OBAFEMI AWOLOWO UNIVERSITY,

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,

ILE-IFE, OSUN STATE, NIGERIA.

LABORATORY ONE

(Language and Grammar Modelling)

By

Group 18 Members

A CPE 510 project work in partial fulfilment of the requirements for the award of Bachelor's degree in the Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife.

Under the Supervision of

Prof. O. A Odejobi

NOVEMBER 2019

GROUP 18 MEMBERS

1.	Sholesi Kofoworola Victoria	CSC/2014/080	Computer Science & Maths
2.	Taiwo Oluwatomisin Oluwafemi	CSC/2014/081	Computer Engineering
3.	Tijani Yetunde	CSC/2015/160	Computer Science & Econs
4.	Talabi Oluwasegun	CSC/2014/082	Computer Engineering
5.	Solademi Francis	CSC/2015/158	Computer Science & Maths
6.	Utulor Ugonna	CSC/2014/084	Computer Science & Econs

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF NOTATION AND TERMS

1. Ajami – A form of Arabic script in which Yoruba is written.

INTRODUCTION

The tasks in this project are based on Human Language Processing which is a branch of Artificial Intelligence. It is concerned with the phenomenon of human language; the native instrument that humans use to communicate within the self and with others.

Definition of Human Language Processing (HLP)

Human Language Processing is a computational study of human concept formulation and expression instrument with the aim to create autonomous material agents that mimic, in part or whole, the underpinning phenomenon.

The object of study in HLP is human languages in all its forms of manifestations as well as the phenomenon underlying it, including the mechanism of the organs responsible for language.

The kind of study here is computational-i.e. creation and manipulation of symbols.

The purpose of the study is to create machines and/or system that mimics human understanding of human languages.

The development of Human Language Processing Systems can be done in six (6) steps which is duly followed in the process of carrying out the various tasks in this experiments. They are:

1. Understand the problem
2. State Assumptions
3. Behaviour Analysis
4. System Design
5. Implementation
6. Evaluation

PROBLEM STATEMENT

LABORATORY 7

7.0.1 Task1

Select an indigenous African language and provide the following:

a. A brief description of:

(i) The native speakers

(ii) Where the language is spoken

(iii) The language forms of expression

b. Its orthography- i.e. its writing-reading system

c. Select any of the following domains {Market, School, Natural environment and landscape, Farm, Home, Health} and list a lexicon comprising:

1. Twenty-five (25) Nouns

2. Ten (10) Verbs

3. Five (5) Preposition

4. Five (5) Adjectives

5. Ten (10) Loanwords.

d. The structure of a simple grammar for the language.

e. Based on d., generate Fifteen (15) sentences using the results of b. and c.

7.0.2 Task2

a. Generate the British English language gloss for items in 1c.

b. Write the structure of a simple grammar for English.

c. Generate a translation for sentences in 1e based on a. using the result of b. in Task 2.

7.0.3 Task3

1. List and discuss the similarities and differences that you observed between your data in Task 1 and those in Task 2.

2. Reflect on your findings in 1, and provide informed advice for the development of a machine translation system that translates:

(i) From the indigenous African language to the British English language.

(ii) From the British English language to indigenous African language.

LABORATORY 8

1. Identify, extract and observe the fundamental frequency F0 in the speech signals of the first and last syllables in each word. Also observe the pattern of the first two formants, i.e. F1 and F2
2. . Identify, extract and observe the pattern of the third and fourth formants, i.e.

   F3 and F4.

3. Record at least two (2) isolated syllables that comprises any of the words in item 2 above and discuss the features of the F0 vis-a-vis the one in the word sample. HINT: study the beginning, middle and end of the F0 waveform.
4. Document experiments 1 to 6 as well as your reflections on your observations.

