Research paper- Regional Hindi Phonetics

A PROJECT REPORT

Submitted by

Stuti Yadav(21BCE10543)

Pritha Dey(21BCE10707)

Sajal Saxena(21BCE10515)

Shashwat Pandey(21BCE11596)

Priyanshu Dubey(21BCE11560)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

VIT®

BHOPAL

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRI KALAN, SEHORE

MADHYA PRADESH - 466114

August 2022

VIT BHOPAL UNIVERSITY,KOTHRI KALAN, SEHORE MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled "**Research paper- Regional Hindi Phonetics**" is the bonafide work of "Stuti Yadav(21BCE10543), Pritha Dey(21BCE10707), Sajal Saxena(21BCE10515),Shashwat Pandey(21BCE11596) Priyanshu Dubey(21BCE11560)" who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR
Dr. Sandip Mal
School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY
UNIVERSITY

PROJECT GUIDE
Dr. Nirmala E.
School of Computer Science and Engineering
VIT BHOPAL

The Project Exhibition I Examination is held on

ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr. Sandip Mal, Head of the Department, School of Computer Science and Engineering for much of his valuable support and encouragement in carrying out this work.

I would like to thank my internal guide Mrs. Dr. Nirmala E., for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.
I would like to thank all the technical and teaching staff of the School of Aeronautical Science, who extended directly or indirectly all support.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

LIST OF ABBREVIATIONS

1.IPA-International Phonetic Alphabet-an alphabet developed in the 19th century to accurately represent the pronunciation of languages.

2.NLP- Natural Language Processing

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Introduction

This article describes the observation through the experiment conducted to compare the results derived from accumulated data(different tokens) and reference journal, in Hindi language, to prove that pronunciation of various vowels has changed accordingly,

and that the originally published journal is no more accurate and the new variations need to be added in the NLP system.

NLP i.e. Natural Language Processing is a branch of Artificial Intelligence under Computer Science, concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text or voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage exceptions, variations in sentence structure—these just a few of the irregularities of human language that take humans years to learn, but that programmers must teach natural language-driven applications to recognize and understand accurately from the start, if those applications are going to be useful.

Hence, NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

– Speech recognition

– Part of speech tagging

– Word sense disambiguation

– Coreference resolution

– Sentiment Analysis

– Natural language generation.

Motivation for the work

As we travel to different parts of India, we get to see that the accent of people, while speaking a specific language (here, Hindi) changes, as it is influenced by their mother tongue, especially if they are multilingual. The difference in linguistics is something that says a lot about the diversity of the way the word is spoken.Although the same word it could be, different background of life makes a slight difference in their accents.

This causes inaccuracy of results in speech to text or similar applications like Alexa or Google Assistant.

Hence, our motive is to find standard variants and deviations of that language, specifically the vowels, and later adding it to NLP system for more accuracy in speech related NLP tasks.

This will lead AI to recognize the words spoken more easily irrespective of the accent to a great extent of accuracy.

Software Implemented: PRAAT

PRAAT is a free computer software package for speech analysis, synthesis and manipulation in phonetics.[4] It was designed, and continues to be developed, by Paul Boersma and David Weenink of the University of Amsterdam. It can run on a wide range of operating systems, including various versions of Unix, Linux, Mac and Microsoft Windows (2000, XP, Vista, 7, 8, 10). The program supports speech synthesis, including articulatory synthesis.

– What is Praat used for?

Praat can read sounds recorded with the program or audio files(mp3 and wav types are commonly recommended) recorded in another way. Once the application is launched, it will generate a graph of waves that indicate intonation, intensity, volume, and other complex details. Proving its handy purpose for deeply learning linguistics, it is able to isolate certain sounds or filter frequencies either manually or using scripts.

You will have access to spectrograms—a visual representation of the sound changes over time. It permits you to produce speech from a pitch curve and filters or from various muscle activities. Furthermore, it grants you the ability to alter existing speech utterances wherein you can customize the pitch, intensity, and duration of the speech.

Here, you have the capacity to custom-label your sample using the IPA. Not only that, you can even annotate your sound segments based on the specific variable you are aiming to examine. It also supports multi-language text-to-speech features that empower you to section the sound into words and phonemes. It's difficult to get to grips with, though. An extensive manual is available but it's aimed mainly at linguistic experts.

Problem Statement

Problem: Insufficient data in NLP regarding variations in values of terms related to phonetics in accents of a language.

Background: As we come from multicultural backgrounds, it is noticed that the pattern of spelling certain words are different for different people. However, the AI is sometimes unable to recognize the words due to this existing difference.

Relevance: Considering a situation, if we spell “kali” /kʌli/ to Google assistant, it considers it as “kaali” /ka:li/. This creates inaccuracy of results especially while using speech recognition systems, like Google Assistant used above.

Objective: This research aims to prove that necessary variations are required in the values of phonetic terms, that NLP has, to increase the accuracy of any AI thus making Human and machine co-ordination and communication much better.

Experimental Data

For this research we have collected data of 8 students from different regions of India present in the campus itself and performed Vowel Formant Analysis on their voices on the specified

sentences in Hindi and analyzed how their voices are different from the originally specified vocal formant frequency.

We referred to Ladefoged(2006) for the theory and positional analysis of vowels as shown in the chart given below which depicts IPA vowel characteristics.

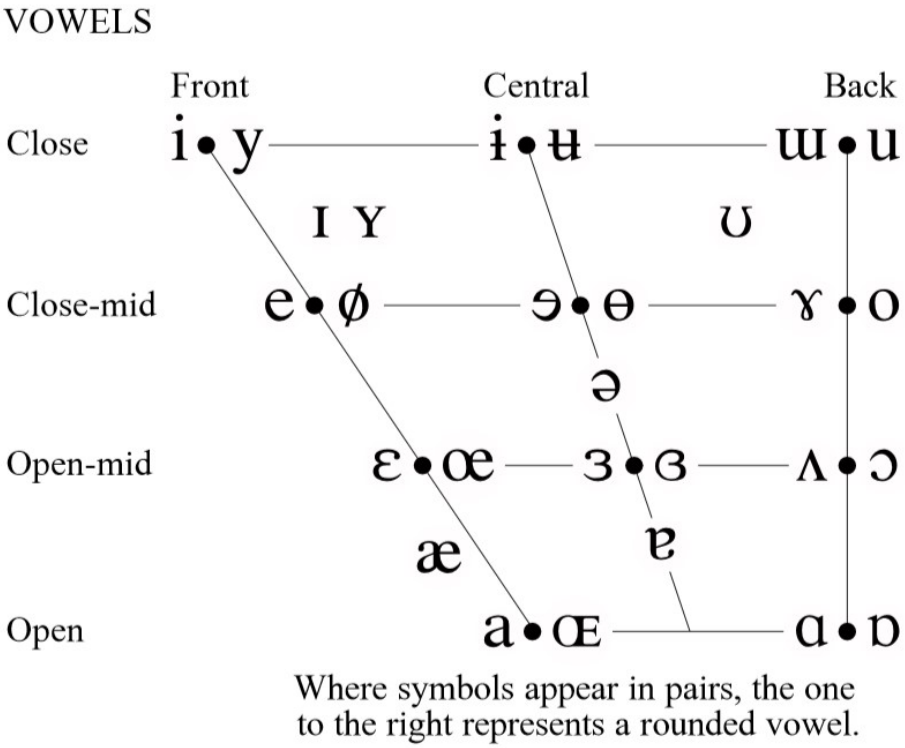

Fig.1

In the sentences used for analysis we have included the vowels from all 3 front, central and back vowels placement.

The following 5 sentence are as follows:

1.**काश केश** होते।

2.**बात बीत** गई।

3.**काली कली** कोमल है।

4.**तेल तोल** लेना।

5.**तिल ताल** के पास है।

So the derived vowels that we considered for the evaluation are in the CVC form that is Consonant Vowel Consonant form. That are अ (ʌ), आ(a:), इ(I), ई(i:), ए(e), ओ (o).

After recording the 40 samples we analyzed each of the samples on PRAAT software and found the F1 And F2 formants.

The first two formants are important in determining the quality of vowels, and are frequently said to correspond to the open/close (or low/high) and front/back dimensions (which have traditionally been associated with the shape and position of the tongue). Thus the first formant F_1 has a higher frequency for an open or low vowel such as [a] and a lower frequency for a closed or high vowel such as [i] or [u]; and the second formant F_2 has a higher frequency for a front vowel such as [i] and a lower frequency for a back vowel such as [u].

Considering the example:

काली कली कोमल है

Here we are checking the Formants of the vowels /a:/ and /ʌ/ that are present in the words काली /ka:li/ and कली /kʌli/ respectively.

Initially, the sentence is recorded and uploaded on PRAAT software in .mp3 or .wav format. Then, every word of the sentence is separated by boundaries created using a text grid.

Below is the image representing the process.

Red dots represent the formants.

