Cairo University

Faculty of engineering

Electronics and communications department

Assignment 1

DSP

Submitted to Dr.: Mohsen Rashwan

Submitted by:

Name	Sec	BN
Mohamed Maged Khalil	3	57
Mohamed Mahmoud Abdelmotaleb	3	63
Mohamed Mostafa Mahrous	3	64
Nada Hamdy Tawfiq	4	33
Norhan Ahmed Hamed	4	41

TABLE OF CONTENTS

List of Figures

QUESTION 1

Explain how the DSP can serve:

a) **Image and Video Compression :** Video compression is a practical implementation of source coding in information theory. In practice, most video codecs are used alongside audio compression techniques to store the separate but complementary data streams as one combined package using so-called *container formats*.
Uncompressed video requires a very high data rate. Although lossless video compression codecs perform at a compression factor of 5 to 12, a typical H.264 lossy compression video has a compression factor between 20 and 200.
The two key video compression techniques used in video coding standards are the discrete cosine transform (DCT) and motion compensation (MC). Most video coding standards, such as the H.26x and MPEG formats, typically use motion-compensated DCT video coding (block motion compensation).

b) The biomedical applications : there's lots of biomedical Applications in which we use dsp mainly like :

- **Hearing AID**
- **MRI(Magnetic Resonance Imaging)**
- **Medicine**
- **ECG(Electro Cardio Graphy)**
- **Ultrasound Equipment**
- **Life Support Systems**
- **Image Enhancement**

c) The military applications : In military field we can use DSP in :

- **Radar Processing**
- **Secure Communications**
- **Sonar Processing**
- **Image Processing**
- **Navigation and Guidance**

d) Human Language Technologies :

- **Speech recognition**
- **Speech Processing**
- **Speech Enhancement**
- **Text-To-Speech**
- **Voice Mail**

COMPARE BETWEEN VERIFICATION AND IDENTIFICATION TASKS IN BIOMETRICS.

WHO ARE YOU?

The key difference between identifying someone and verifying them is that identification is asking "who are you?" In biometrics terms, this known as 1-to-n matching. You're taking the individual and comparing their biometrics to a database of possible identities in order to match them and discover their identity. This is how law enforcement and border control often uses biometrics – scanning a latent print or pulling someone's fingerprint and running it against a database to see if it matches against a previously captured print.

Identifying someone with their biometrics in this way is very useful in many instances, but not in the average enterprise use case. When using biometrics for access control at a bank or hospital, you don't need to identify the user, you need to see if they are who they say they are.

ARE YOU… YOU?

Verification, on the other hand, is the process of asking "are you who you say you are?" Proving your identity. This is the actual basis of access management and biometrics-based security. Whether deploying biometrics in a mobile banking app or setting up a biometric multi factor authentication system for accessing a secure server, the user will be claiming the identity of someone already known to the system. This is known as 1-to-1 matching in biometrics. I already know who you claim to be, I just need to verify that it's true.

The best example of this is how we use biometrics with modern smartphones. Every time you use your fingerprint to unlock your smartphone, you're verifying that it's you against the fingerprint you previously scanned. It's distinctly different from identifying you based on your fingerprint, and the distinction is important for security.

SKETCH W_{Rec} , W_{Han} , W_{Ham} IN THE FREQUENCY DOMAIN ON THE SAME GRAPH

- Assumption: Window length is equal 512

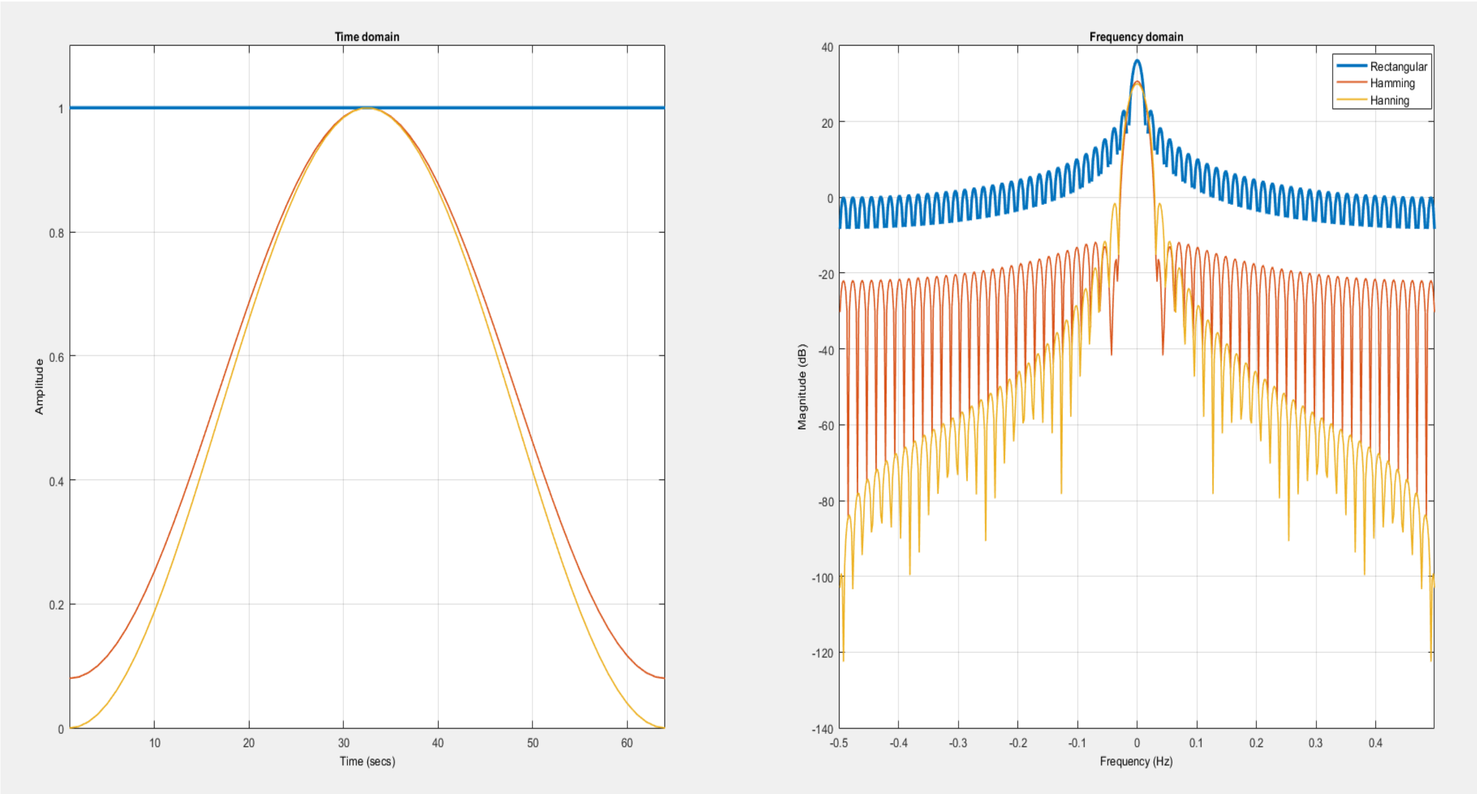


Figure 1 Freq response of Rect window,Hamming and Hanning

Comment:

- **In previous different windows we notice that there is a trade off between "Width in main lobe" and "Attenuation in side lobes"**

- **The Hanning window has a wider main lobe and more side lobe attenuation than**

  Rectangular and Hamming windows

- **The rectangular window has minimal side lobe attenuation which is undesirable so that is the poor window**

WHAT ARE THE COMPONENTS OF THE SPEECH PRODUCTION IN THE HUMAN?

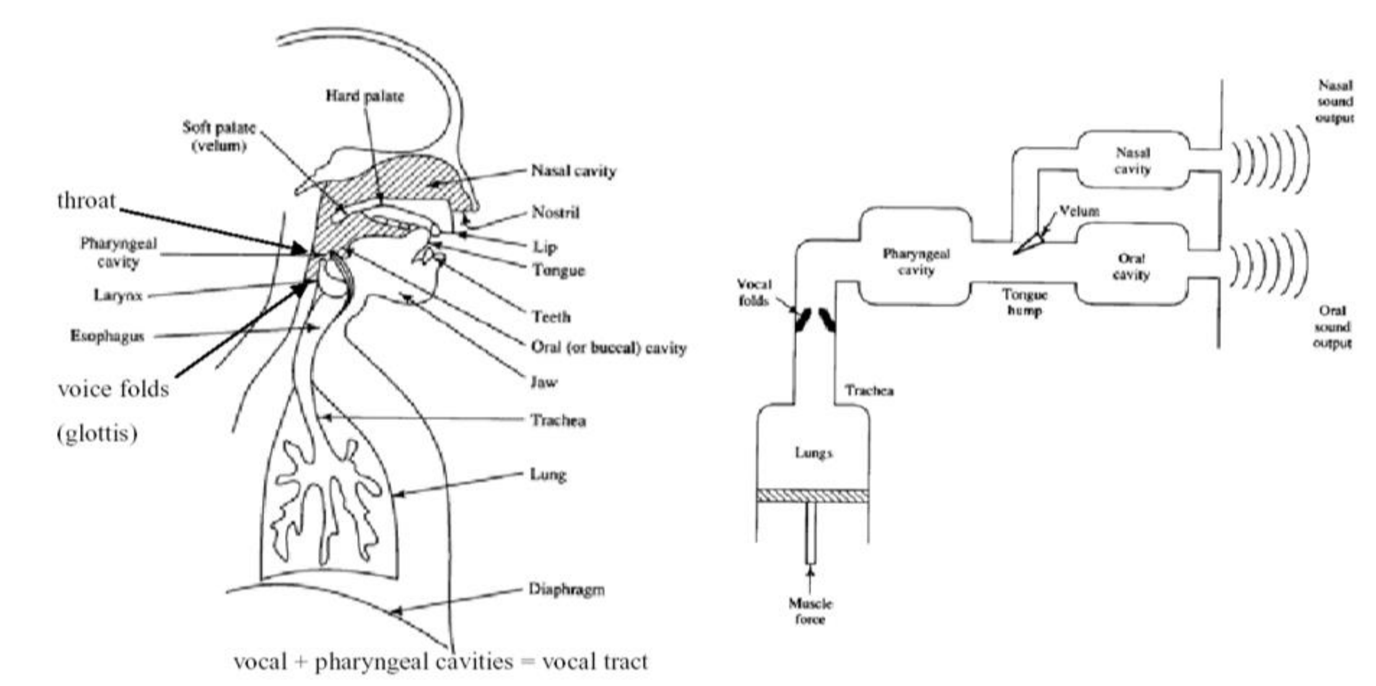


Figure 2 COMPONENTS OF THE SPEECH PRODUCTION IN THE HUMAN

- Lungs : Produce the air by Inhalation and Exhalation
- Throat : The air pass through it and it can contribute to produce the speech
- Oral Cavity : Make the sound comes out from mouth
- Nasal Cavity : Make the sound comes out from nose
- Tongue : Change the spaces which the air passed through
- Lips : Produce different phonemes by its formation

CALCULATE THE LINEAR PREDICTIVE COEFFICIENTS (A1 AND A2) AND THE MEAN SQUARED ERROR BY SOLVING THE MATRIX EQUATION:

R(0) = 1 , R(1) = 0.8 , R(2) = 0.5

$$\begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix}$$

$a_1 + 0.8a_2 = 0.8 \quad \Rightarrow \mathrm{I}$

$0.8a_1 + a_2 = 0.5 \quad \Rightarrow \mathrm{II}$

By solving I & II :

$$a_1 = \frac{10}{9} = 1.1111 \quad , \ a_2 = -\frac{7}{18} = -0.38889$$

$$Mean\ Squared\ Error = 1^2 + (0.8 - \left(\left(\frac{10}{9} * 1\right)\right)^2 + (0.5 - \left(\left(\frac{10}{9} * 0.8\right) + \left(\frac{-7}{18} * 1\right)\right)^2 = 0.81543$$

QUESTION 6:

The frame size is 5 so the first frame is ⇒ [1,2,3,1,4]

d)

$$r_0 = ((1*1) + (2*2) + (3*3) + (1*1) + (4*4)) = 31$$

$$r_1 = \left((1*2) + (2*3) + (3*1) + (1*4)\right) = 15$$

$$r_2 = \left((1*3) + (2*1) + (3*4)\right) = 17$$

$$\begin{bmatrix} 31 & 15 \\ 15 & 31 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 17 \end{bmatrix}$$

$31a_1 + 15a_2 = 15 \quad \Rightarrow \mathrm{I}$

$15a_1 + 31a_2 = 17 \quad \Rightarrow \mathrm{II}$

By solving I & II : $a_1 = \frac{105}{368} = 0.28532 \ , \ a_2 = \frac{151}{368} = 0.41033$

e) If the number of overlapping samples for two frames is 2 so Second frame is [1,4,1,2,4]

$$r_0 = ((1*1) + (4*4) + (1*1) + (2*2) + (4*4)) = 38$$

$$r_1 = \left((1*4) + (4*1) + (1*2) + (2*4)\right) = 18$$

$$r_2 = \left((1*1) + (4*2) + (1*4)\right) = 13$$

$$\begin{bmatrix} 38 & 18 \\ 18 & 38 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 18 \\ 13 \end{bmatrix}$$