LABORATORY 9

9.2.3 Task 1

Figure 9.1 is the screenshot of JFLAP representation and processing of grammar. The grammar in this case is meant to model the expression $\{Y = a^n b^n \mid n \geq 1\}$. This expression can be interpreted as follows: “one or more a’s followed by exactly the same number of b’s”. The grammar for this expression is specified as a four (4) tuple $G_0 = <\Sigma, V, P, S>$. Where:

1. $\Sigma = \{a, b\}$

2. $V = \{S, A, B\}$

3. P: $S \rightarrow ASB$

   $A \rightarrow a$

   $B \rightarrow b$

   $S \rightarrow \lambda$

4. S the string start symbol.

9.2.4 Task 2

1. Analyze and discuss the grammar G_0 of this language using the rail-road diagram.

2. Using the JFLAP platform, implement the grammar G_0 above.

a. Test your model by generating the parse trees for the following expressions:

(i) a^3b^3

(ii) a^4b^4

(iii) a^6b^6

b. What did your system produce for the expression a^2b^3? And why?

9.2.5 Task 3

Repeat the experiment in Task 1 using the following:

(i.) $G_1 = <\Sigma, V, P, S>$.

Where: • $\Sigma = \{a, b\}$

• $V = \{S\}$

• P: $S \rightarrow aSa$

$S \rightarrow bSa$

$S \rightarrow \lambda$

• S the string start symbol.

(ii.) Discuss the language generated by the grammar G_1 (if any)

G2 = <Σ, V, P, S>.

Where:

• $\Sigma = \{a, b\}$

• $V = \{S\}$

• P: $S \rightarrow aSbb$

$S \rightarrow abb$

• S the string start symbol.

(iii.) Discuss the language generated by the grammar G_2 (if any)

G3 = <Σ, V, P, S>.

Where:

• $\Sigma = \{a, b\}$

• $V = \{S, A\}$

• P: $S \rightarrow As$

$S \rightarrow A$

$A \rightarrow ab$

$A \rightarrow aAb$

• S the string start symbol.

(iv). Discuss the language generated by the grammar G_3 (if any)

G4 = <Σ, V, P, S>.

Where:

• $\Sigma = \{a, b\}$

• $V = \{S, B\}$

• P: $S \rightarrow Sb$

$S \rightarrow Bb$

$B \rightarrow aBb$

$B \rightarrow ab$

- S the string start symbol.

9.2.6 Task 4

Design the grammar for a language that models the generation of strings in the base 6 number system. Test your system with at least six valid and six invalid string and discuss the language of your system.

LABORATORY 10

10.3.1 Task1

1. Using the above grammar, analyze and discuss any twelve (12) English sentences using: (i) the parse trees OR (ii) Rail Road diagram. Your sentence should have the following features:

(a) Composed from the vocabulary generated in Laboratory I.

(b) Each sentence should have not more than six words.

(c) Use declarative sentences only.

2. Using the NLTK grammar tool in Python, explore the English language with respect to correctness of the statements generated.

10.3.2 Task2

1. Based on the indigenous African language selected in Laboratory I, discuss the grammar for generating corresponding statements to that in Table 10.1.

2. Repeat the tasks you executed for the English language using the indigenous language selected.

3. Illustrate a situation of ambiguity and suggest how it can be managed. 4. Discuss your observation and reflections on the translation process.

10.3.3 Task3

Using the Python programming language (you could use NLTK toolkit):

1. Develop a software for checking the correctness of English statements, based on the grammars defined above.

2. Develop a software for checking the correctness of indigenous language selected statements, based on the grammars defined above.

3. Test your system with at least ten examples of correct and incorrect grammatical statements. Your evaluation should be limited to the database generated in Laboratory I. Observe and document the kind of sentences that your system will fail to correctly it grammar.

4. Discussyourobservationandreflectionsonthetranslationsystemdevelopment process.

LABORATORY 11

Task 1

Using the Python programming language:

1. Develop a Dialogue and Action System to interact with Ap´alar´ a in its world. (HINT: Follow the steps in Section 1.1).

2. Test your system extensively and discuss its limitations.

3. Discuss the language of the machine that models Ap´alar´ a's behaviour.

4. Discuss your reflection on the above exercises by looking into specific things that Ap´alar´ a does very well, how you can improve on its performance as well as the things you hope Ap´alar´ a should be able, but not able, to do.

5. Based on your reflections on the above experiment, design a grammar for the language of Ap´alar´

OBJECTIVES OF EXPERIMENT

General Objective(s) of the laboratory work

1. To help understand the general concepts and principles discussed in Human Language Processing.

2. To help guide private studies, by allowing us to explore specific concepts in the subject matter of Human Language Processing (HLP) techniques and their applications.
3. To help improve proficiency in HLP systems development and computational problem solving in general.

