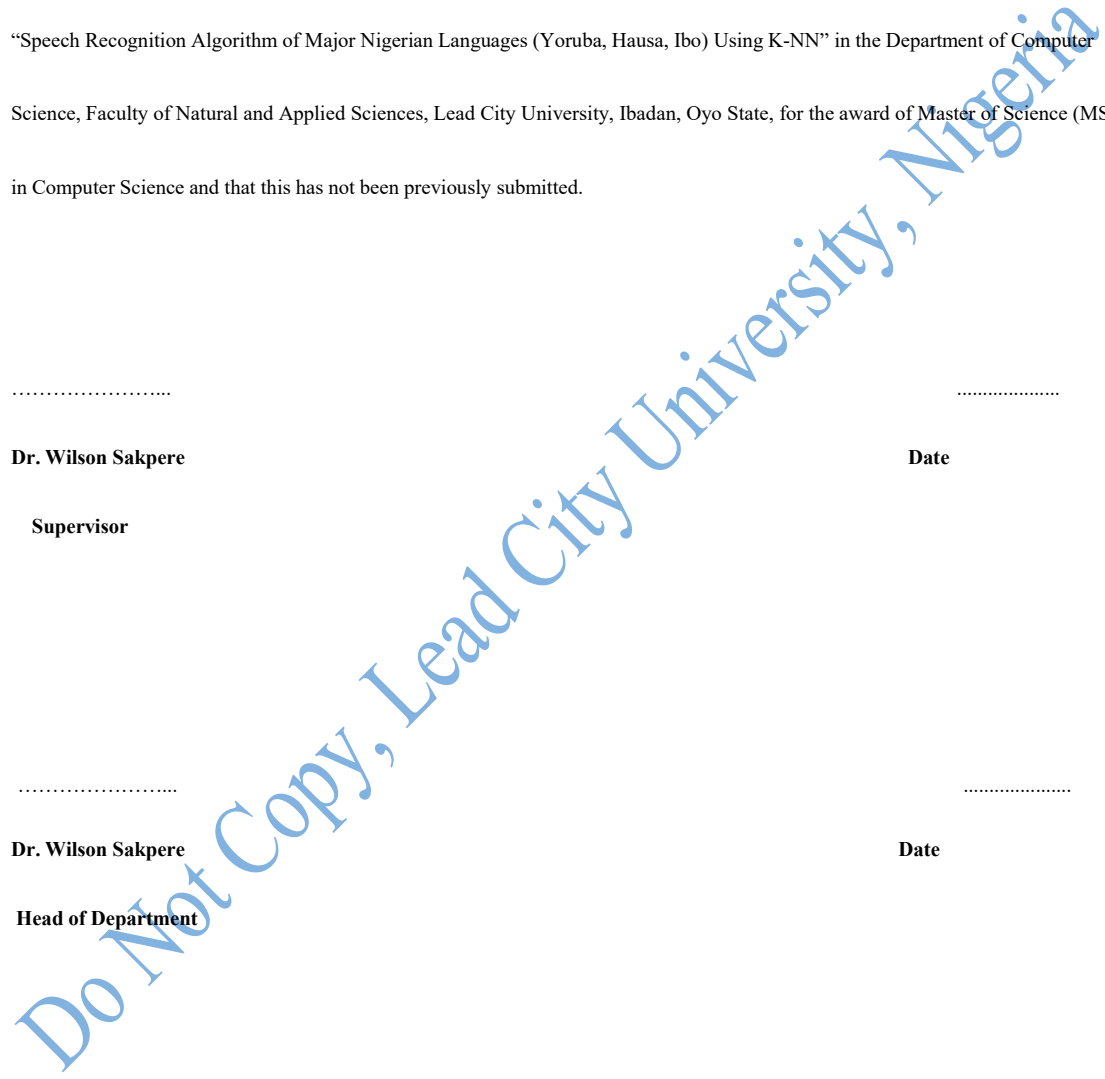
This is to certify that Taiwo Mauyon KUPONU with matriculation number LCU/PG/002632 carried out this research work titled “Speech Recognition Algorithm of Major Nigerian Languages (Yoruba, Hausa, Ibo) Using K-NN” in the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, for the award of Master of Science (MSc) in Computer Science and that this has not been previously submitted.

.....  
**Dr. Wilson Sakpere**  
**Supervisor**

.....  
**Date**

.....  
**Dr. Wilson Sakpere**  
**Head of Department**

.....  
**Date**



49

50

51

**Dedication**

52

This research work is dedicated to God, my parents, and my siblings.