$38a_1 + 18a_2 = 18 \quad \Rightarrow \mathrm{I}$

$18a_1 + 38a_2 = 13 \quad \Rightarrow \mathrm{II}$

By solving I & II : $a_1 = \frac{45}{112} = 0.401785$, $a_2 = \frac{17}{112} = 0.15178$

f) If the pre-emphasis constant is 0.95 :

[1 2-(0.95*1) 3-(0.95*2) 1-(0.95*3) 4-(0.95*1) 1-(0.95*4) 2-(0.95*1) 4-(0.95*2) 3-(0.95*4)]

New speech waveform = [1 , 1.05 , 1.1 , -1.85 , 3.05 , -2.8 , 1.05 , 2.1 , -0.8]

- **First Frame = [1 , 1.05 , 1.1 , -1.85 , 3.05]**

$r_0 = ((1*1) + (1.05*1.05) + (1.1*1.1) + (-1.85*-1.85) + (3.05*3.05)) = 16.0375$

$r_1 = \big((1*1.05) + (1.05*1.1) + (1.1*-1.85) + (-1.85*3.05)\big) = -5.4725$

$r_2 = \big((1*1.1) + (1.05*-1.85) + (1.1*3.05)\big) = 2.5125$

$$\begin{bmatrix} 16.0375 & -5.4725 \\ -5.4725 & 16.0375 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -5.4725 \\ 2.5125 \end{bmatrix}$$

$16.0375a_1 - 5.4725a_2 = -5.4725 \quad \Rightarrow \text{I}$

$-5.4725a_1 + 16.0375a_2 = 2.5125 \quad \Rightarrow \text{II}$

By solving I & II : $a_1 = -0.325696$, $a_2 = 0.045526$

- **Second Frame = [-1.85 , 3.05 , -2.8 , 1.05 , 2.1]**

$r_0 = ((-1.85*-1.85) + (3.05*3.05) + (-2.8*-2.8) + (1.05*1.05) + (2.1*2.1)) = 26.0775$

$r_1 = \big((-1.85*3.05) + (3.05*-2.8) + (-2.8*1.05) + (1.05*2.1)\big) = -14.9175$

$r_2 = \big((-1.85*-2.8) + (3.05*1.05) + (-2.8*2.1)\big) = 2.5025$

$$\begin{bmatrix} 26.0775 & -14.9175 \\ -14.9175 & 26.0775 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -14.9175 \\ 2.5025 \end{bmatrix}$$

$26.0775a_1 - 14.9175a_2 = -14.9175 \quad \Rightarrow \text{I}$

$-14.9175a_1 + 26.0775a_2 = 2.5025 \quad \Rightarrow \text{II}$

By solving I & II : $a_1 = -0.768692$, $a_2 = -0.3437626$

QUESTION 7

$$H(z) = \frac{G}{1 + \sum_{k=1}^{8} a_k z^{-k}}$$

If $\alpha_1 = -2.12$, $\alpha_2 = 2.89$, $\alpha_3 = -3.4$, $\alpha_4 = 3.55$, $\alpha_5 = -3.13$, $\alpha_6 = 2.25$, $\alpha_7 = -1.2$, and $\alpha_8 = 0.47$.

The magnitude response of the previous system is as following

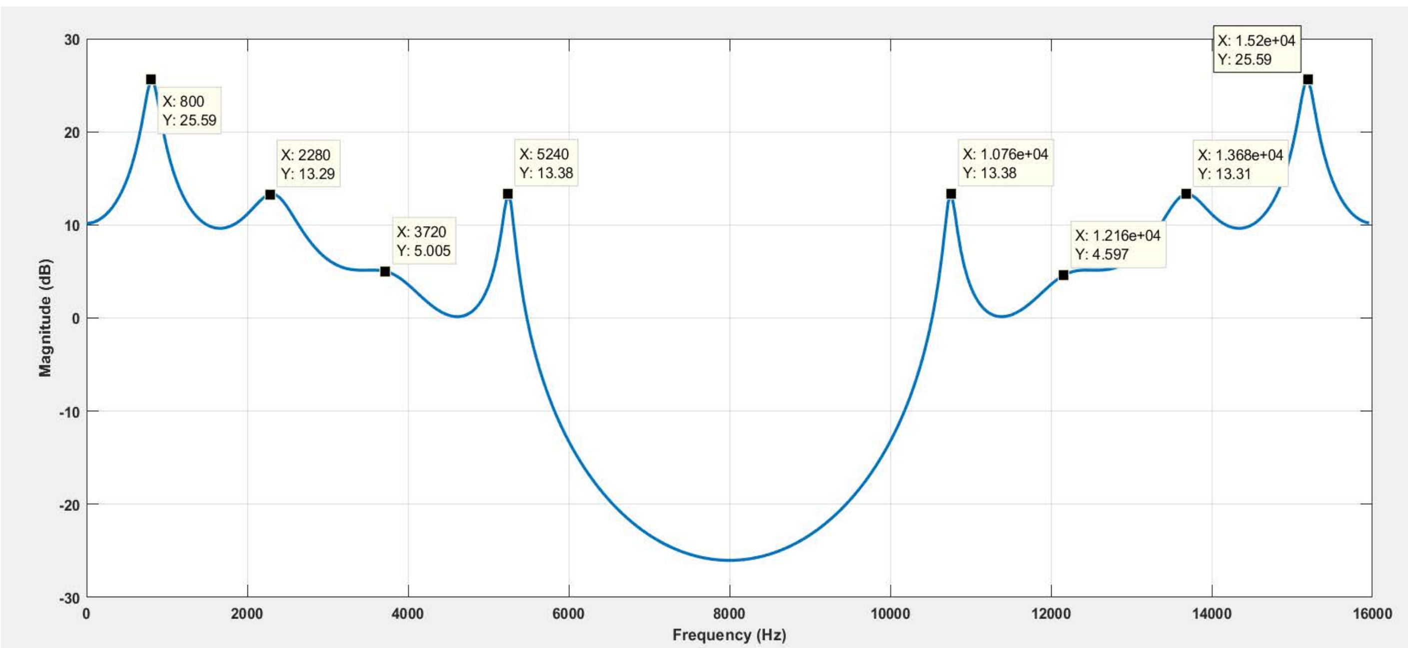


Figure 3 Magnitude response of the system

As we see the following system has 4 formants (Max peak) as it has 4 conjugate poles

QUESTION 8

$$A(z)=z^{-10} - 1.8779z^{-9} + 1.2491z^{-8} - 0.444z^{-7} + 0.5018z^{-6} - 0.9232z^{-5} + 0.5867z^{-4} - 0.0706z^{-3} - 0.016z^{-2} - 0.1158z^{-1} + 0.1483$$

LPC spectrum ,Formants and BW are as follow

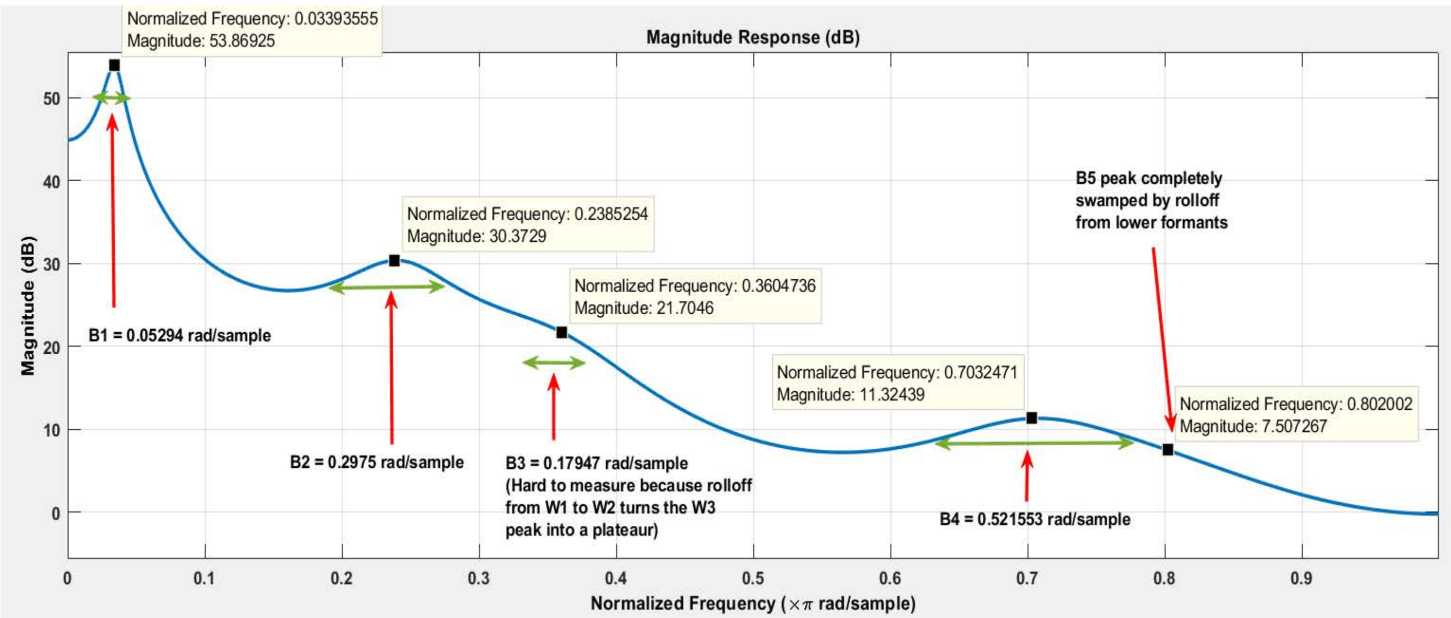


Figure 4 LPC spectrum , BW and Formants of the system

This system should has 5 Formants(Max peaks) as we saw there are 4 formants are obviously cleared but the fifth peak is completely swamped by rolloff from lower formants

QUESTION 9

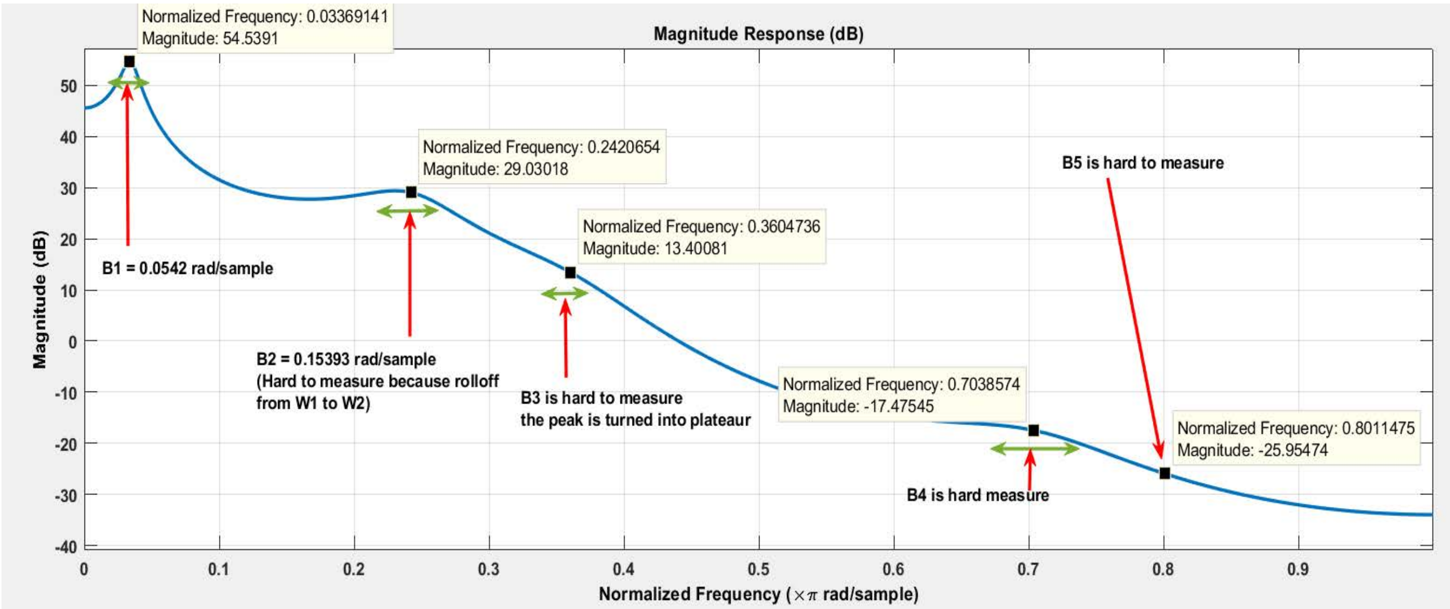


Figure 5 BW and Formants of the system

As we see when we have roots of the system and we plot the LPC the spectrum we find that BW and Formants are the same and there is no any difference from Question 8 .