Specific Objective(s) of the laboratory work

Laboratory 7 is to expose us to translation and grammar from English Language to an Indigenous Language and verse versa and to note the difference in their structure.

Laboratory 8 is to expose us to different frequencies when speech are generated in order to get notable difference between each word and person

Laboratory 9 is to expose us to grammar generation and identification using the tool JFLAP

Laboratory 10 is to expose is to grammar correctness and differences using the NLTK Python library.

Laboratory 11 is to expose us to the practical aspect of artificial intelligence by building a Robotic arm.

EXPERIMENT PROCEDURES

Procedures for Laboratory 7

Materials and tools

Data collected from different individuals.

Experiment Processes

The indigenous language selected is the Yoruba language.

Task 1

a. The Yoruba People are one of the largest African Ethnic group in the Sahara Desert concentrated in the south western part of Nigeria. Yoruba mythology holds that all Yoruba people descended from a hero called Oduduwa. The Yoruba have shared a common language and culture for centuries but were probably never a single political unit. Their towns became densely populated and eventually grew into the present-day cities of Oyo, Ile-Ife, Ilesha, Ibadan, Ilorin, Ijebu-Ode, Ikere-Ekiti, and others.

b. Orthography: In the 17th century, Yoruba was written in the Ajami script, a form of Arabic script. Modern Yoruba orthography originated in the early work of Church Mission Society missionaries working among the Yoruba of Freetown.

The Yoruba has two alphabets which are classified into tone and phone.

The phone consists of 25 letters which are:

A B D E Ẹ F G Gb I J K L M N O Ọ P R S Ṣ T U W Y

a b d e ẹ f g gb i j k l m n o ọ p r s ṣ t u w y

The tone consists of High, Middle and Low tone marks.

c. Domain selected is Home

1. Twenty-five Nouns
    i. Ile
    ii. Omokunrin

iii. pepe

iv. Ibi idana

v. Baluwe

vi. Ife

vii. Awo

viii. Ibusun

ix. Ikowe

x. Aga

xi. Eni

xii. Igbale

xiii. Ferese

xiv. Orule

xv. Ose

xvi. Apo

xvii. Ikoko

xviii. Iwe

xix. Kokoro

xx. Aso

2. Ten (10) Verbs

i. Wa

ii. Lo

iii. Ti

iv. Gbe

v. Wo

vi. Ra

vii. Sun

viii. Se

ix. Fo

x. Gba

3. Five (5) Prepositions

i. Abe

ii. Ori

iii. Fun

iv. Ni

v. Ninu

4. Five (5) Adjectives

i. Tuntun

ii. Nla

iii. Dudu

iv. Kekere

v. Tutu

5. Ten (10) Loan words

i. Komputa

ii. Tabili

iii. Adiresi

iv. Foonu

v. Gilasi

vi. Sinima

vii. Fiimu

viii. Dokita

ix. Redio

x. Noosi

d. The simple grammar structure for the Yoruba language is SVO.

Where S is the subject (Performer of an action)

V is the verb (The action)

O is the object (The thing that suffers the action)

e. Sentences

i. Ade joko sori Aga naa.

ii. Kofo wo aso dudu.

iii. Ajakaye fo gbogbo aso inu ile.

iv. Lawale gba oju ferese bota.

v. Mama Ojo sun sori eni.

vi. Ugo ti fo ikoko kerere naa.

vii. Yetunde gbe apo losi oja.

viii. Solademi lo ri dokita.

ix. Tomisin ti fo gbogbo awo and ife inu ile.

x. Segun ti ra Komputa tuntun.