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

*Do Not Copy, Lead City University, Nigeria*

68

69

70

### Acknowledgement

71

Foremost, I would like to express my gratitude to the leadership of the Lead City University, Ibadan and also acknowledge the

72

libraries used for creating a medium for us to acquire knowledge for self-reliance.

73

I acknowledge my supervisor Dr. Wilson Sakpere, for the continuous support for M.Sc study and research, for his motivation,

74

enthusiasm, to guide me through the research. The immense knowledge and plentiful experience have encouraged me in all the time

75

of my academic research and daily life. Besides my supervisor, my sincere thanks go to the Head of Department Dr. Wilson Sakpere,

76

the Postgraduate Coordinator Dr. Azeez Waheed, all other lecturers and staff members in the department of computer science for

77

their guidance encouragement, and insightful comments.

78

I thank my fellow coursemates, Mr. Oladejo Samuel Adetunji, Mr. Ayomide Feyi-Robinson, Mrs. Temilade Fashina, Mr. Kingsley

79

Efekodedo, Mr. Wasiu Olayinka and others numerous to mention. I also thank Mr. Oluseye Akinmoluwa and Mr. Faruq Ayoade for

80

their kind support on this project.

81

I thank my parents Rt. Rev'd Dr. S.G Kuponu and Mrs Victoria Kuponu for their guidance, parental advice and commitment instilled

82

in me towards learning. May you both live long in good health to enjoy and the reap fruits of your labour. My sincere thanks to my

83

friends and loved ones for the understanding, sacrifice, and prayers during my study. Also, to my siblings Mrs. Funmi Oni, Mrs.

84

Elizabeth Akinyanmi, Mrs. Esther Olowookere and Mrs. Kehinde Akinlabi for their continuous support and advice.

85

“Even though the above-mentioned institutions and persons have assisted in the process of this research work, I alone stand

86

responsible for the errors, if any, found in the work”

87

88

89

90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122

**Abstract**

This study addresses the crucial need for enhanced voice recognition systems in the realm of human-machine interfaces, particularly with a focus on accent identification algorithms, and their application in the context of Nigerian English speakers. The research aims to improve the accuracy and efficiency of speech recognition for these major Nigerian languages using KNN to increase the efficiency and accuracy of accent identification by using a trained data with the three major Nigerian language. Voiced audio samples of speakers from these tribes speaking English and various platforms such as news media and radio recordings was scrapped and recorded and extracted. The audio data was then preprocessed and transformed from the time domain to the frequency domain using the Fourier transform. Matlab R2015A was employed for model training, encompassing input reading, window size and hop size definition, and noise reduction techniques such as high-pass filtering and spectral subtraction. For feature extraction, Mel-Frequency Cepstral Coefficients (MFCC) were computed for each audio frame, subsequently aggregated to create fixed-length representations for each dialect sample for about sixty seconds in order to ensure uniformity in the inputs. The model underwent training with a classification algorithm KNN, followed by evaluation, which gave an accuracy rate of 84%. This result indicates that the model proficiently predicts the dialects within the context of English speech. The study's outcomes signify substantial progress in the development of an accent detection model tailored to the major Nigerian tribes: Yoruba, Hausa, and Igbo. The research is a significant stride toward more precise and effective voice recognition systems for Nigerian English speakers, contributing to the broader advancement of human-machine interfaces in an increasingly technology-driven world. It is recommended that future research explores alternative feature extraction techniques, particularly deep learning-based approaches capable of automatically learning relevant features from raw audio data.

**Keywords:** Accent, Accuracy, Algorithm, Dialect detection, Feature extraction, Fourier transform, Performance, Speech recognition, Voice recognition

**Word Count:** 298 Words

123

124

**Table of Contents**

125 Title page

126 Certification

127 ii

128 Dedication

iii

129 Acknowledgement

iv

130 Abstract

vi

131 Table of Content

vii

132 List of Tables

133 xiii

134 List of Figures

135 xiv

136 List of Acronyms

xv

137

138

139

140

141

142

143

144

145

*Do Not Copy, Lead City University, Nigeria*

146			
147			
148			
149			
150	<b>Chapter One: Introduction</b>		
151	Background to the Study		
152	1		
153	1.2 Statement of the Problem		7
154	1.3 Aim and Objectives of the Study		7
155	1.6 Significance of the Study		8
156	1.6 Scope of the Study		8
157	1.7 Operational Definition of Terms		9
158			
159	Endnotes	10	
160			
161	<b>Chapter Two: Literature Review</b>		
162	2.1. Conceptual Review		12
163			
164	2.1.1 Speech		
165	12		
166	2.1.1.1 Properties of Human Voice		14
167	2.1.2 Speech Recognition		15
168			