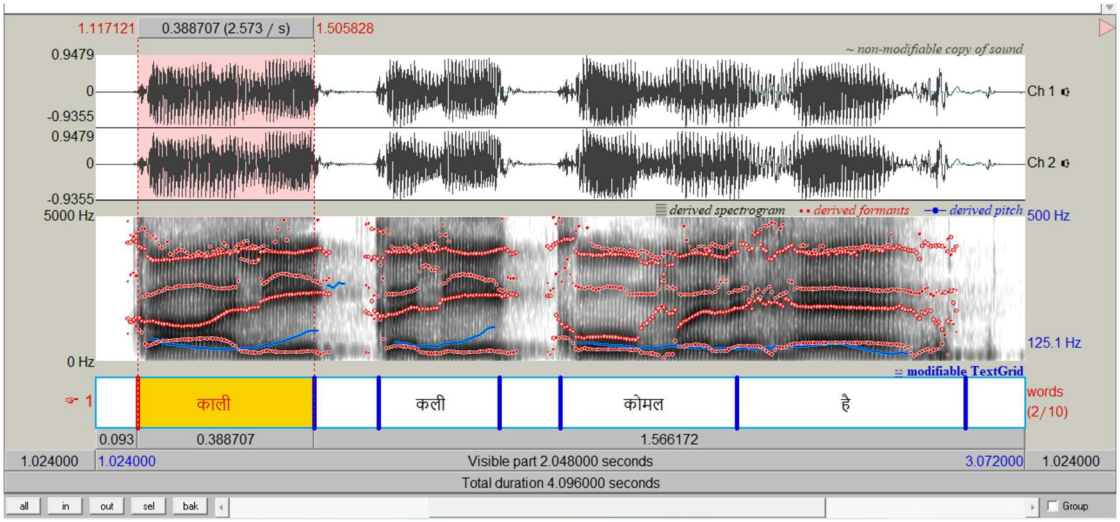

Fig.2

Now, we select the first word, काली and then check the formants of the segment that spells /a:/.

Once the segment is selected, we look for the lowermost dark patch in the segment. This lowermost dark patch represents **F1**.

To check the F1 of vowel we select the point that is stable across the larger part in the selected area.

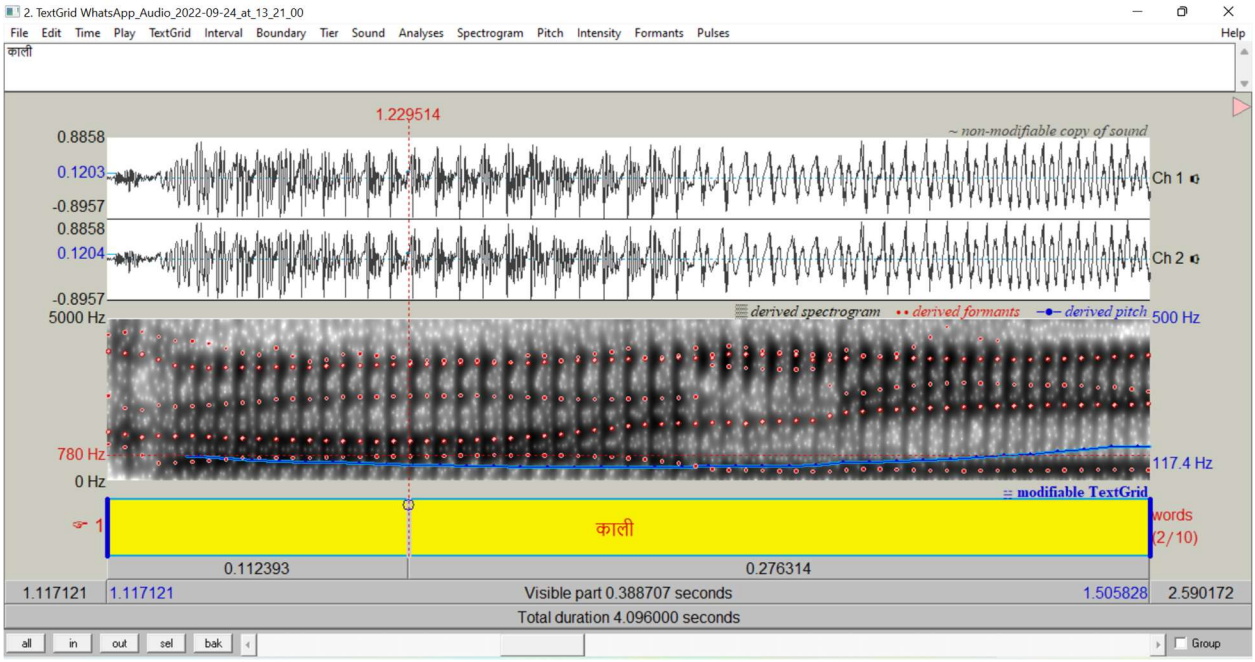

Fig.3

F1: 780 Hz

The second-most dark patch is that of **F2**. The value of F2 is taken by considering the highest value among all the formant dots. But we need to make sure that the renge is taken in the mid region such that consonants do not affect the sound. While for diphthongs, we consider the starting section of formant.