Task 2

a. British English Language Gloss

1. Twenty-five (25) Nouns

    i. House

    ii. Boy

    iii. Shelf

    iv. Kitchen

    v. Bathroom

    vi. Cup

    vii. Plate

    viii. Bed

    ix. Pen

    x. Chair

    xi. Mat

    xii. Broom

    xiii. Window

    xiv. Roof

    xv. Soap

    xvi. Bag

    xvii. Pot

    xviii. Book

    xix. Key

    xx. Clothes

2. Ten (10) Verbs

i. Come

ii. Go

iii. Lock

iv. Carry

v. Enter

vi. Buy

vii. Sleep

viii. Cook

ix. Wash

x. Take

3. Five (5) Prepositions

i. Under

ii. Over

iii. For

iv. At

v. in

4. Five (5) Adjectives

i. New

ii. Big

iii. Black

iv. Small

v. Cold

5. Ten (10) Loan words

i. Computer

ii. Table

iii. Address

iv. Phone

v. Glass

vi. Cinema

vii. Film

viii. Doctor

ix. Radio

x. Nurse

b. The simple grammar structure for the English language is SVO.

Where S is the subject (Performer of an action)

V is the verb (The action)

O is the object (The thing that suffers the action)

c. <u>Sentences</u>

i. Ade sat on the head of the chair.

ii. Kofo wore cloth black.

iii. Ajakaye washes all the clothes inside the house.

iv. Lawale passes the eye of the window outside.

v. Mama Ojo slept on the head of the mat.

vi. Ugo has broken the pot small.

vii. Yetunde carried the bag to the market.

viii. Solademi went to see the doctor.

ix. Tomisin has broken all the plates and cups inside the house.

x. Segun has bought computer.

Task 3

1. Similarities and Differences

a. Similarities between Data in Task 1 and Task 2

i. They both have the same grammatical structure which is SVO.

ii. The prepositions still maintain the same position during translation.

b. Difference between Date in Task 1 and Task 2

i. Task 1 has tonal alphabets while Task 2 does not.

ii. The adjectives in Task 1 come after the object or subject while the adjectives in Task 2 come before the subject or object.

iii. The verbs in Task 1 could have different translations when translated to verbs in Task 2.

iv. Task 1 sentences are more prone to sentential ambiguity than Task 2 sentences.

v. Identifiers such as ‘a’ and ‘the’ are not properly cared for in Task 1.

2. Advise for the development of machine translation system

a. From Yoruba Language to English Language

i. Identifiers should be properly cared for during translation

ii. Adjectives should also be monitored that they are placed in the right places.

iii. Take note of the context of the sentence to avoid sentential ambiguity.

b. From English Language to Yoruba Language

i. Translation should not be verbatim when translating from English to Yoruba.

ii. The position of the adjectives should also be considered.

iii. Take into consideration the tonal alphabets of the Yoruba language.

Procedure for Laboratory 8

Materials and tools

Praat Software

Experiment Processes

APO (female)