Hand analysis is in the next page

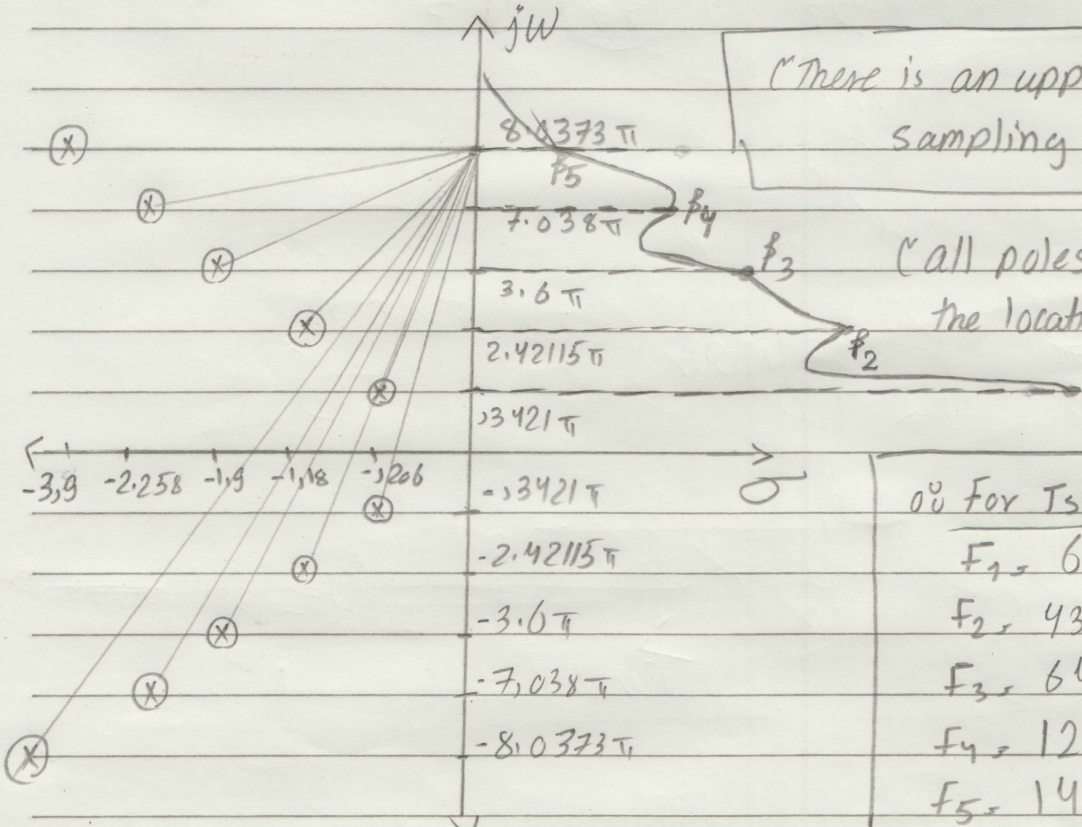
Z - domain . التاريخ: 5 - domain roots الموضوع:

roots	r	Ø rad	σ	ω (rad/sec)
,9739 ± ,1051j	,97955	± ,03421π	-,2066	± ,3421 π
,6432 ± ,6121j	,887903	± ,242115 π	-1.18892	± 2,4211 π
,3492 ± ,744j	,821916	± ,36028 π	-1.96117	± 3.6028 π
-,4768 ± ,6397j	,79784	± ,703882 π	-2.2584	± 7,03882 π
15508 ± ,3904j	,675127	± ,80373 π	-3,9285	± 8,0373 π

⊛ $Z = |e^{\sigma T_s}| \angle \pm \omega T$

⊛ $\omega = \frac{\text{Ø rad}}{T_s \text{ sec}}$, $\sigma = \frac{\ln(r)}{T_s}$,

⊛ , $T_s = ,1$ sec (Assumed)



(There is an upper limit due to sampling)

(all poles contribute to the location of formant)

o° For $T_s = ,1$ sec

$F_1 = 61,578$ HZ

$F_2 = 435,798$ HZ

$F_3 = 648,504$ HZ

$F_4 = 1266,98$ HZ

$F_5 = 1446,714$ HZ

⊛ For a pole to be close to jw axis → the formant is far from jw axis

⊛ for each of 2 conjugate poles we have 1 peak.

QUESTION 10

a. Define these terms:

Phoneme: the smallest unit (indivisible unit) of the phonetic system of a language that can be used to make one word different from another word as it's correspond to a set of similar speech sounds that are known to be a single characteristic sound in a language.

Vowel: a speech sound produced by humans when the breath flows out through the mouth without being blocked by the teeth or tongue or lips as the oral part of the breath channel is not blocked.

Constant: it represents sounds that are made when part of or the whole vocal tract is closed as it requires a specific position of lips, tongue and so on.

Voiced phoneme and Unvoiced phoneme: Voiced signals are produced when the vocal cords vibrate during the pronunciation of a phoneme but Unvoiced signals don't use the vocal cords vibration. For example: the phoneme /s/ is unvoiced but /z/ is voiced as it requires vibration of vocal cords.

Fricative consonant: they are consonants produced when air passes through a narrow gap as it leaves the body. There are different types of fricative consonants depending on the parts involved in the emission of the sounds.
For example:

1) Labiodental Fricative /f/ as in "<u>f</u>ive": it's made by using the lower lip and the upper teeth.
2) Dental Fricative /θ/ as in "<u>th</u>ing": The tip of the tongue against the teeth creates these fricatives.
3) Alveolar Fricative /z/ as in "<u>z</u>oo": here the tip of the tongue moves up against the gum line behind the upper teeth.

Stop consonant: a stop consonant has three parts to be created. First, either the tongue or the lips close off the air flow in the vocal tract. Second, this closure causes a buildup of pressure. Third, there is a release of this built up air.

For example:

1) /b/ as in "<u>b</u>oy" or /p/ as in "<u>p</u>en": Stop the airflow by holding lips together. Then part the lips quickly to release the airflow with sudden pressure.

2) /t/ as in "<u>t</u>op" or /d/ as in "<u>d</u>o": Place tongue tip on the alveolar ridge to stop the airflow. Then lower the tip to release the airflow.

3) /k/ as in "<u>c</u>at" or /g/ as in "<u>g</u>o": Raise the back of tongue against the palate to stop the airflow. Drop the tongue quickly to release the airflow with sudden pressure.

b. Give examples (at least 2) for:

Voiced fricative:

1) Labiodental Fricative /v/ as in "<u>v</u>ine" , "<u>v</u>iolet".

2) Dental Fricative /ð/ as in "<u>th</u>ese" , "<u>th</u>at".

3) Alveolo-Palatal Fricative /ʒ/ as in "plea<u>su</u>re"

Unvoiced fricative:

1) Labiodental Fricative /f/ as in "<u>f</u>ive" , "<u>f</u>our".

2) Dental Fricative /θ/ as in “<u>th</u>ing”.

3) Alveolo-Palatal Fricative / ʃ / as in “pre<u>ss</u>ure” , “mo<u>t</u>ion”.

Nasal phoneme: to be created the air is completely blocked from leaving the mouth, and is instead released out through the nose.

1) The bilabial /m/ as in “<u>m</u>ode”: lips block the air in this case.

2) The alveolar /n/ as in “<u>n</u>eck”: the tip of the tongue at the alveolar ridge blocks the air.

3) The velar /ŋ/ as in “so<u>ng</u>”: the back of the tongue at the velum blocks the air.

Stop, voiced phoneme:

1) /b/ as in “<u>b</u>ee”.

2) /g/ as in “<u>g</u>ift”.

3) /d/ as in “<u>d</u>og”.

In order to make model for auditory process we need to understand the nature of auditory for human knowing its structure and how it works.

The structure(figure6**):**

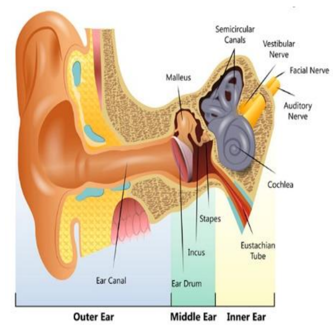


Figure 6:human auditory system

when the sound wave goes through pinna(modifies the incoming sound, particularly at high frequencies)down to auditory canal until it hits the eardrum(the source of vibrations), then vibrations will transmitted to the three bones attached to (malleus,incus,stapes) to the oval window of the cochlea which contains fluid it will move to the tip of cochlea and returns back to back to circular window until the energy is disappeared through this process the hair cell will be pushed so it sends electrical impulses to the brain.

Note Because of the resonance of the outer ear, we are more sensitive to sound frequencies between 1000 and 6000 Hz and as shown in figure2 the transfer function of the ear canal. Noticed the middle ear acts as an impedance-matching device improves sound transmission and reduces the amount of reflected sound.

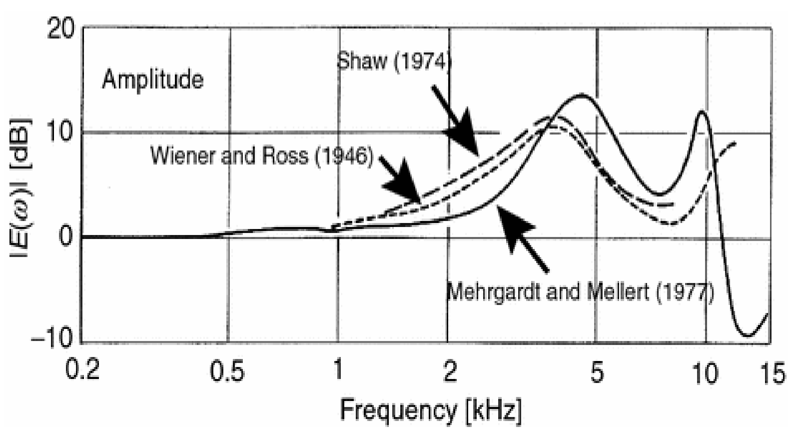


Figure 7:Transfer functions (amplitude) of the middle ear

Now we could summarize and build the model and get the main components of auditory process (as shown in next figures)

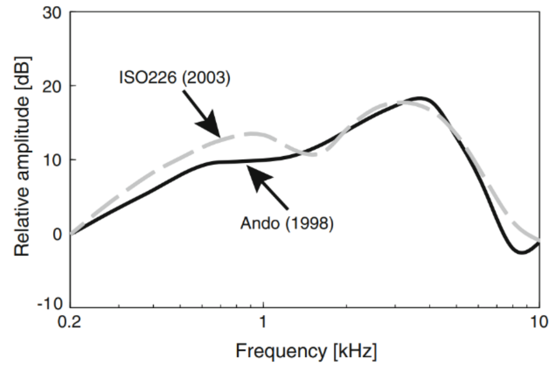


Figure 8: Sensitivity of the human ear to a sound source in front of the listeners estimated from transformation characteristics between the sound source and the cochlea