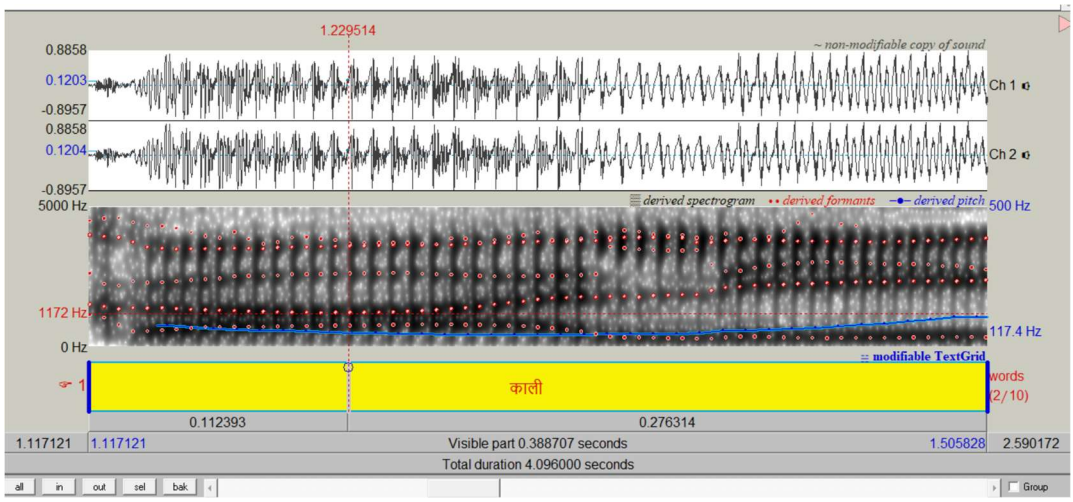

Fig.4

F2: 1172 Hz

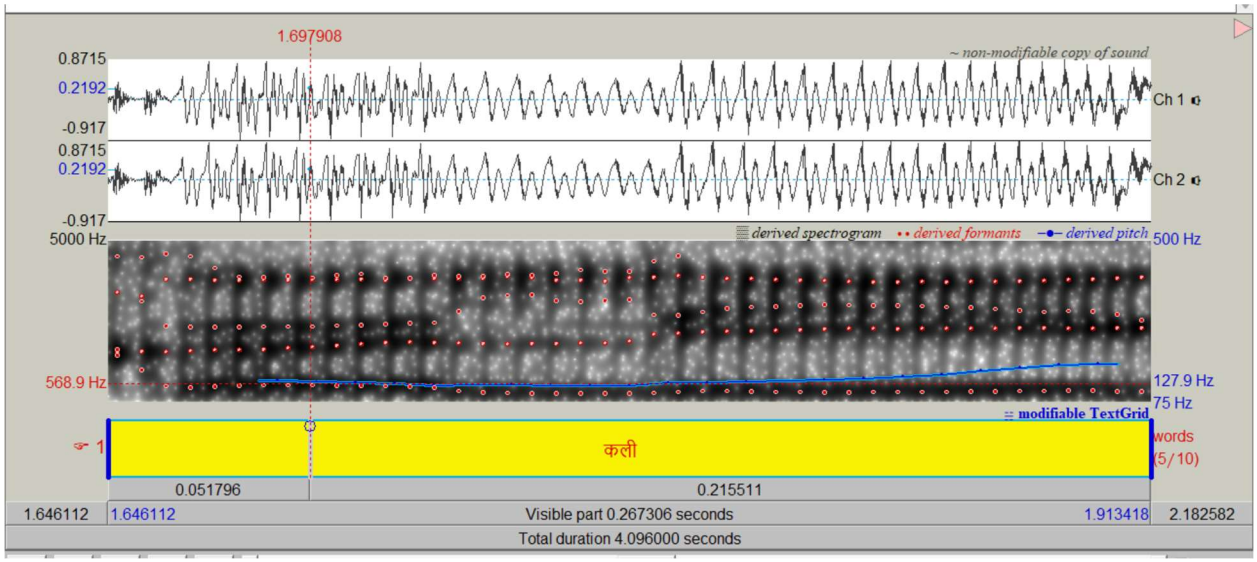

Fig.5

F1: 568.9 hz

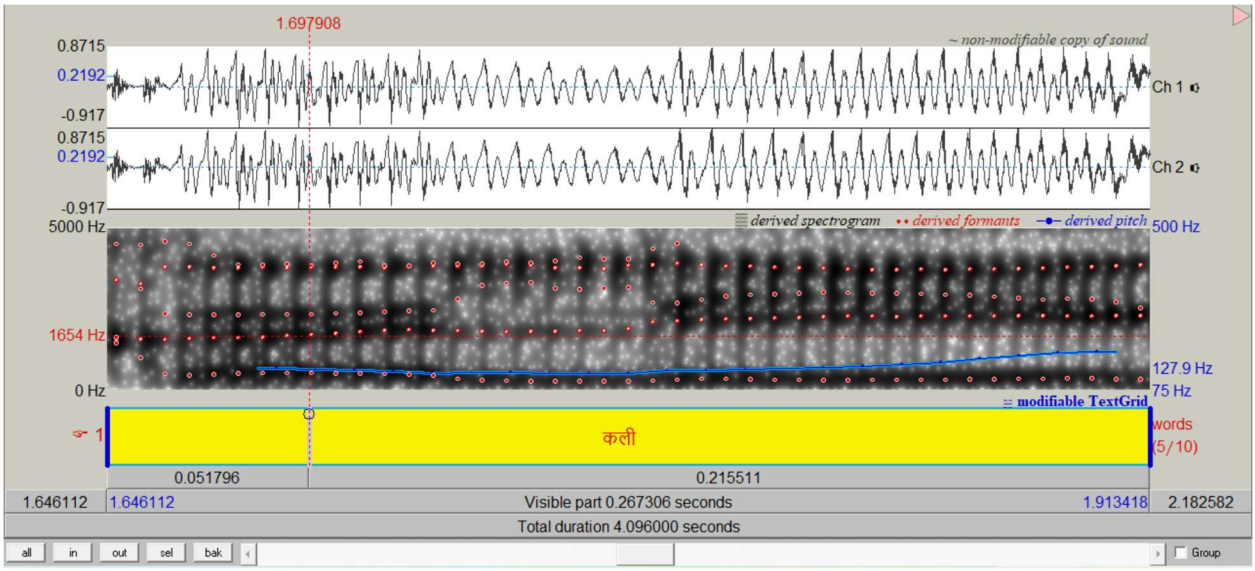

Fig.6
F2: 1654 hz

Similarly, we can analyze other sentences.