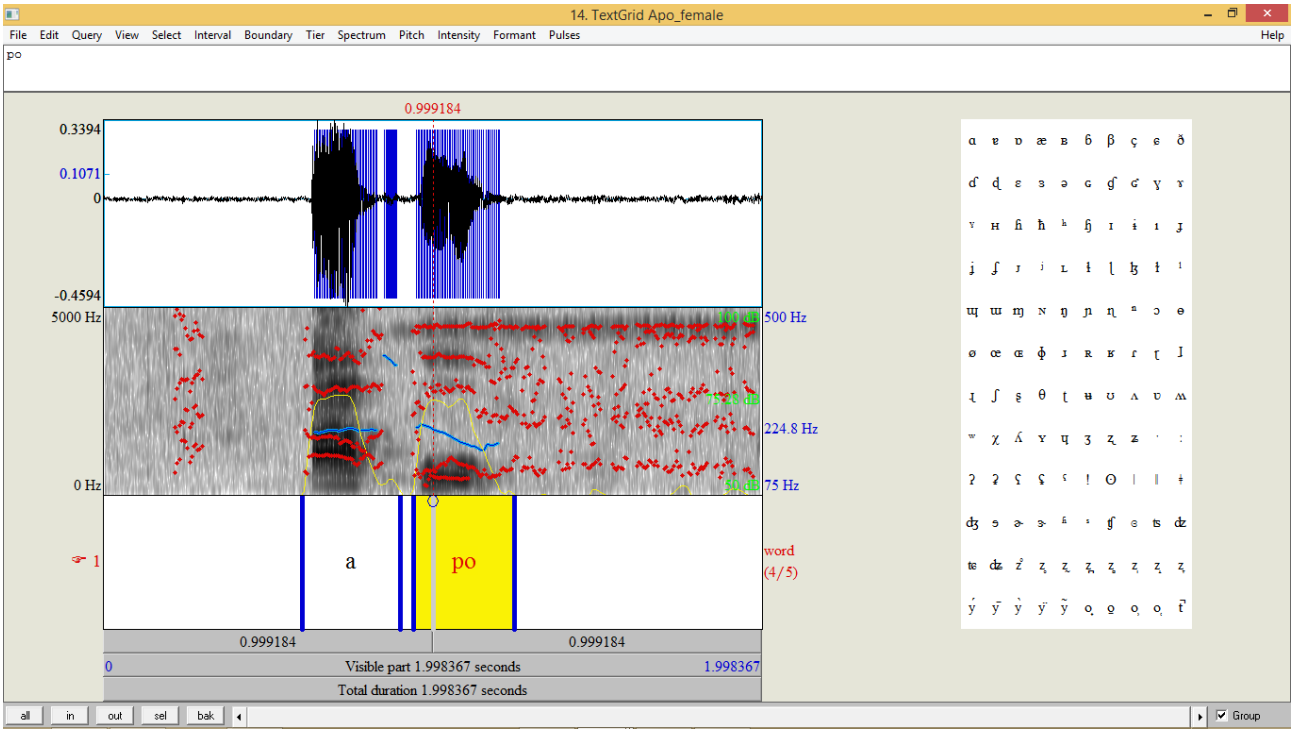

Figure 8.1 Analysis for Apo (Female)

F_0 for the first vowel– 222.2Hz	F_0 for the second vowel - 190.3Hz
F_1 – 470.81Hz	
F_2 – 2113.68Hz	
F_3 – 2824.96Hz	
F_4 – 4242.99Hz	

Table 1: Analysis for Apo (Female)

APO (male)

F_0 for the first vowel– 109.6Hz	F_0 for the second vowel – 111.6Hz
F_1 – 893.99Hz	
F_2 – 1180.71Hz	
F_3 – 2896.92Hz	
F_4 – 3584.78Hz	

Table 2: Analysis for Apo (Male)

AWO (Female)