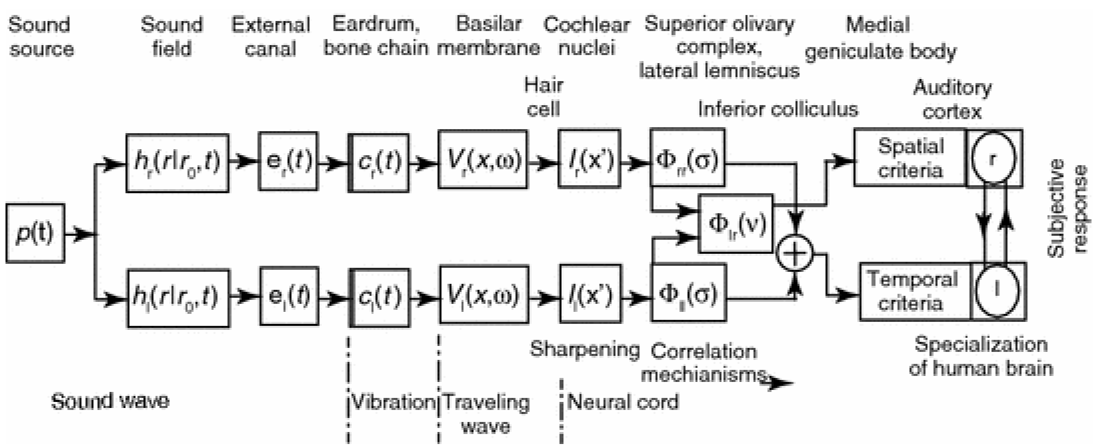


Figure 9:modeling auditory process

QUESTION 12

A.RECORDING OF THE DIGITS

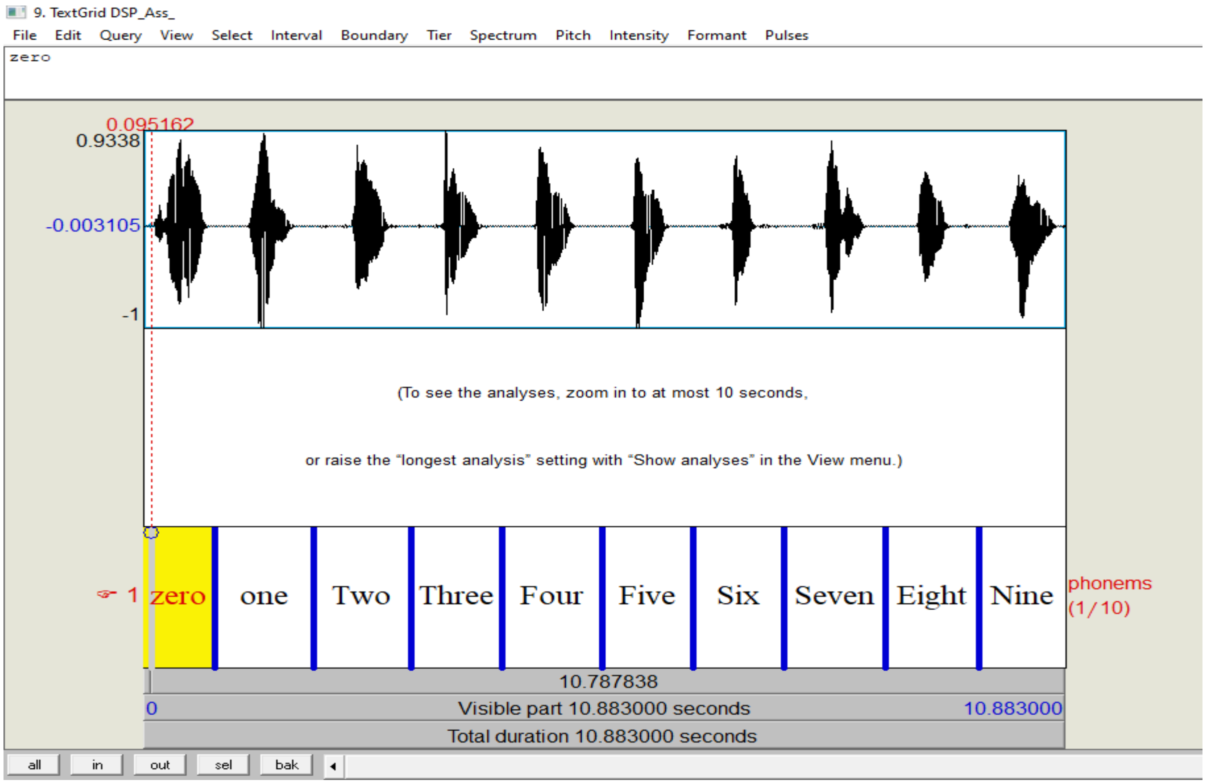


Figure 10 Recordings of digits from zero to nine

Note: For points b, c and d in this question I put the phonemes , segment and the whole word in the same place , for pitch and formants I draw each of them in separate graph as following