बात बीत गई

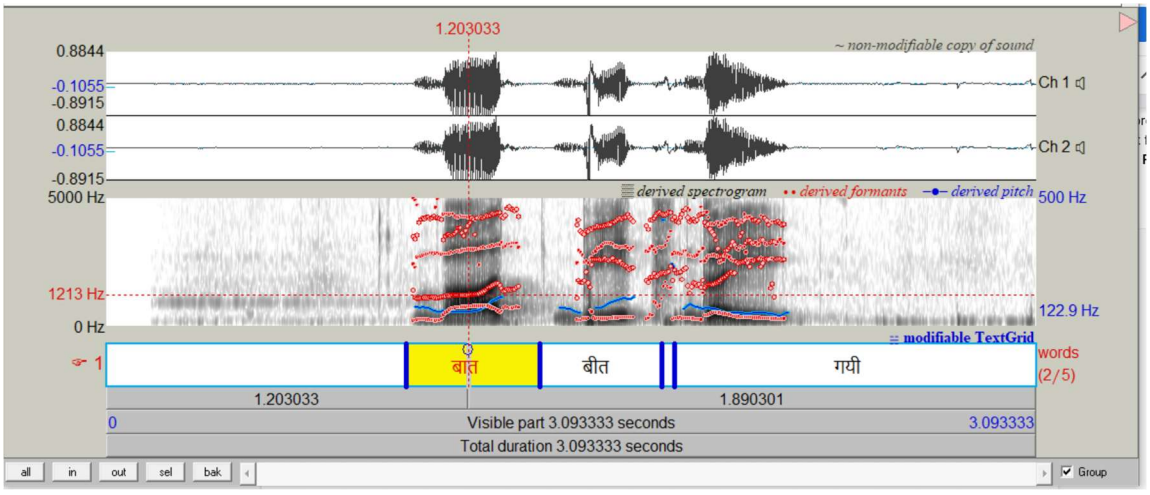

Fig.7

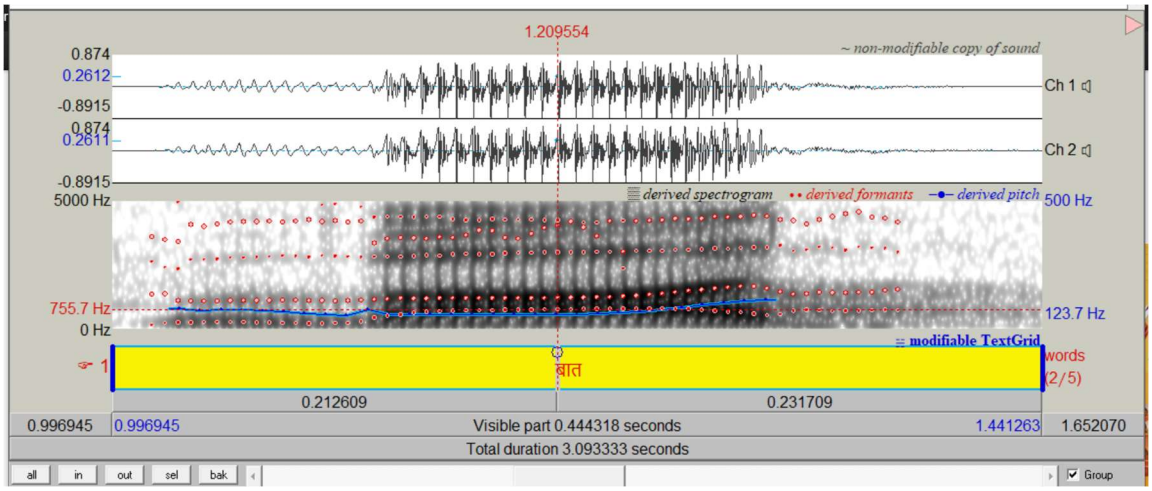

Fig.8

F1: 755.7 Hz

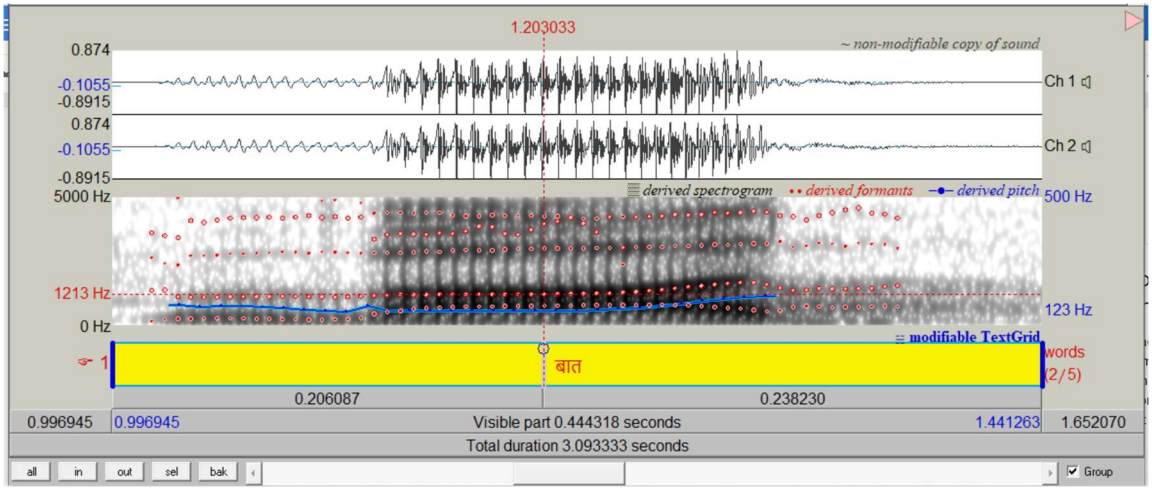

Fig.9

F2: 1213 Hz

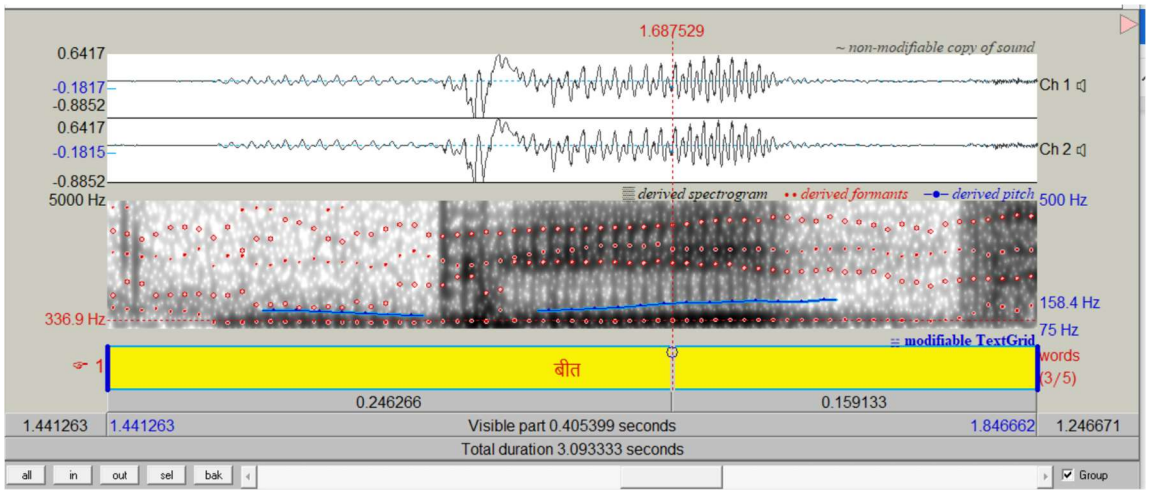

Fig.10

F1: 336.9 Hz

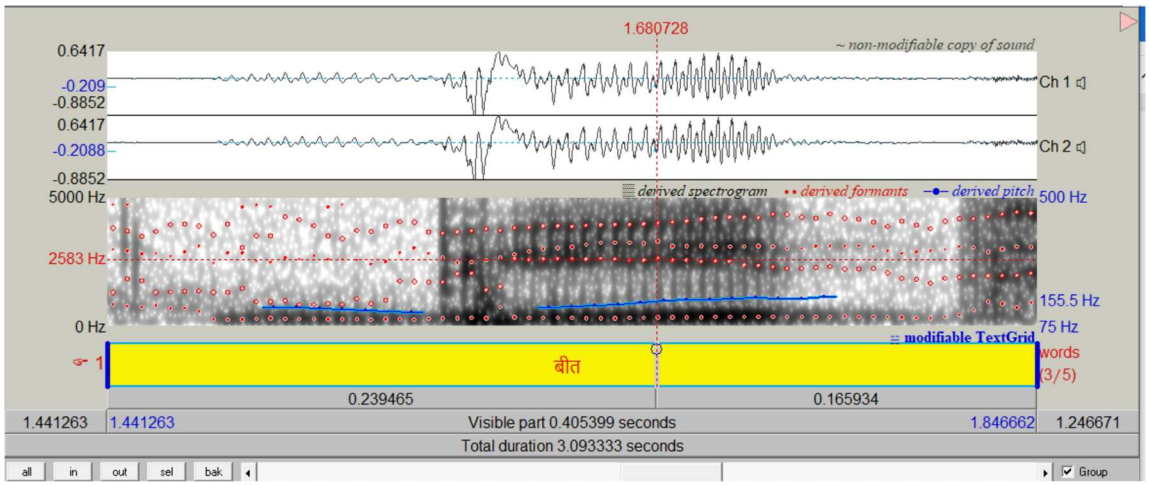

Fig.11

F2: 2583 Hz

काश केश होते

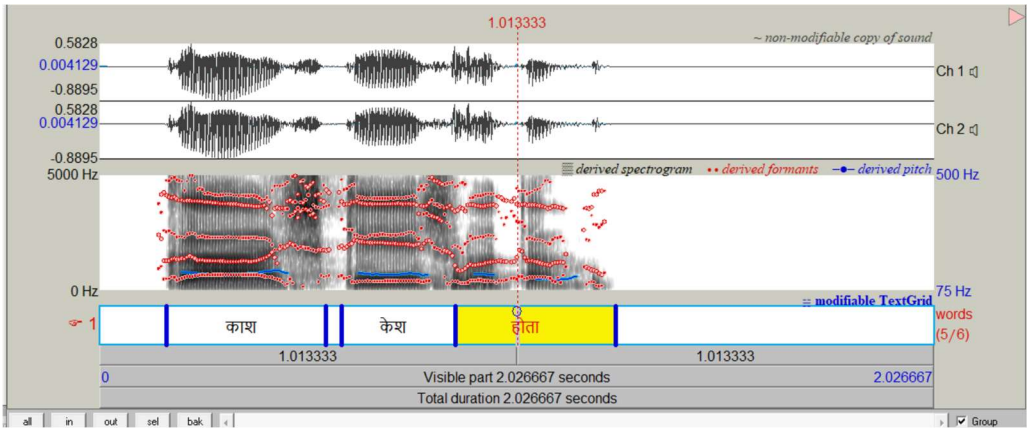

Fig.12

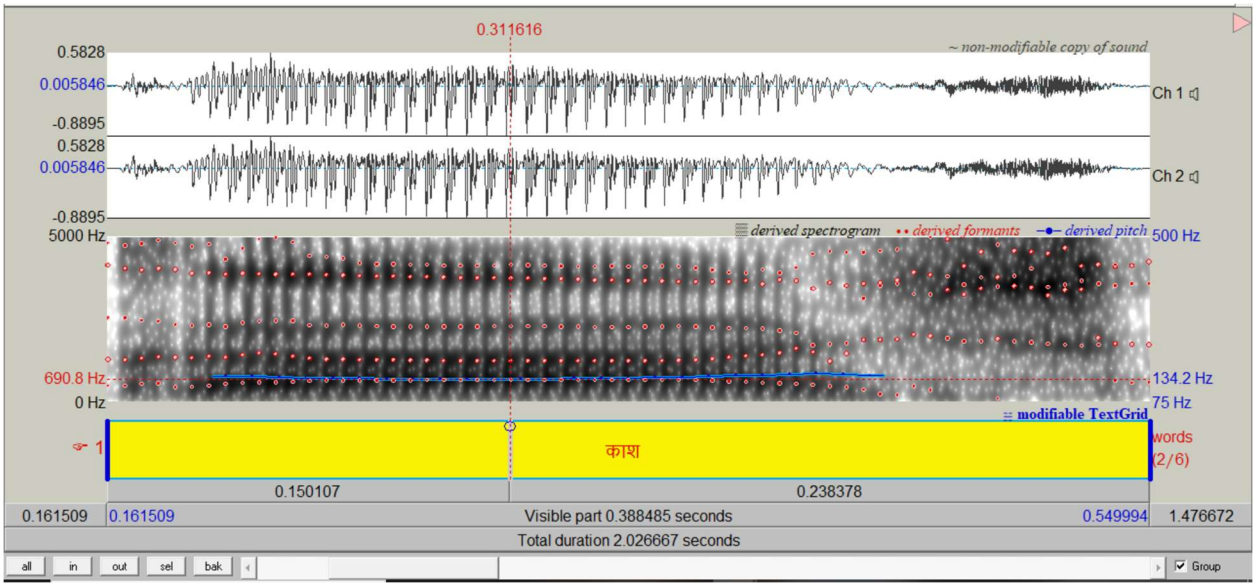

Fig.13

F1: 690.8 Hz

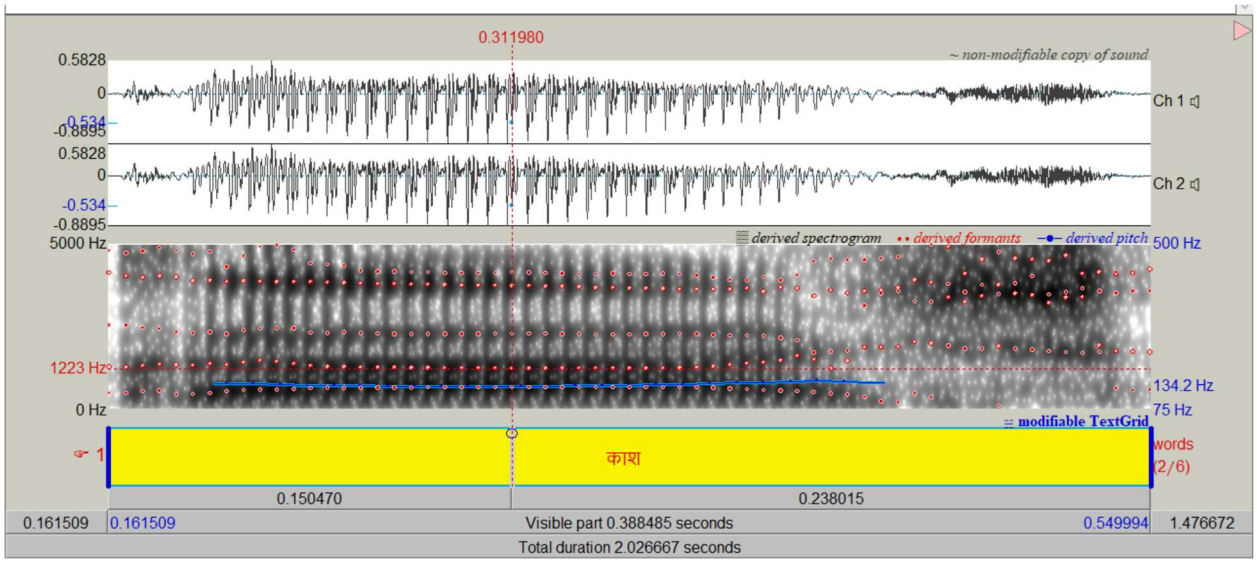

Fig.14
F2: 1223 Hz

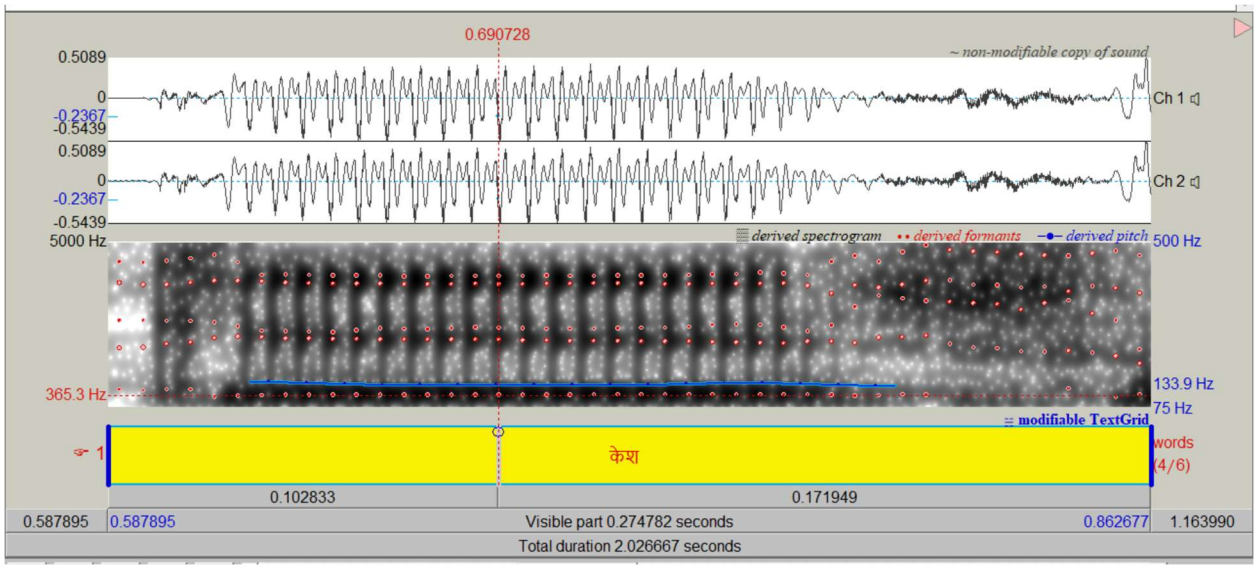

Fig.15
F1: 365.3 Hz

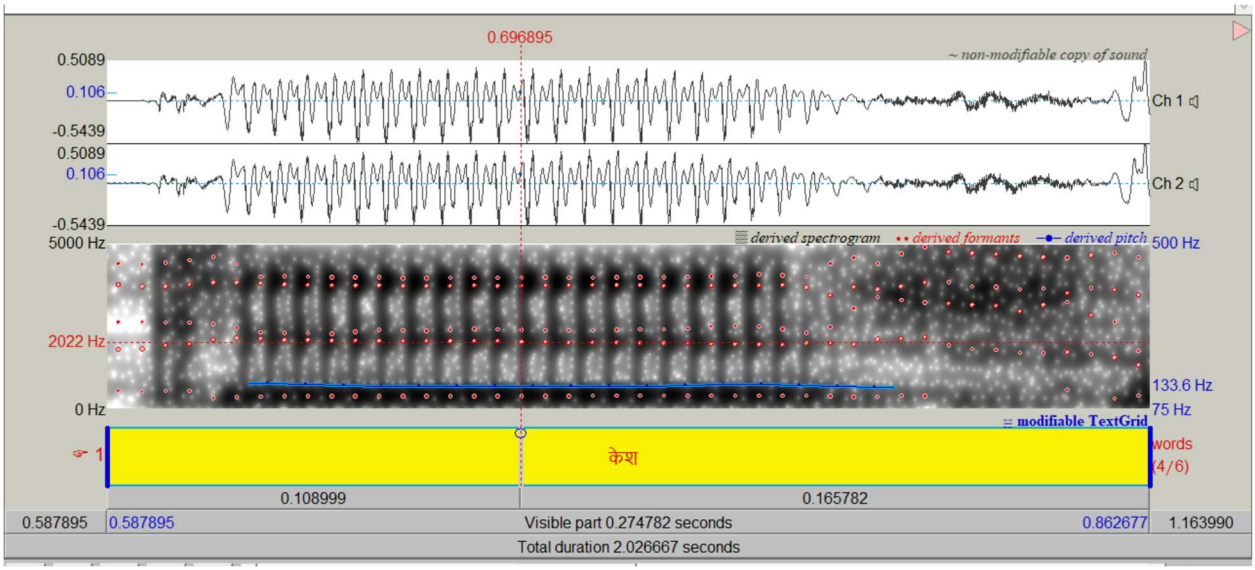

Fig.16
F2: 2022 Hz

तेल तोल लेना

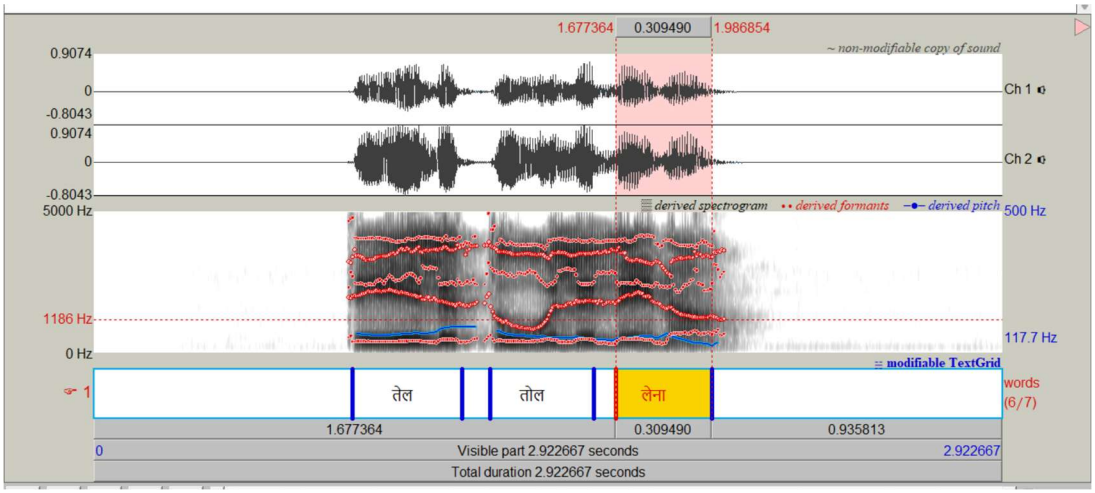

Fig.17

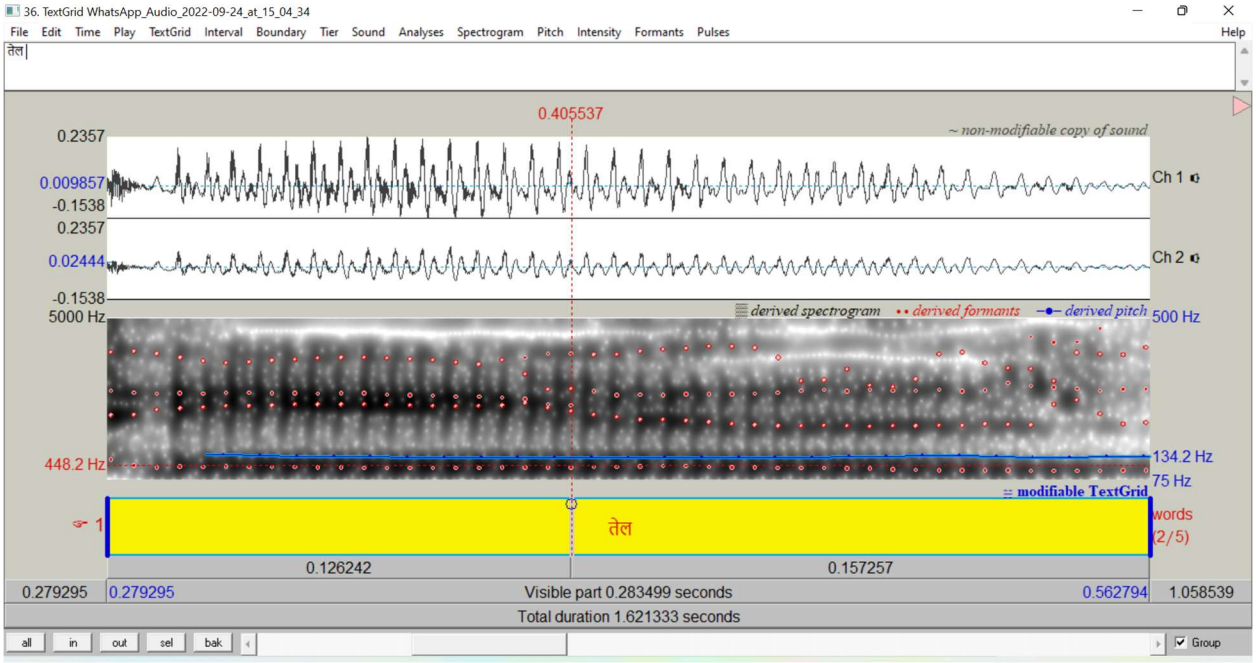

Fig.18
F1:448.2 hz

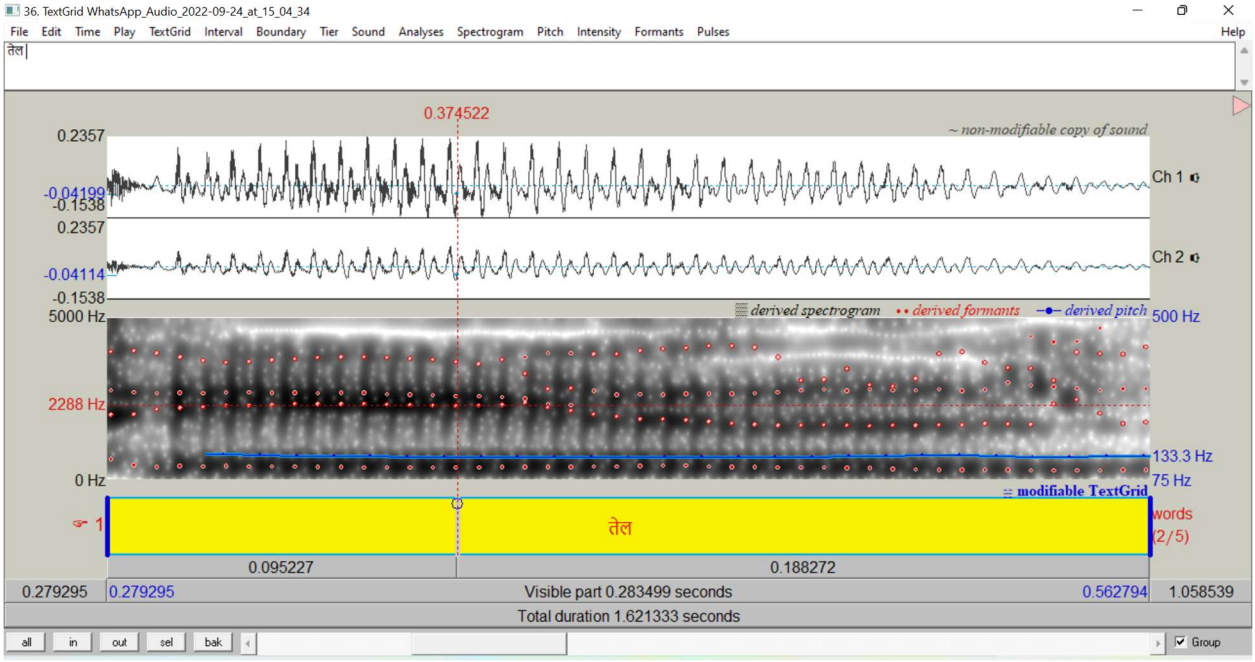

Fig.19
F2:2288 hz

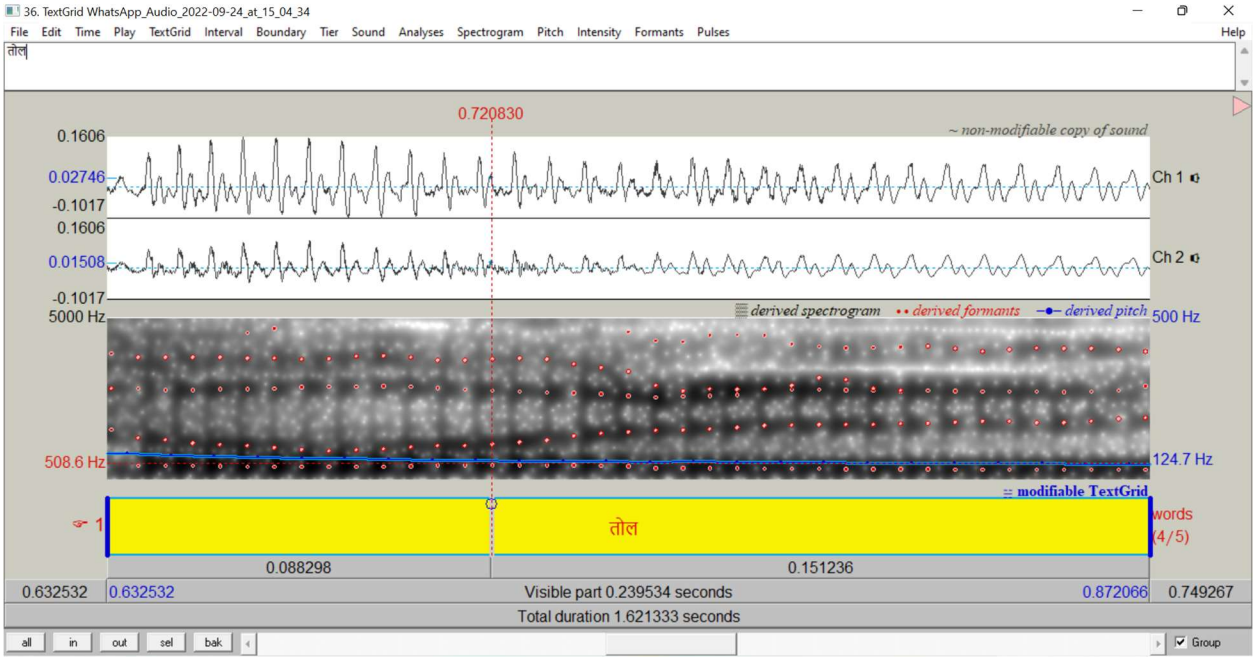

Fig.20

F1:508.6 hz

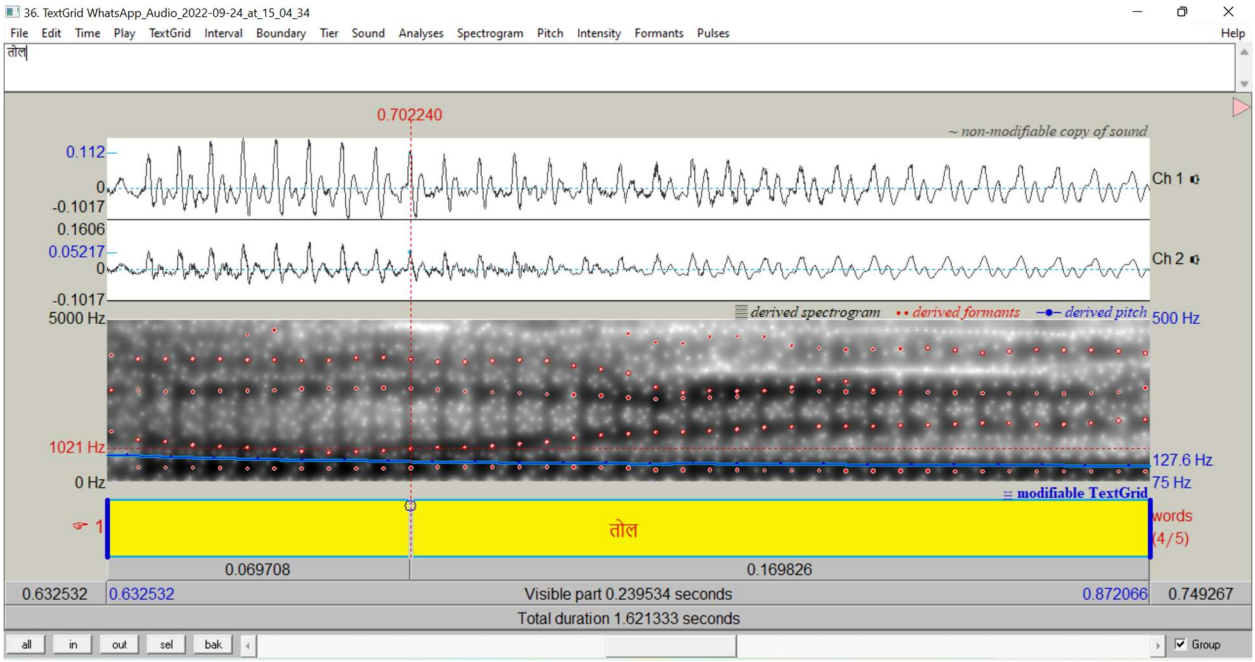

Fig.21

F2:1021 hz

तिल ताल के पास है

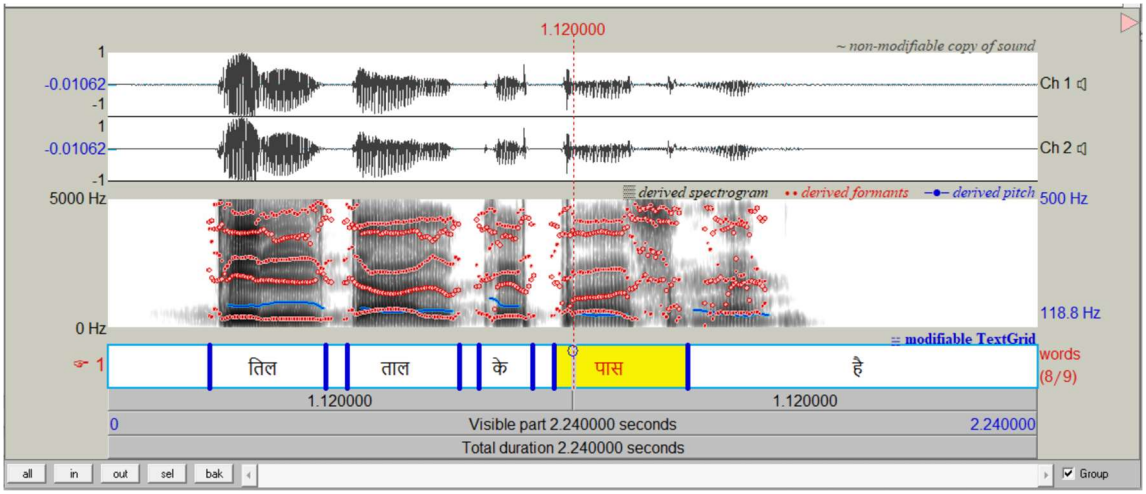

Fig.22

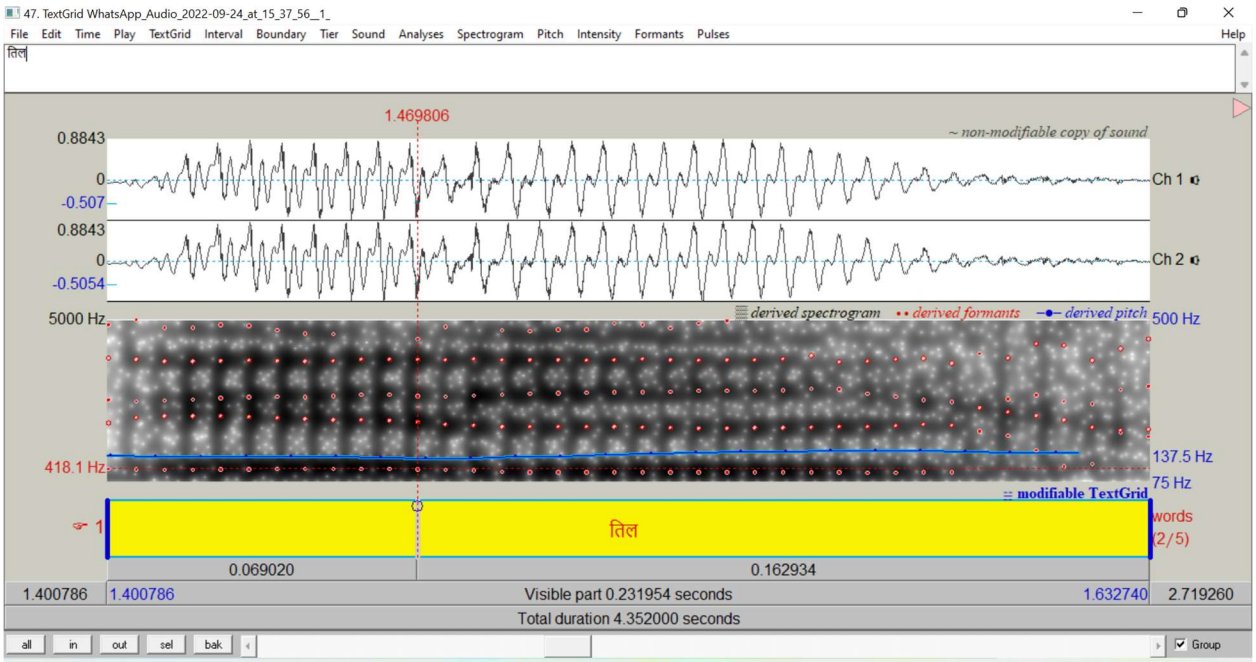