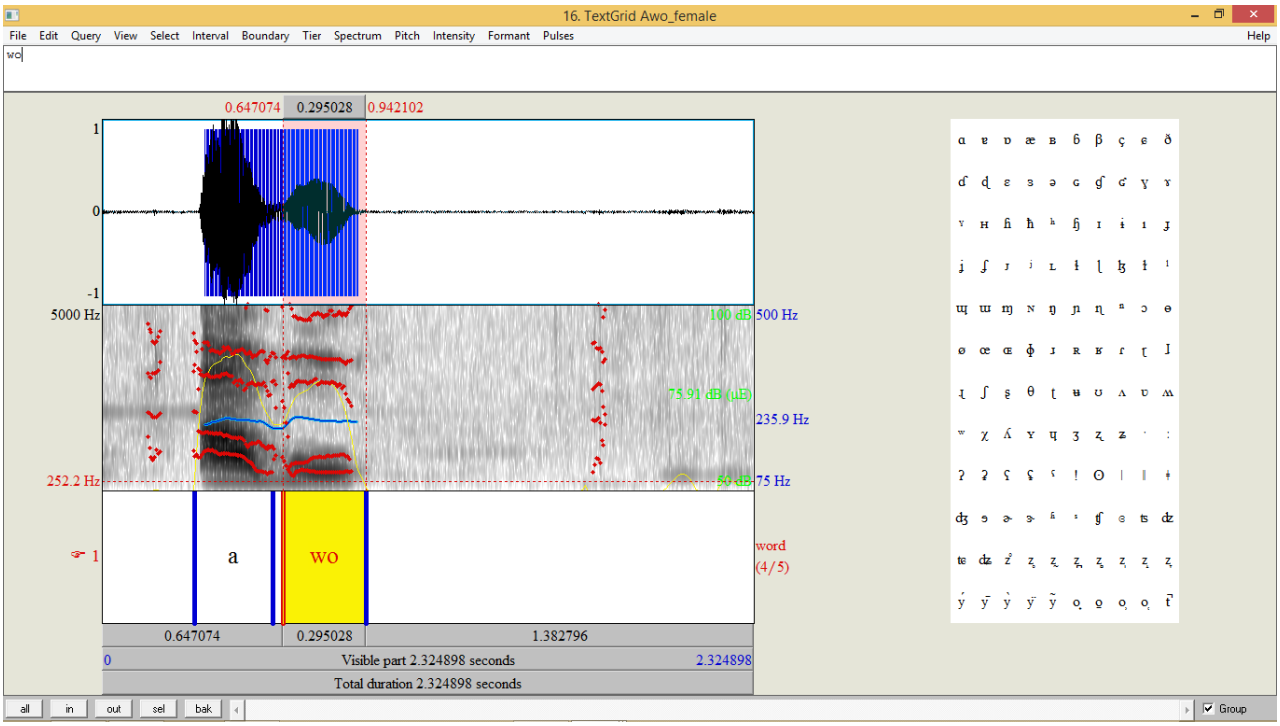

Figure 8.3 Analysis for Awo (Female)

F_0 for the first vowel– 1801Hz	F_0 for the second vowel – 232.5Hz
F_1 – 530.11Hz	
F_2 – 1918.92Hz	
F_3 – 2665.45Hz	
F_4 – 3447.50Hz	

Table 3: Analysis for Awo (Female)

Awo (Male)

F_0 for the first vowel– 646.9Hz	F_0 for the second vowel – 131.6Hz
F_1 – 335.84Hz	
F_2 – 1820.59Hz	
F_3 – 3101.37Hz	
F_4 – 4275.94Hz	

Table 4: Analysis for Awo (Male)

Eni (Female)

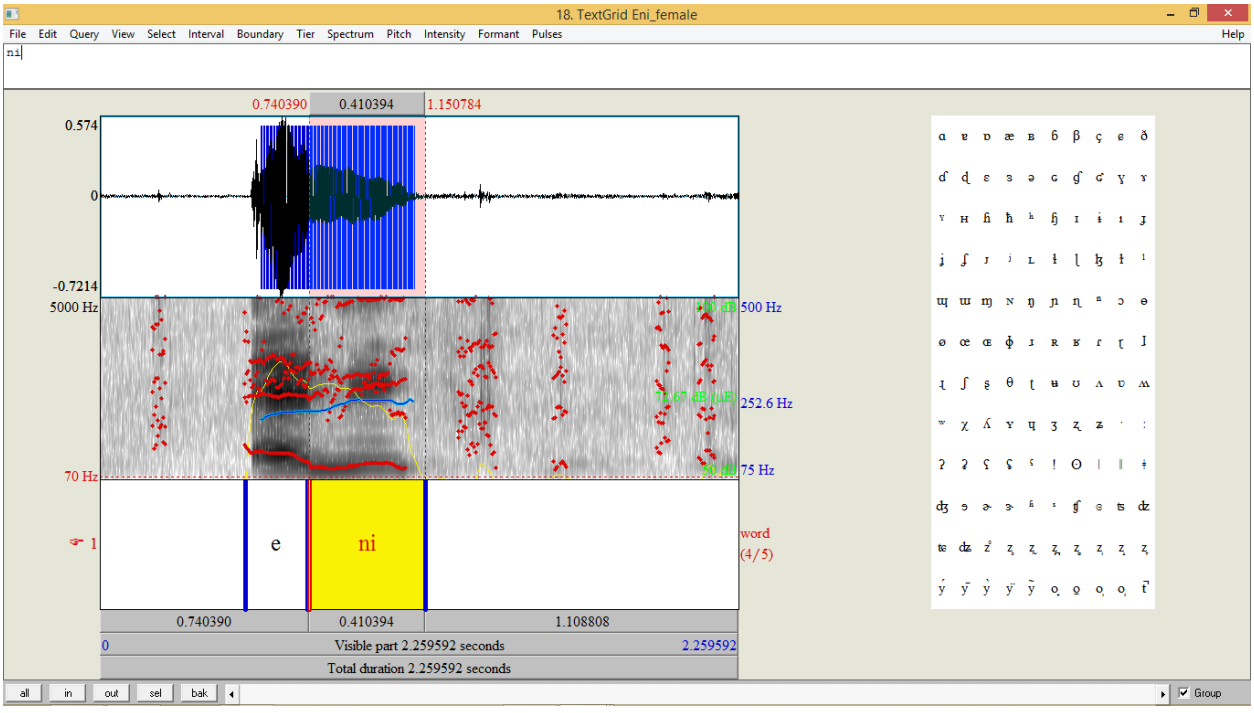

Figure 8.5 Analysis for Eni (Female)