C. DRAW PITCH CONTOUR OF THE SEGMENT

- **Zero**

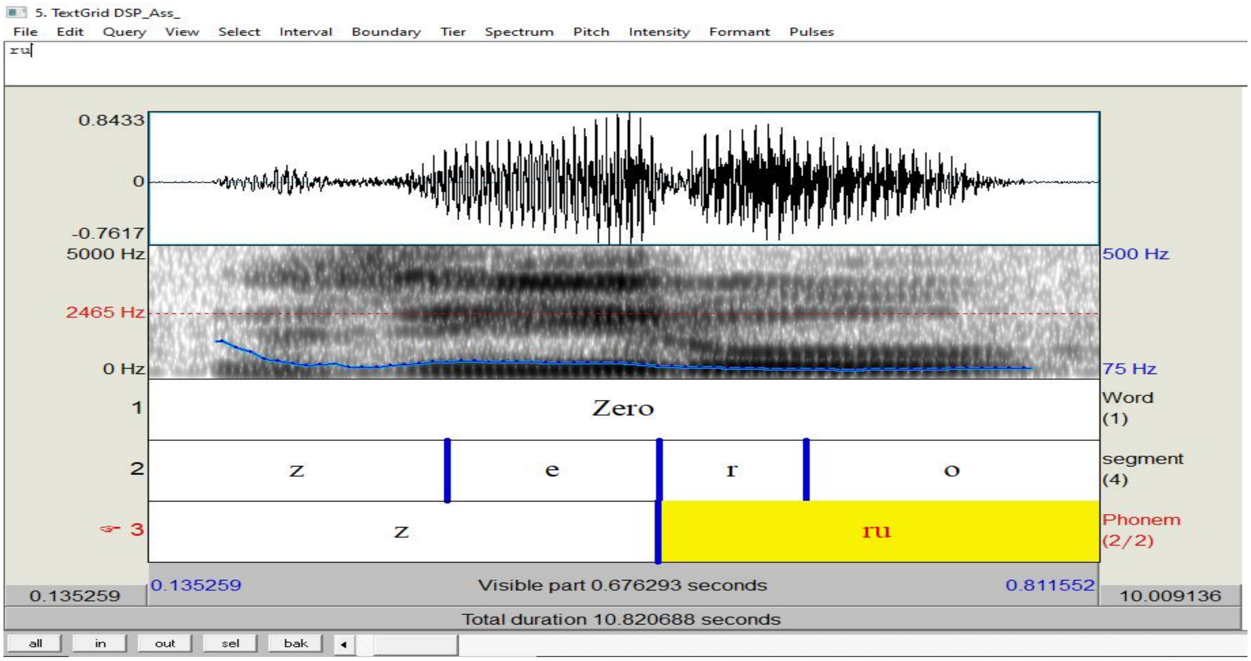


Figure 11 Pitch of Zero

- **One**

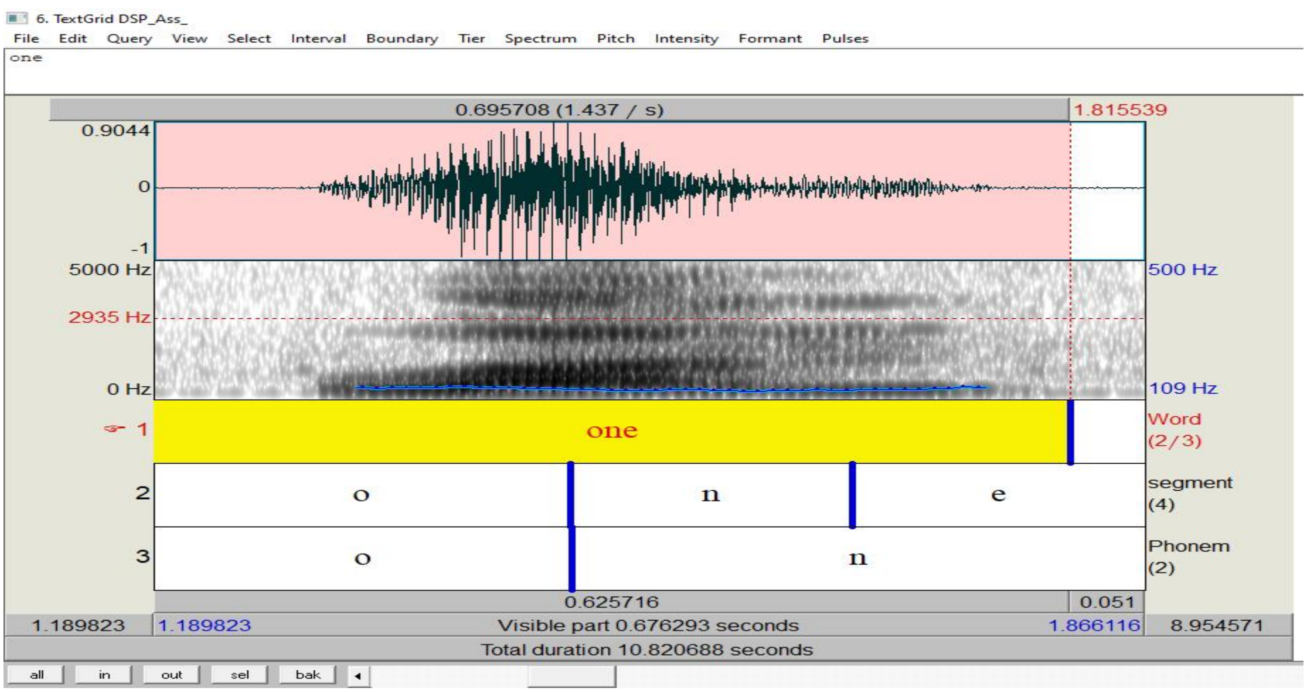


Figure 12 Pitch of One

- **Two**

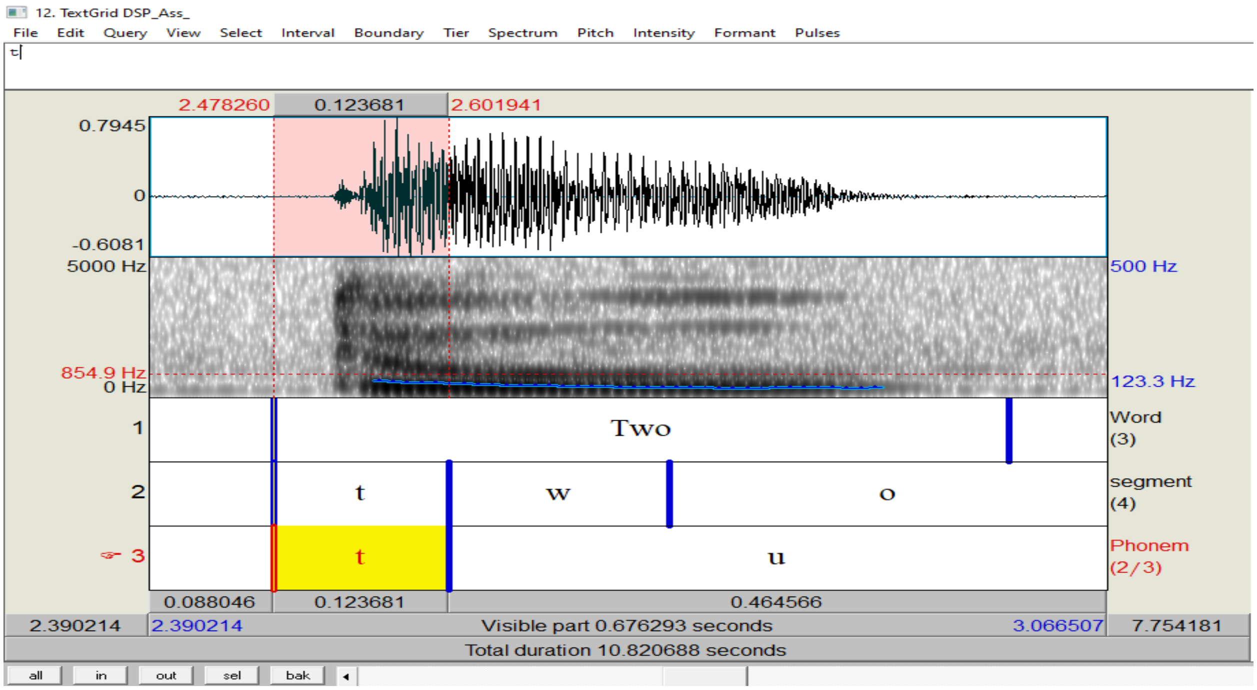


Figure 13 Pitch of Two

- **Three**

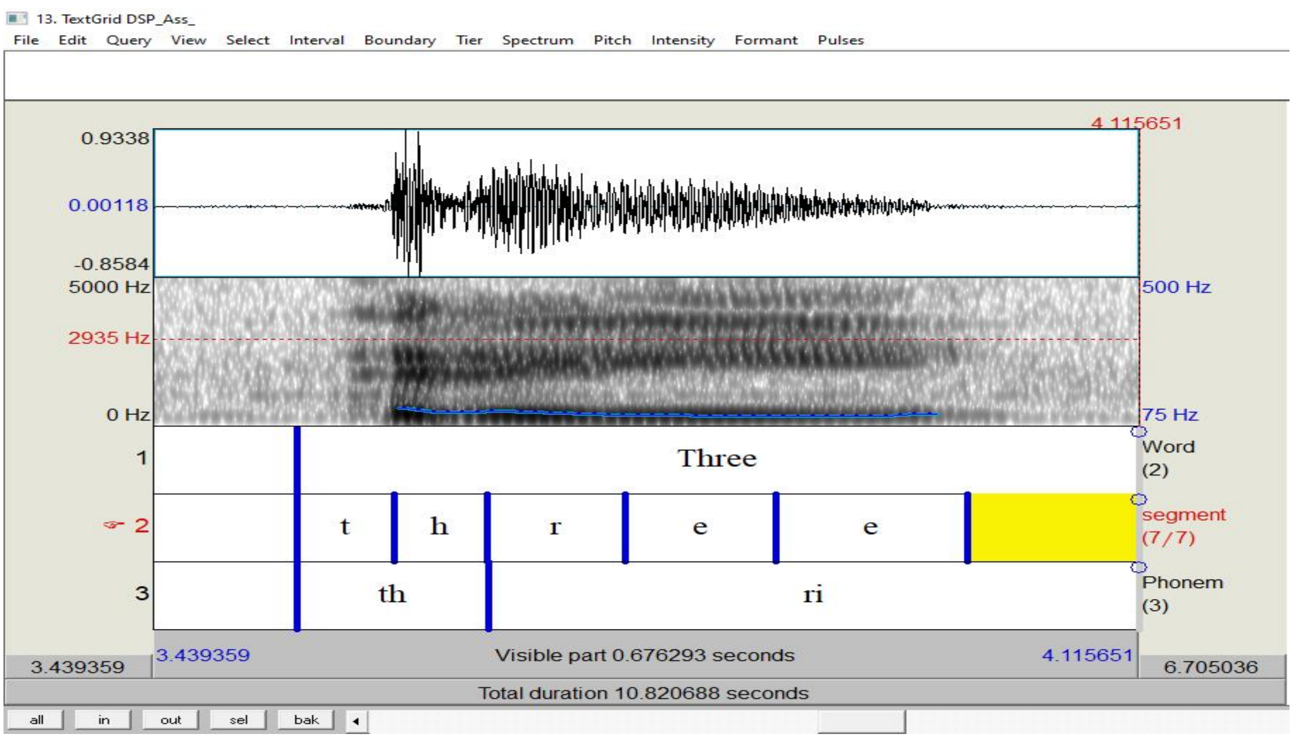


Figure 14 Pitch of Three

- **Four**

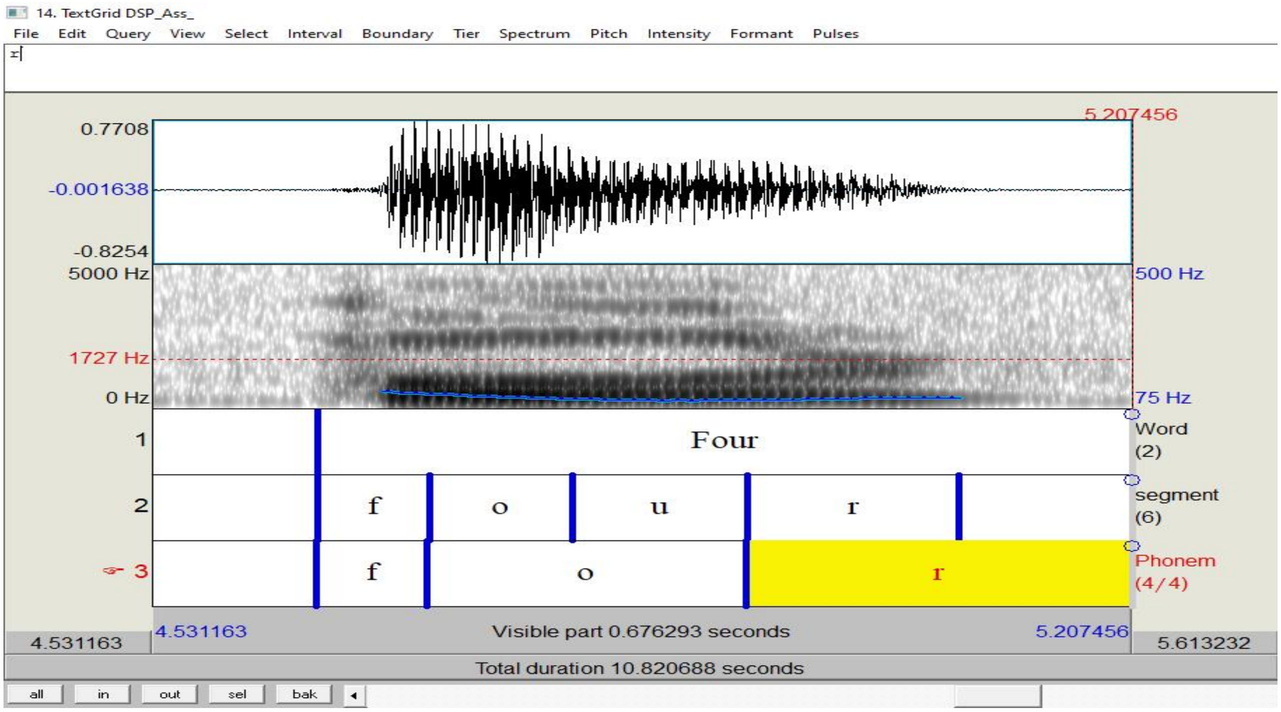