169	2.1.2.1	Accuracy of Speech Recognition	20
170	2.1.2.2	Speech Coding	21
171	2.1.2.3	Speech Synthesis	22
172	2.1.2.4	Speech Classifier	23
173	2.1.2.5	The Speech Recognition Process	23
174	2.1.3	Speech Recognition Models	29
175	2.1.3.1	Hidden Markov Models	29
176	2.1.3.2	Gaussian Mixture Models	29
177	2.1.3.3	Deep Neural Network Model	31
178	2.1.3.4	Language Models	32
179	2.1.4	Nigerian Major Languages	37
180	2.1.4.1	Hausa Language	37
181	2.1.4.2	Yorùbá Language	39
182	2.1.4.3	The Igbo Language	40
183	2.2	Methodological Review	41
184	2.2.1	Feature Extraction	42
185	2.3	Related Works	47
186	2.4	Summary of Gaps in Literature Reviewed	91
187		Endnotes	92
188			

189	<b>Chapter Three: Methodology</b>		
190	3.1.	Research Approach	105
191	3.2.	System Design	105
192	3.3	Requirement Specification	106
193	3.4	Research Method	108
194	3.4.1	Data Collection	108
195	3.4.2	Preprocessing Audio Data	109
196	3.4.3	Feature Extraction	110
197	3.4.4	Training/Testing	111
198	Endnotes	112	
199			
200			
201			
202			
203			
204	<b>Chapter Four: Results and Discussion</b>		
205	4.1	Result on Acquiring Speech Data	1154.2
206		Training the Dataset	116
207	4.2.1	Reading In the Voiced Input	116
208	4.2.2	Define Window Size and Hop size	117
209	4.2.3	Noise Reduction	117
210	4.2.4	Feature Extraction Using MFCC	118
211	4.3	Testing	
212		119	
213			

214	4.3.1	Classification	
215		119	
216			
217	4.4	Performance Evaluation	120
218			
219	4.5	Discussion of Results	123
220			
221	Endnote		124
222			
223	<b>Chapter Five: Conclusion</b>		
224	5.1	Summary of Results	
225		125	
226	5.2	Recommendations	
227		126	
228	5.3	Contribution to Knowledge	
229		129	
230	5.4	Suggestions for Further Research	130
231	Bibliography		
232		132	
233	Appendix		
234	Bio Data		
235	University Compliance Form		
236			
237			

238

**List of Tables**

239 **Table** **Title** **Page**

240 4.1 Performance Evaluation Table 119

241

242 4.2 Result of Precision, F - Score and Recall 120

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

*Do Not Copy, Lead City University, Nigeria*

**List of Figures**

264

265

266

**Figure**

**Title**

**Page**

267

2.1.

Speech Recognition Process

17

268

2.2.

A Three-State Hidden Markov Model

29

269

2.3

Representation of a Deep Neural Network

31

270

2.4

The Standard Orthographical Graphemes for Igbo

40

271

2.5

Components in an ASR System

41

272

3.1

Conceptual Model of the Proposed Design

103

273

3.2

Flowchart of the proposed Accent Classification Process

105

274

3.3.

Speech Data Recording Process Using Audacity

107

275

3.4

MFCC Block Diagram

108

276

3.5

Flowchart of K-Nearest Neighbors Algorithm

111

277

4.1

Training Model

114

278

279

4.2

Matlab Interface Showing Training Completed

118

280

4.3

Matlab Interface Showing Hausa Dialect

118

281

4.4

Confusion Matrix

120

282

283

284

285

286

287

288

289

290

**List of Acronyms**

291

292 AI- Artificial Intelligence

293 AM- Acoustic Model

294 ASR- Automatic Speech Recognition

295 CMS- Cepstral Mean Subtraction

296 DNN- Deep Neural Networks

297 DTW- Dynamic Time Warping

298 GMM- Gaussian Mixture Model

299 HMM- Hidden Markov Model

300 KNN - K-Nearest Neighbour

301 LPC- Linear Predictive Coding

302 LRE- Language Resources and Evaluation

303 MFCC- Mel Frequency Cepstral Coefficient

304 ML- Machine Learning

305 MLP- Multi-Layer Perceptron

306 NLP- Natural Language Processing

307 RASTA-PLP- Relative Spectral Transform-Perceptual Linear Prediction

308	SI-ASR-	Speaker-Independent Automated Speech Recognition
309	SVM-	Support Vector Machine
310	VQ-	Vector Quantization

*Do Not Copy, Lead City University, Nigeria*