F_0 for the first vowel– 1678Hz	F_0 for the second vowel – 258.2Hz
F_1 – 384.75Hz	
F_2 – 2179.73Hz	
F_3 – 2759.34Hz	
F_4 – 3873.64Hz	

Table 5: Analysis for Eni (Female)

Eni (Male)

F_0 for the first vowel– 677.3Hz	F_0 for the second vowel – 147.5Hz
F_1 – 429.46Hz	
F_2 – 1561.30Hz	
F_3 – 2978.43Hz	
F_4 – 4101.47Hz	

Table 6: Analysis for Eni (Male)

Ife (Female)

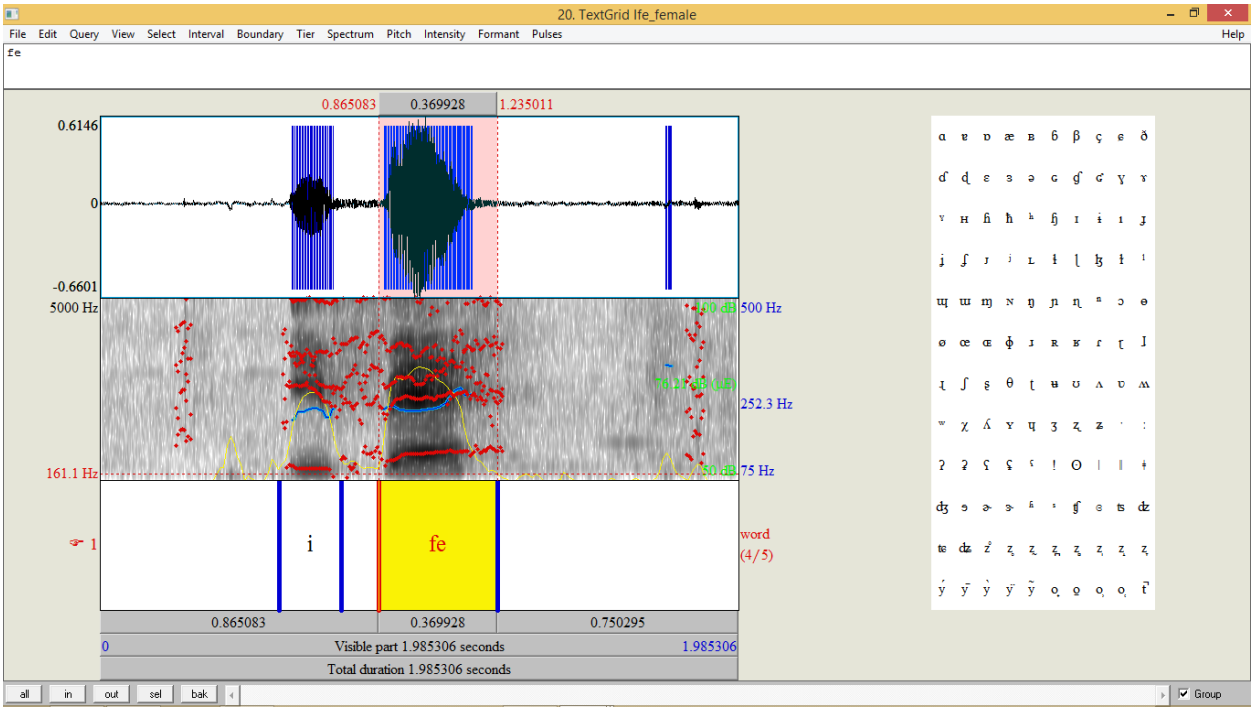

Figure 8.7 Analysis for Ife(Female)