Figure 15 Pitch of Four

- **Five**

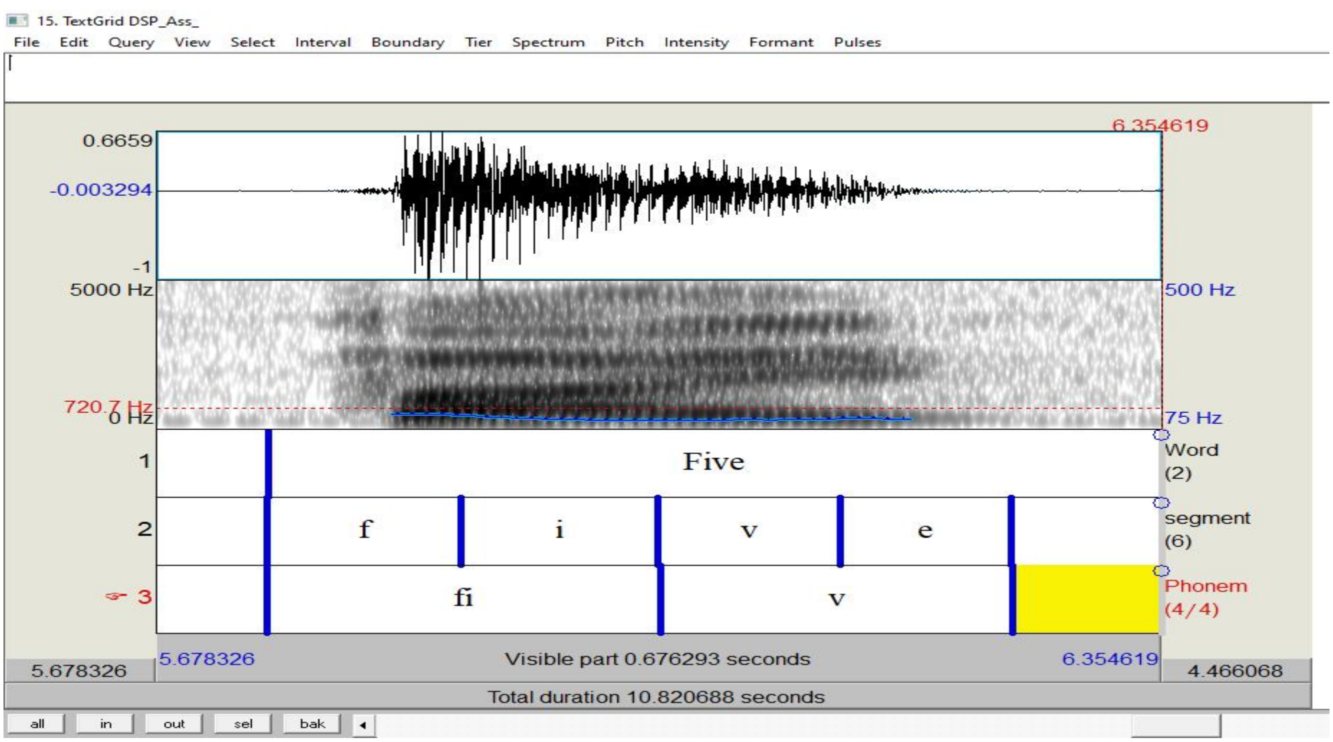


Figure 16 Pitch of five

- **Six**

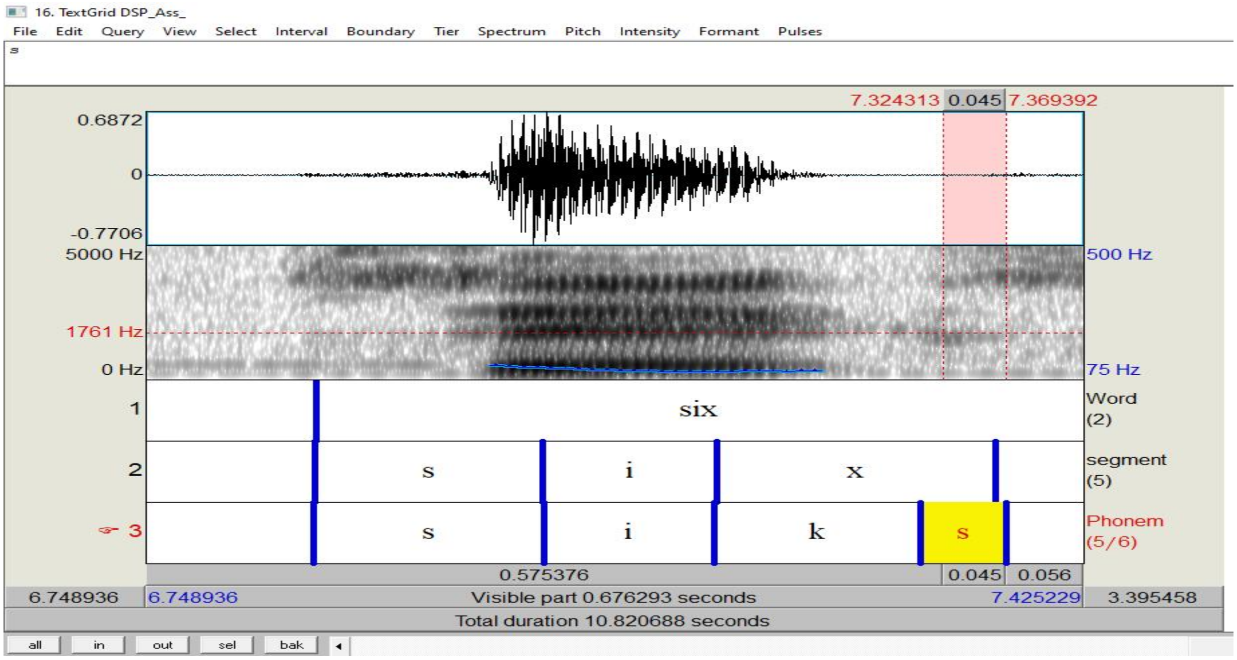


Figure 17 Pitch of six

- **Seven**

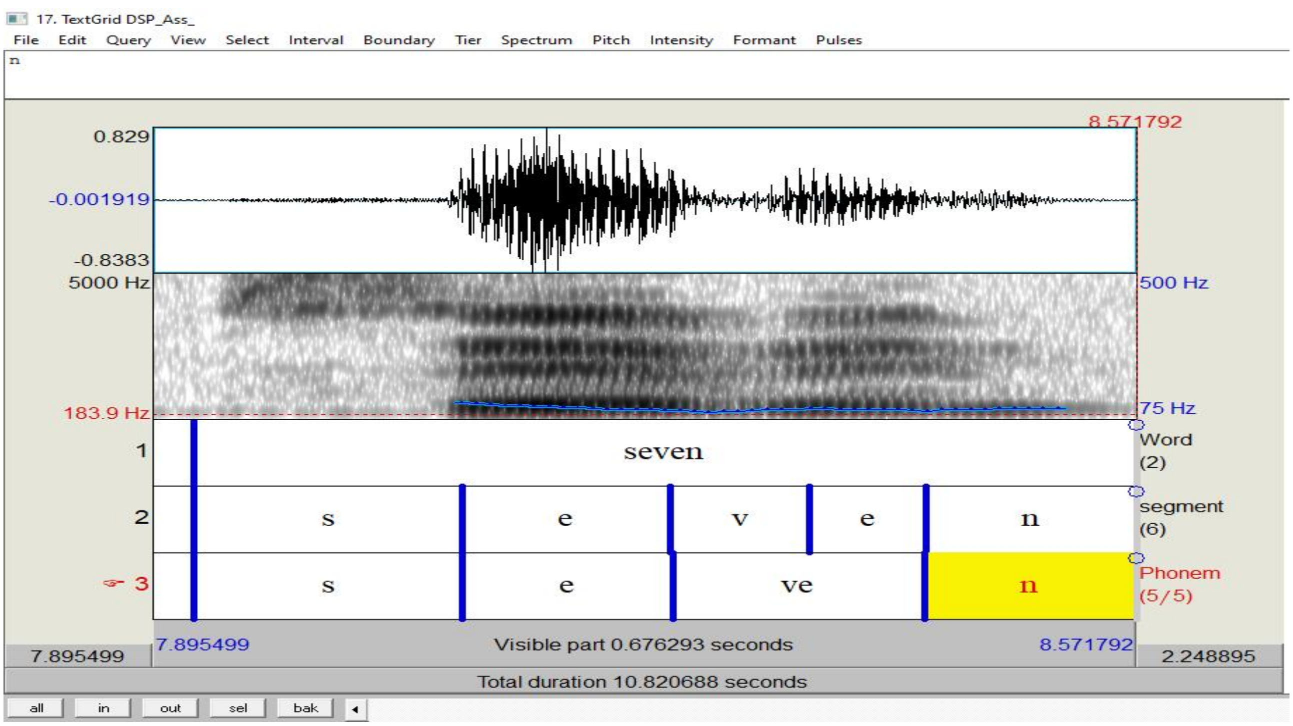


Figure 18 Pitch of Seven

- **Eight**

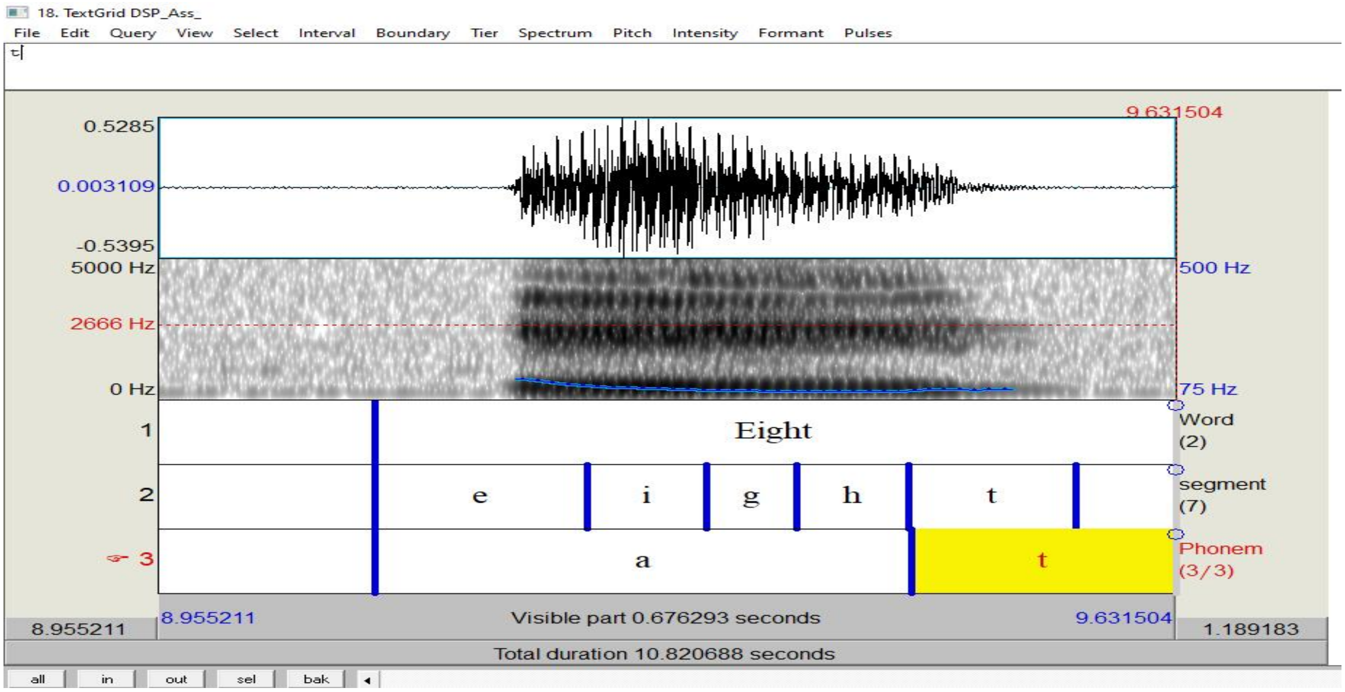


Figure 19 Pitch of Eight

- **Nine**

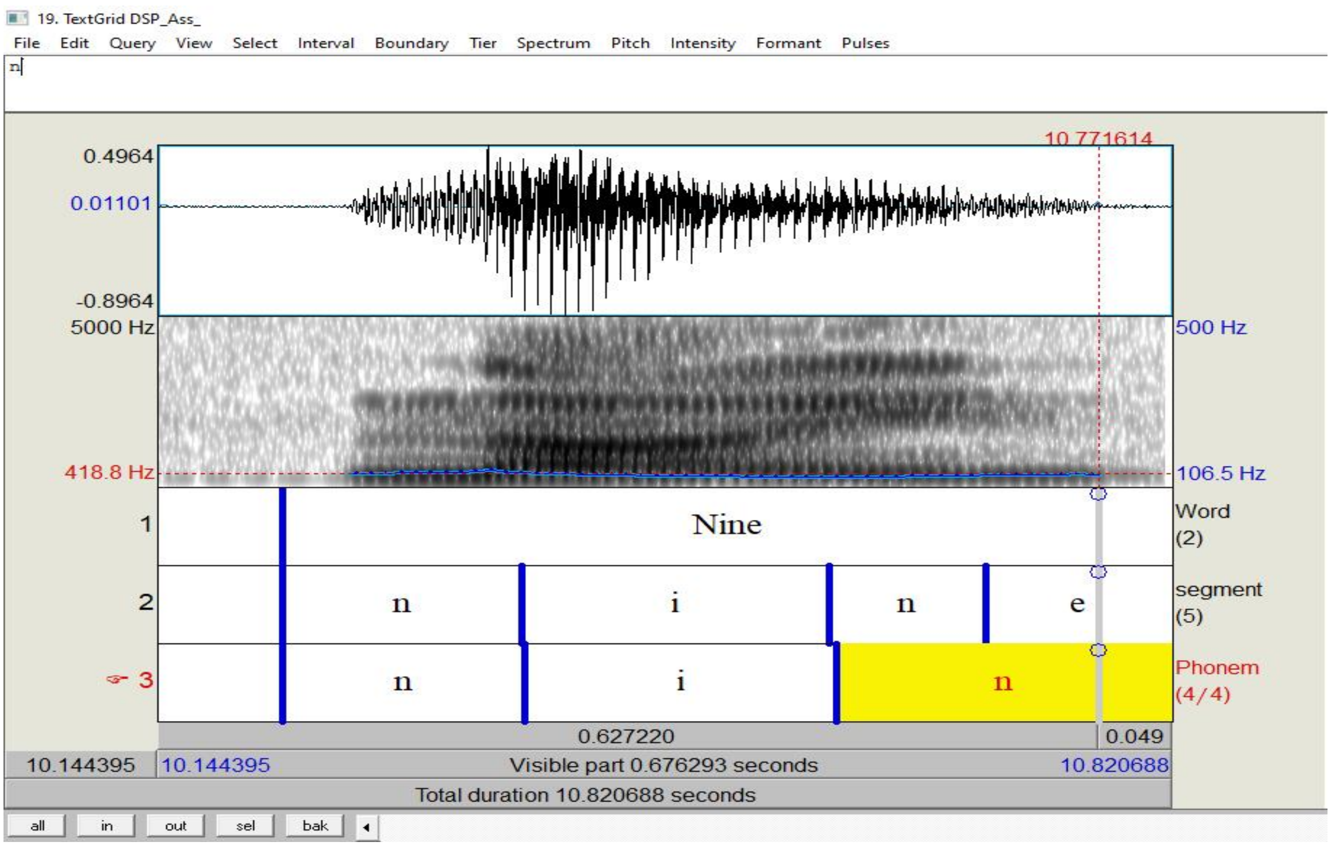


Figure 20 Pitch of Nine