Fig.23
F1:418.1hz

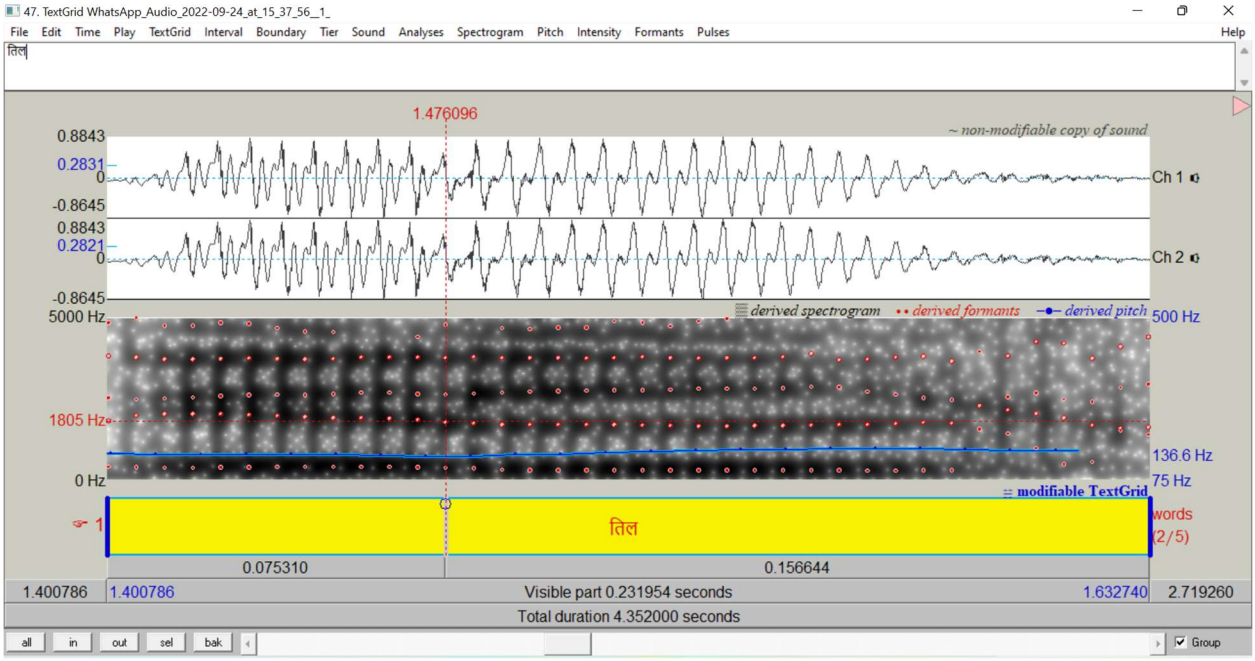

Fig.24

F2:1805 hz

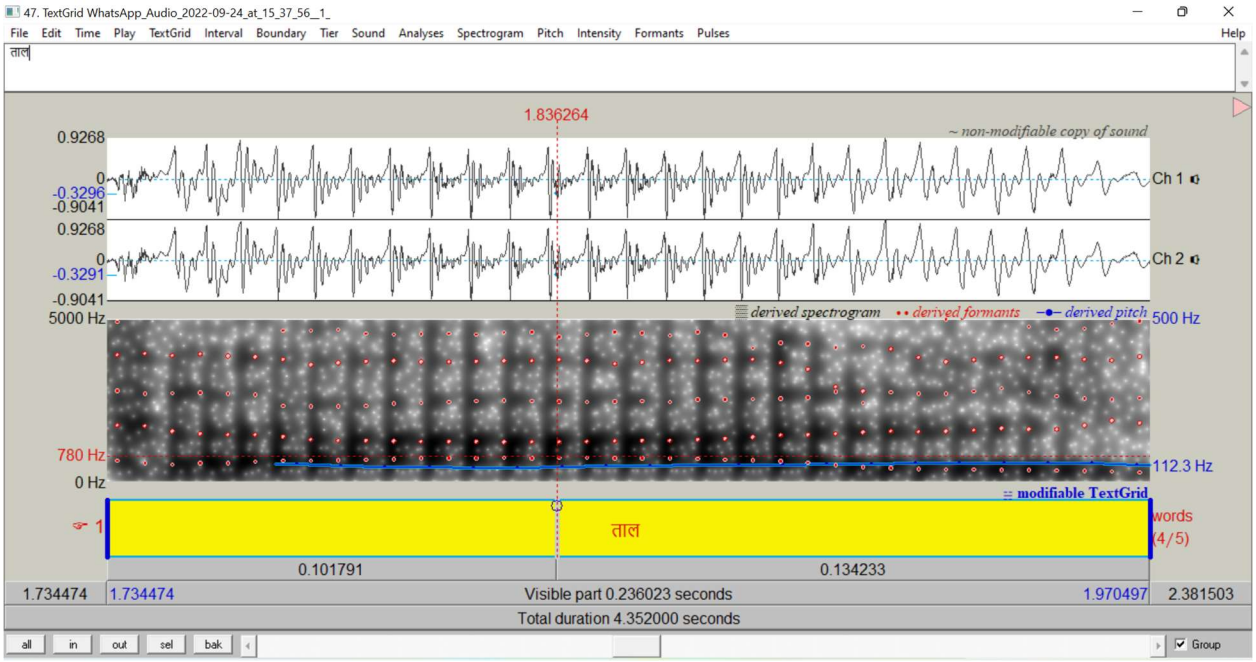

Fig.25

F1:780hz

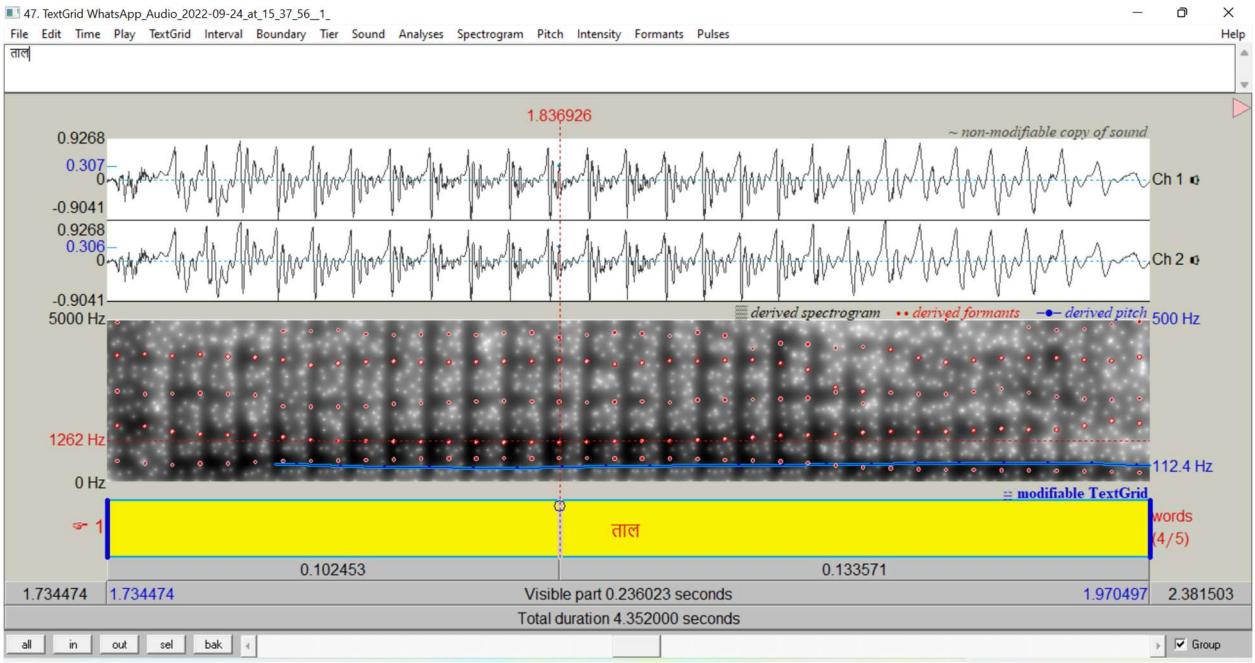

Fig.26
F2:1262 hz

Similarly we perform the formant analysis on all the samples and retrieve the data that is further compiled into an excel file. We have ignored the data of ई because it is a hindi phoneme of इ.

Results and Discussion

For all mathematical calculations, the MS Excel function of AVERAGE for arithmetic mean calculation and STDEV.S for standard deviation is applied. If there are more than one data set for a particular vowel, arithmetic mean is considered of all the observed values and further function upon.

It was further noted that F1 and F2 formant standard deviation should not be more than 100 units.

The standard F1 and F2 values used for the comparison are derived from Catford, J.C. (1988) *A Practical Introduction to Phonetics*, Oxford University Press, p. 161. ISBN 978-0198242178 these are as follows:

Average vowel formants for a male voice[6]

Vowel (IPA)	Formant F_1 (Hz)	Formant F_2 (Hz)	Difference $F_1 - F_2$ (Hz)
i	240	2400	2160
y	235	2100	1865
e	390	2300	1910
ø	370	1900	1530
ɛ	610	1900	1290
œ	585	1710	1125
a	850	1610	760
œ	820	1530	710
ɑ	750	940	190
ɒ	700	760	60
ʌ	600	1170	570
ɔ	500	700	200
ɤ	460	1310	850
o	360	640	280
ɯ	300	1390	1090
u	250	595	345

Fig.27

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	F1/F2													
1	participant	अ	आ				mean	इ	ई	ए		mean	ओ	
2	Aditya	520.3	670.4	527.3	690.8	679.6	642.025	375	298.8	365.3	365.3	365.3	394.9	
3	Lokesh	527.3	908	755.7	831.9	831.9	831.875	375	336.9	375	375	375	413	
4	Aaditya	599	780	659.3	749.8	780	742.275	418.1	357.8	508.6	448.2	478.4	448.2	
5	Priyanshu	641.5	793.8	755.7	690.8	793.8	758.525	375	298.8	394.9	375	384.95	451.1	
6	Sajal	568.9	780	835.5	780	780	793.875	508.6	387.9	387.9	448.2	418.05	478.4	
7	Shashwat	445.3	749.8	689.5	689.5	719.6	712.1	448.2	297.5	418.1	448.2	433.15	508.6	
8	Sanjay	508.6	780	870.4	870.4	840.3	840.275	387.9	418.1	508.6	508.6	508.6	538.7	