F_0 for the first vowel– 1679Hz	F_0 for the second vowel – 275.3Hz
F_1 – 734.31Hz	
F_2 – 2196.49Hz	
F_3 – 2822.29Hz	
F_4 – 3633.49Hz	

Table 7: Analysis for Ife (Female)

Ife (Male)

F_0 for the first vowel– 495.1Hz	F_0 for the second vowel – 157Hz
F_1 – 566.82Hz	
F_2 – 1829.90Hz	
F_3 – 2836.23Hz	
F_4 – 3601.06Hz	

Table 8: Analysis for Ife (Male)

Ile (Female)

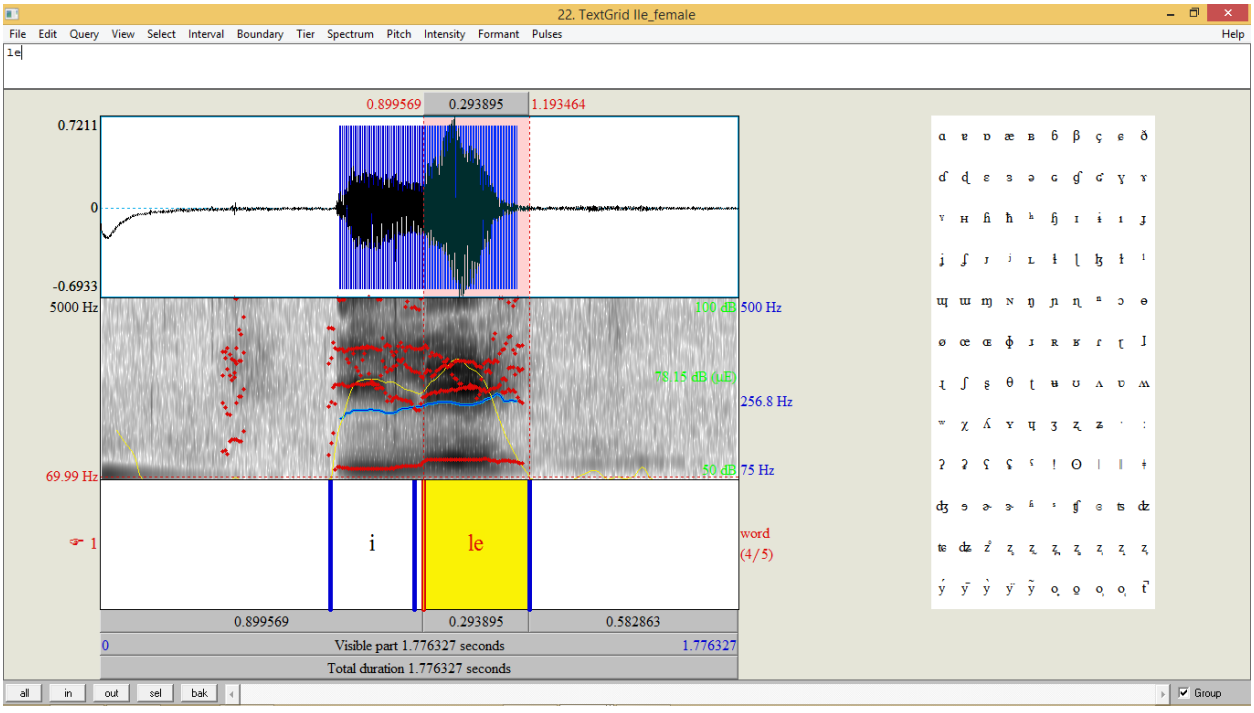

Figure 8.9 Analysis for Ile (Female)

F_0 for the first vowel– 1701Hz	F_0 for the second vowel – 257.7Hz
F_1 – 505.50Hz	
F_2 – 2369.07Hz	
F_3 – 2844.37Hz	
F_4 – 3446.66Hz	

Table 9: Analysis for Ile (Female)

Ile (Male)

F_0 for the first vowel– 889.8Hz	F_0 for the second vowel – 149.7Hz
F_1 – 384.74Hz	
F_2 – 2147.60Hz	
F_3 – 3024.61Hz	
F_4 – 3655.99Hz	

Table 10: Analysis for Ile (Male)