- **Zero**

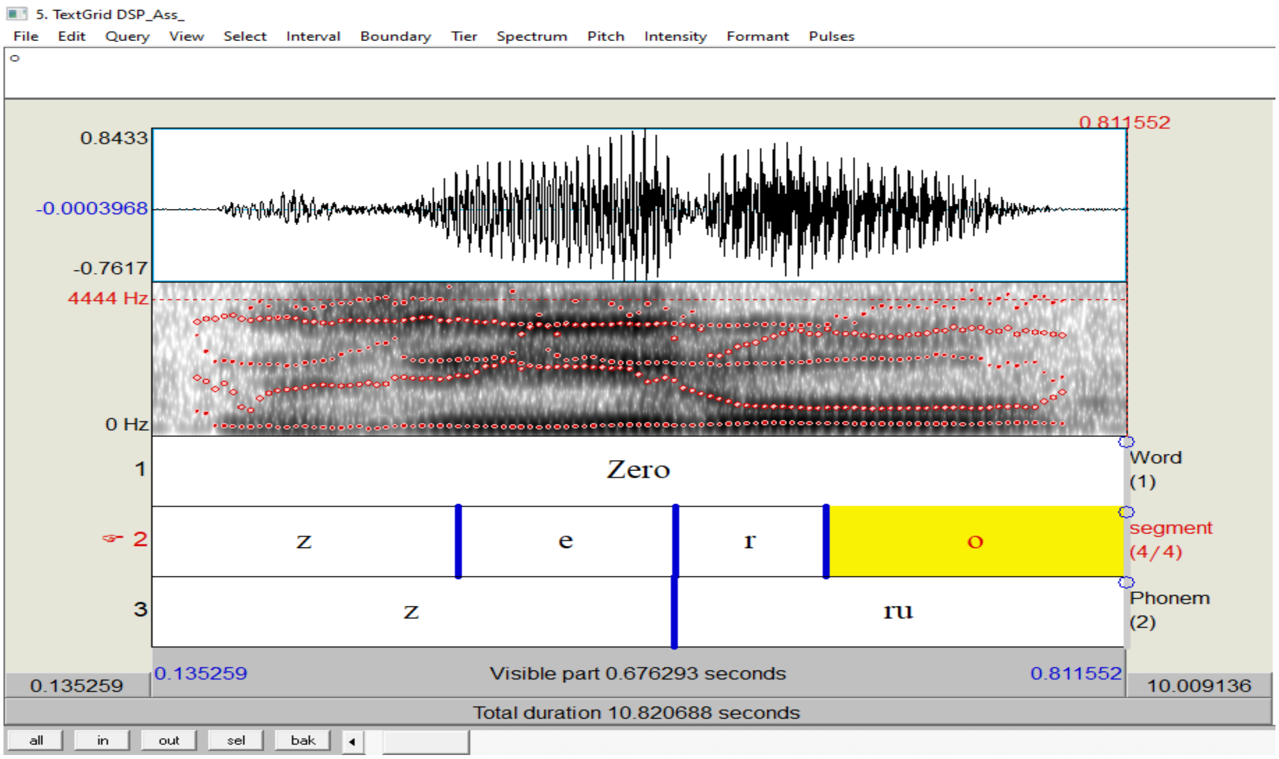


Figure 21 Formants of Zero

- **One**

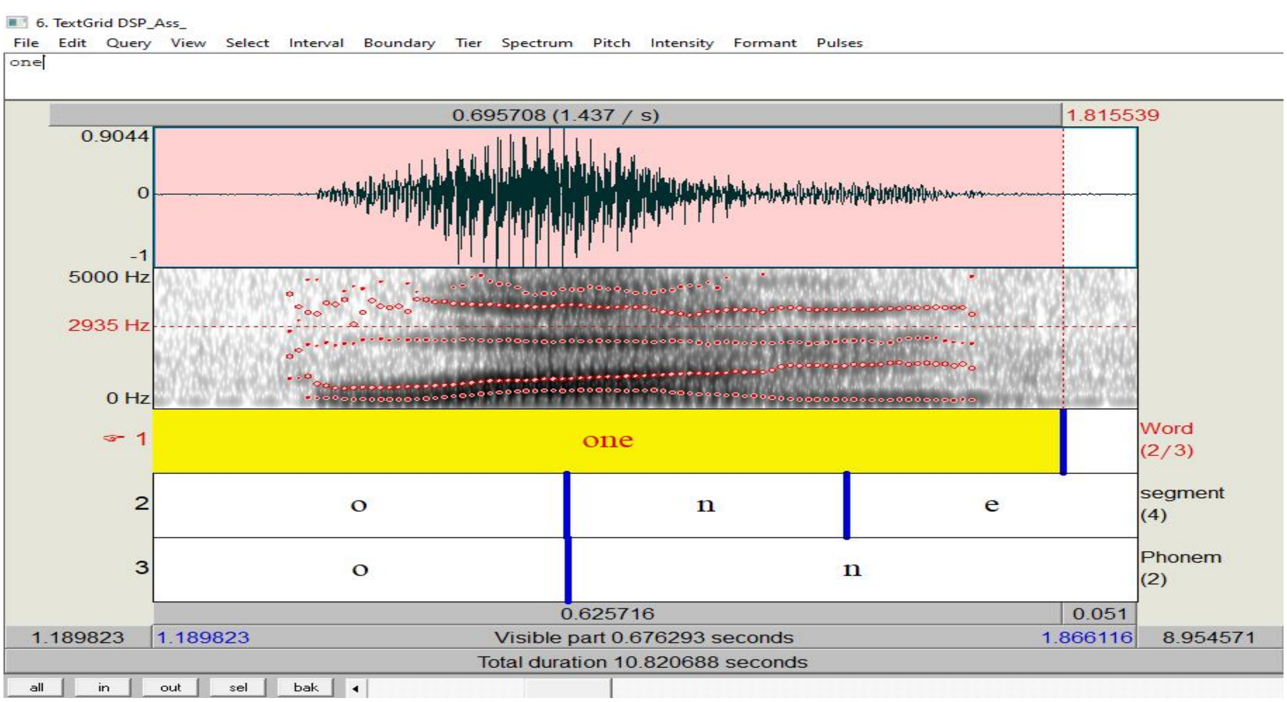


Figure 22 Formants of One

- **Two**

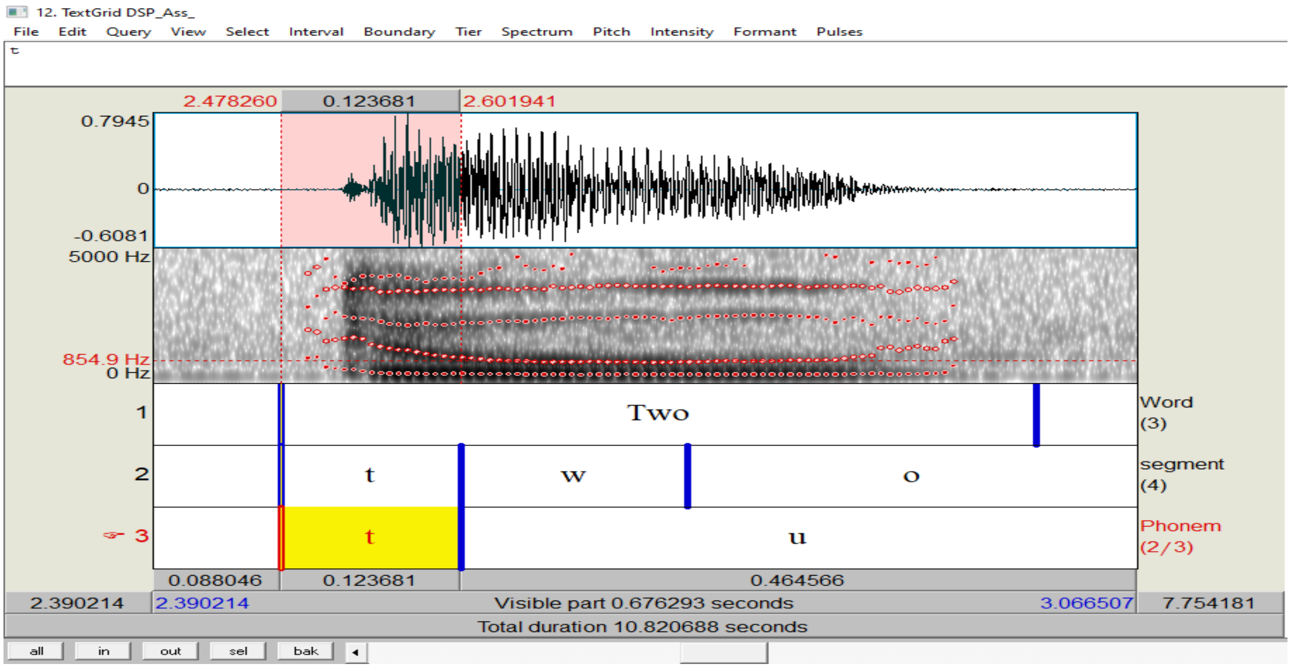


Figure 23 Formants of Two

- **Three**

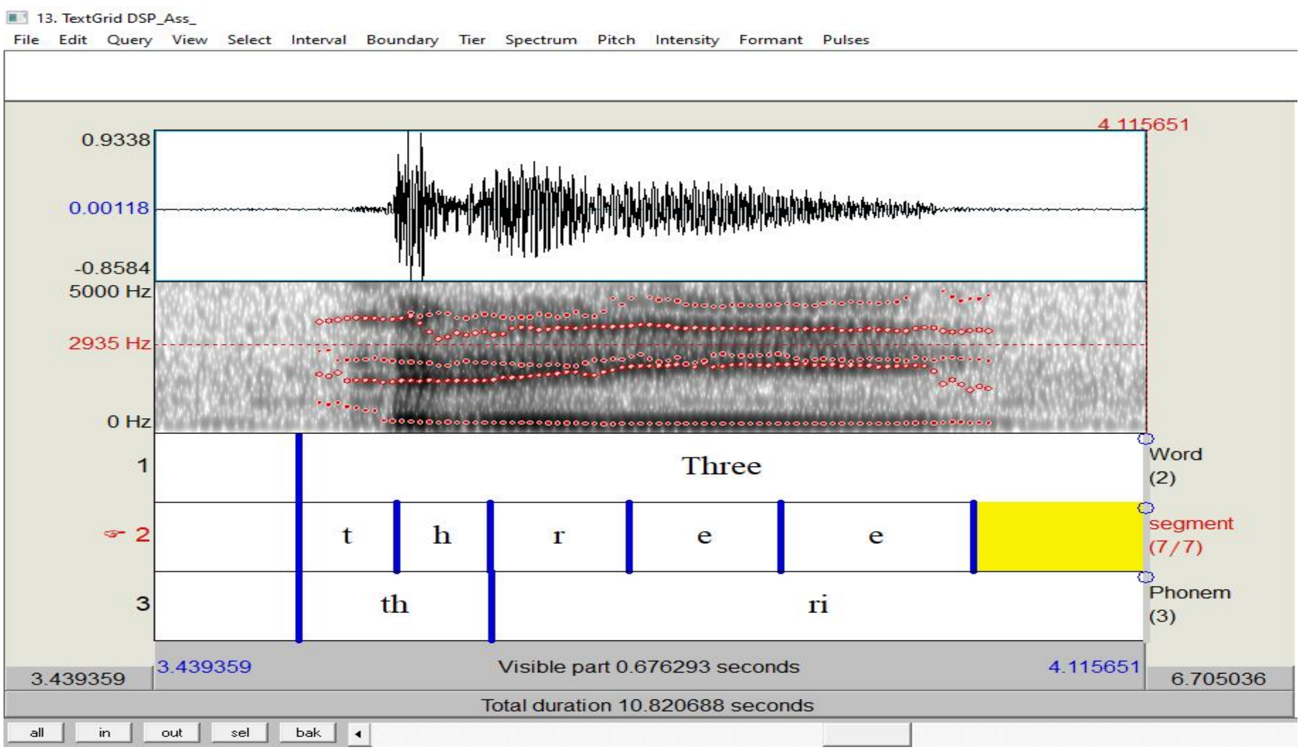


Figure 24 Formants of Three

- **Four**

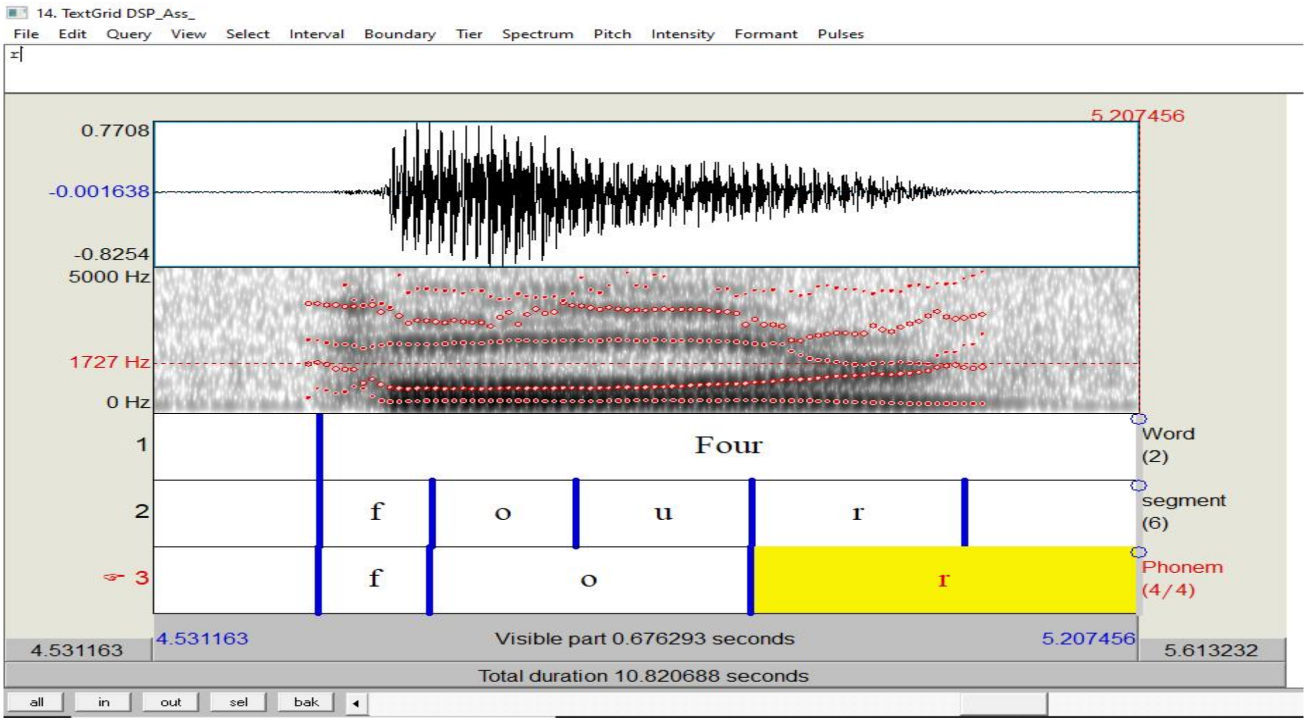


Figure 25 Formants of Four