9	F1 mean	543.4					741.15	441.8	357.8			436.95	466.8	
10	F1 standard	600					750	240				390	360	
11	Standard devi	40.0222					6.2579	142.694				33.1987	75.519	
12	Priyanshu	1571	1383	1213	1223	1327	1286.5	2012	2203	2022	2052	2037	897.9	
13	Shashwat	1403	1403	1327	1342	1327	1349.75	2127	2469	2259	2241	2250	908	
14	Aditya	1654	1172	1172	1142	1262	1187	1805	2378	2077	2167	2122	930.7	
15	Sajal	1708	1479	1213	1213	1479	1346	2355	2583	2127	2317	2222	946	
16	Lokesh	1896	1443	1172	1293	1353	1315.25	1926	2378	2348	2288	2318	1021	
17	Sanjay	1594	1232	1112	1172	1262	1194.5	1805	2197	2137	2167	2152	1112	
18	Aaditya	2137	1413	1323	1564	1685	1496.25	1896	2589	2408	2378	2393	1112	
19	F2 mean	1709					1310.75	1989.43	2393			2213.43	989.657	
20	F2 standard	1170					940	2400				2300	640	
21	Standard devi	381.131					262.16	290.318				61.2152	247.245	
22														

Table.1

As we have observed that F1 standard deviation is less than 100 thus no significant changes are observed hence no need for rectification and we can say that we have verified the existing frequencies using T-test.

A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related.

Similarly analyzing all the data points we can see that drastic changes are observed in F1 standard deviation of इ, F2 standard deviation of अ (ʌ), आ(a:), इ(I), ओ (o). All are greater than 100 units.

Conclusion

Hence we can conclude that the current NLP system is no longer accurate according to the current diverse scenario and thus needs to be updated to the latest values for more efficient and accurate results.

We have also verified that the F1 values of अ (ʌ), आ(a:), ए(e), ओ (o) and F2 values of ए(e) are still valid and hence not needed to be rectified.

References

1.https://en.wikipedia.org/wiki/Formant#cite_ref-6

2.Catford, J.C. (1988) *A Practical Introduction to Phonetics*, Oxford University Press, p. 161. ISBN 978-0198242178

3. A course in phonetics by Ladefoged (2006)

4.https://corpus.eduhk.hk/english_pronunciation/index.php/1-2-introduction-of-praat/#:~:text=Praat%20is%20a%20freeware%20program,uva.nl%2FPraat%2F.