- **Five**

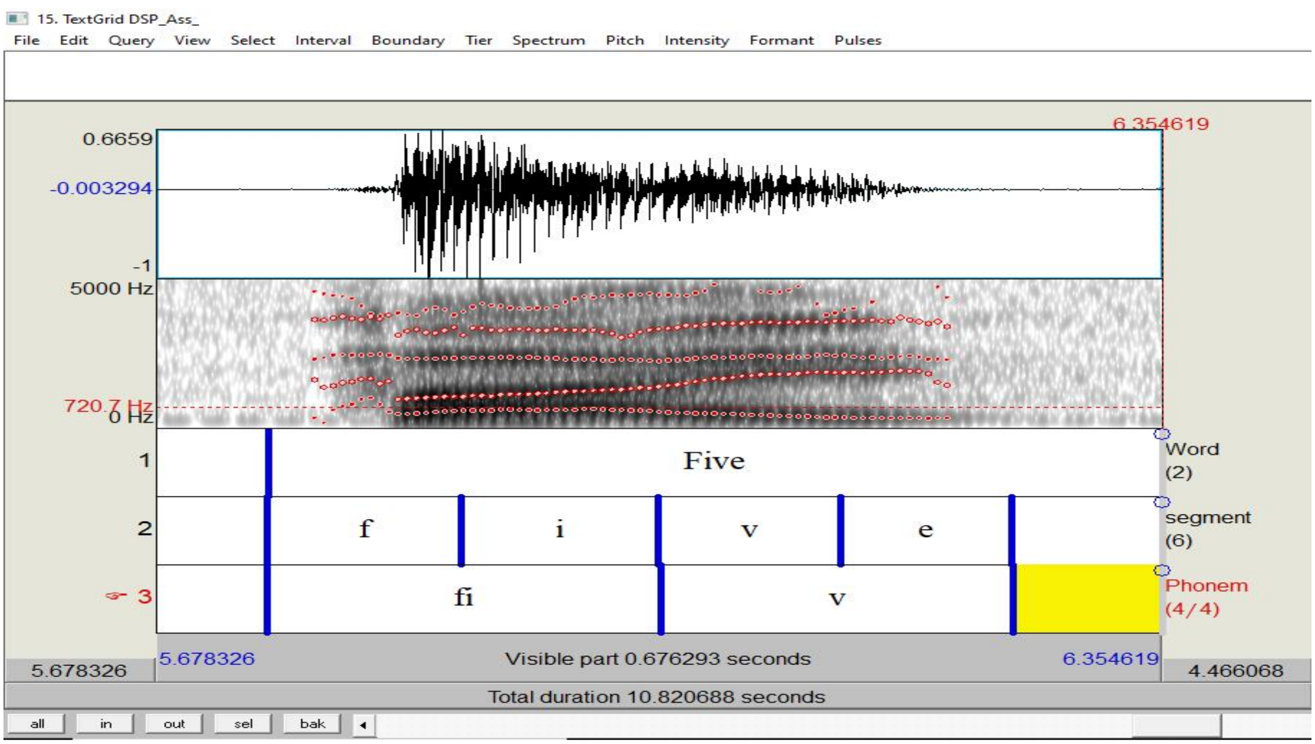


Figure 26 Formants of Five

- **Six**

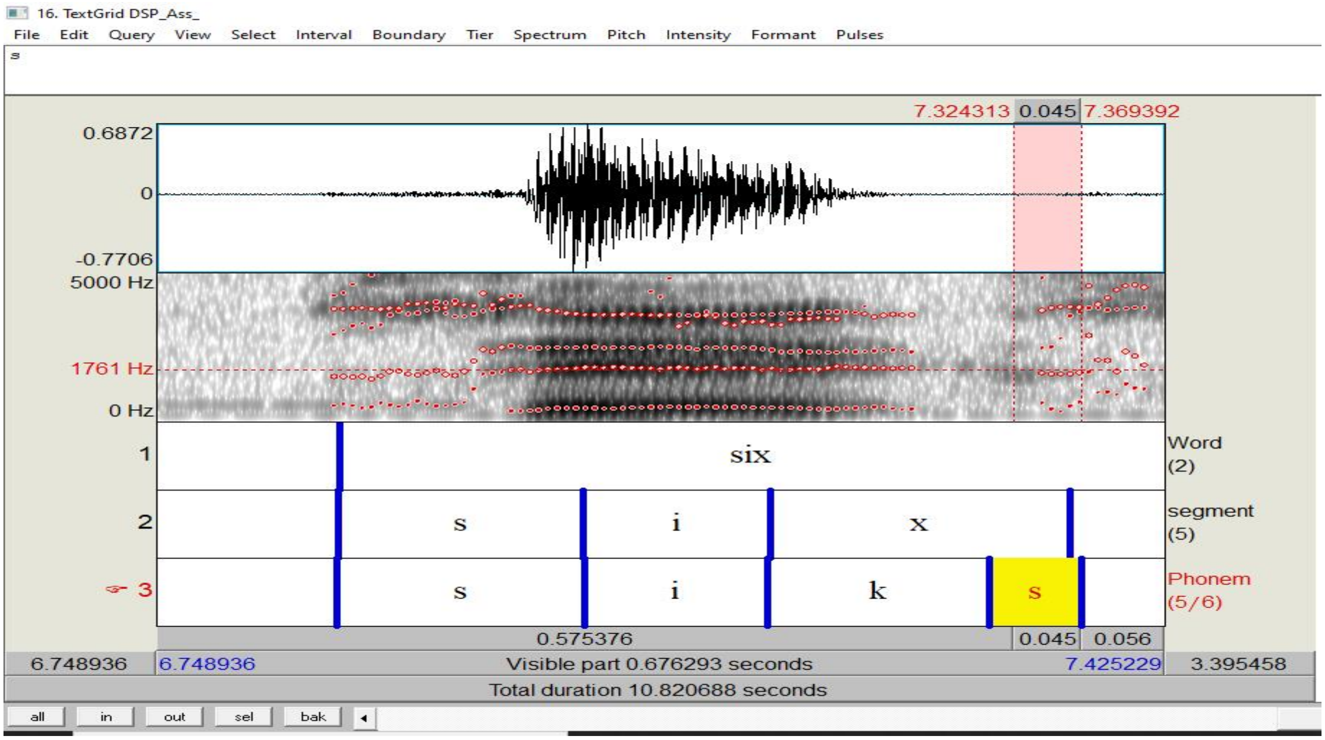


Figure 27 Formants of Six

- **Seven**

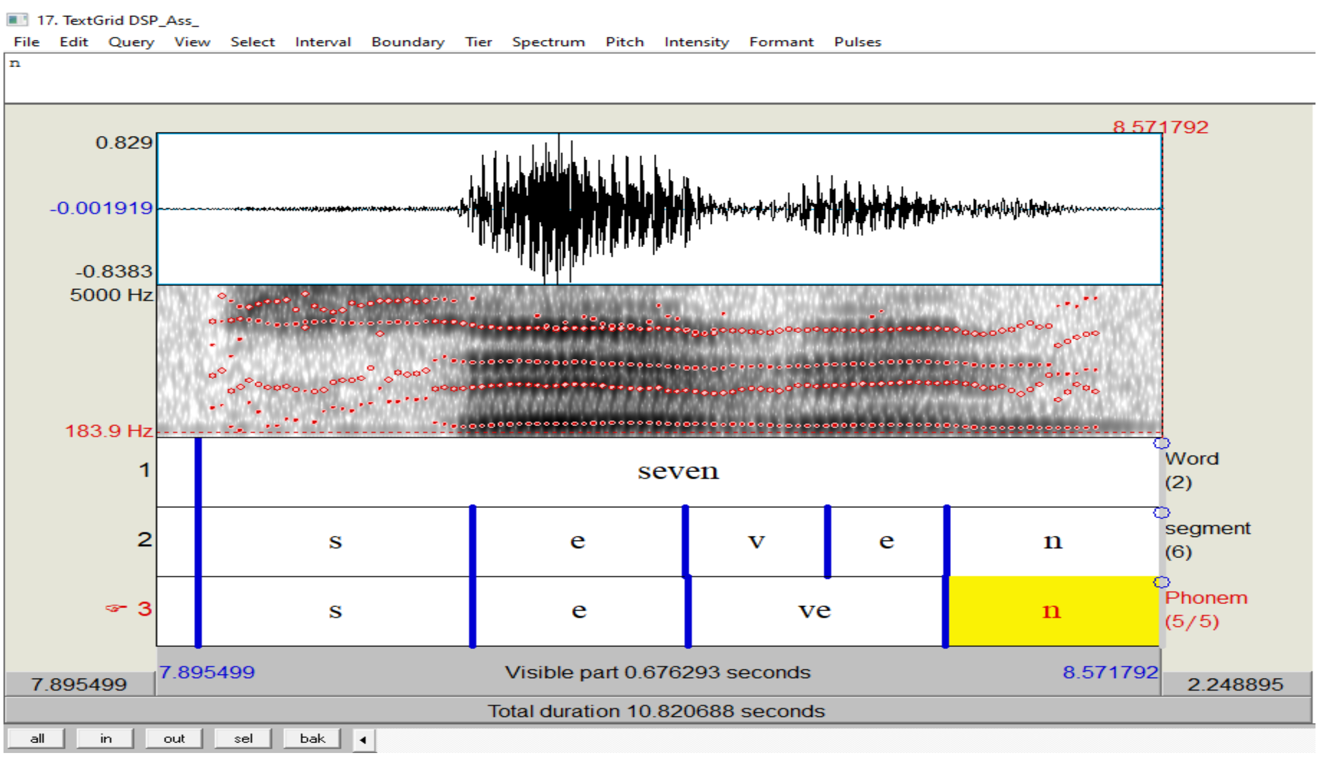


Figure 28 Formants of Seve

- **Eight**

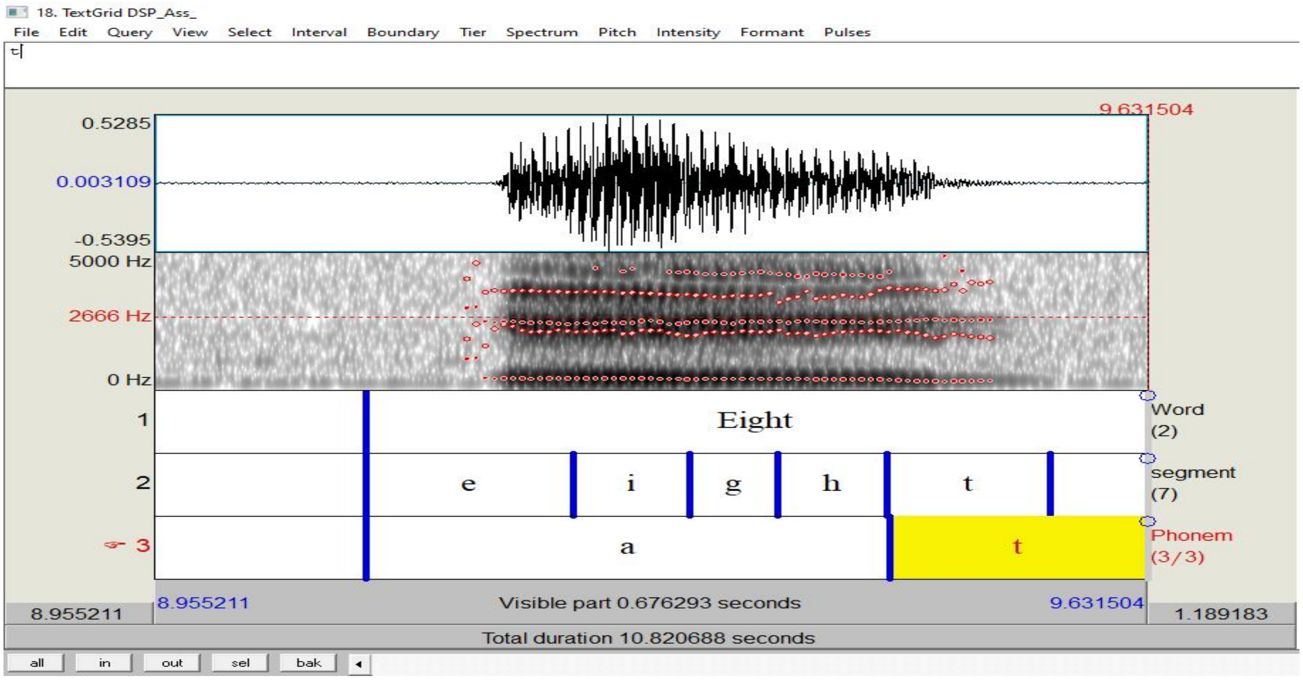


Figure 29 Formants of Eight

- **Nine**

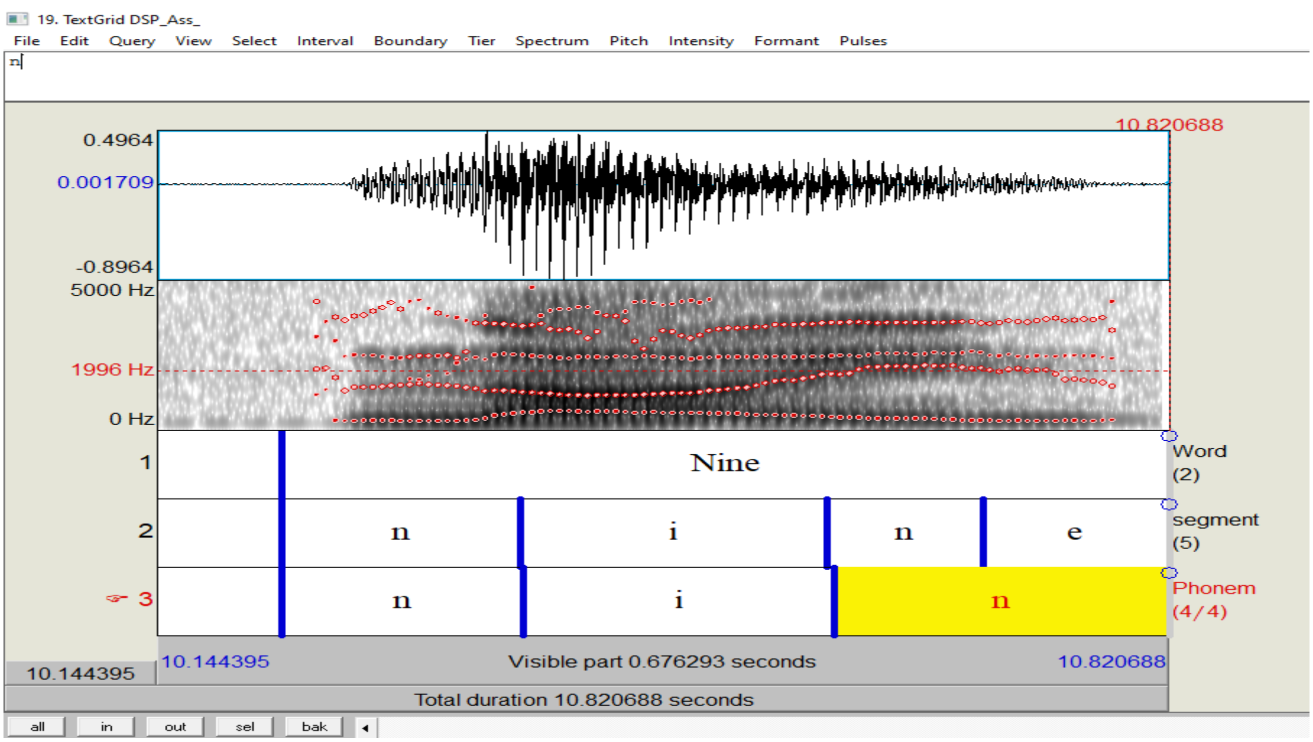


Figure 30 Formants of Nine