

AUDIO-VIDEO FORENSICS

**B.V.S. Siva Prasad
Assistant Director
APFSL, MANGALAGIRI**

FORENSIC SPEAKER IDENTIFICATION AND VOICE AUTHENTICATION

INTRODUCTION

Voice is frequently encountered as evidence in cases of kidnapping for ransom, threatening calls, anonymous calls, hoax calls, drug pedalling, exchange of information across the border, trap investigation etc.,.

The scientific investigation of the recorded voice collaborates scientific evidence in respect of speaker identification.

HISTORY

- The history of speech sound analysis goes back a little more than one hundred years to Alexander Bell who developed a visual representation of the spoken word.
- This visual display of the spoken word conveyed much more information about the pronunciation of that word than the dictionary spelling could ever suggest.

- It was in the early 1940's that a new method of speech sound analysis was developed.
- Potter, Kopp & Green working for Bell Laboratories in Murray Hill, New Jersey began work on a project to develop a visual representation of speech using a sound spectrograph.
- This machine, an automatic sound wave analyzer, produced a visual record of speech portraying three parameters frequency, intensity and time.

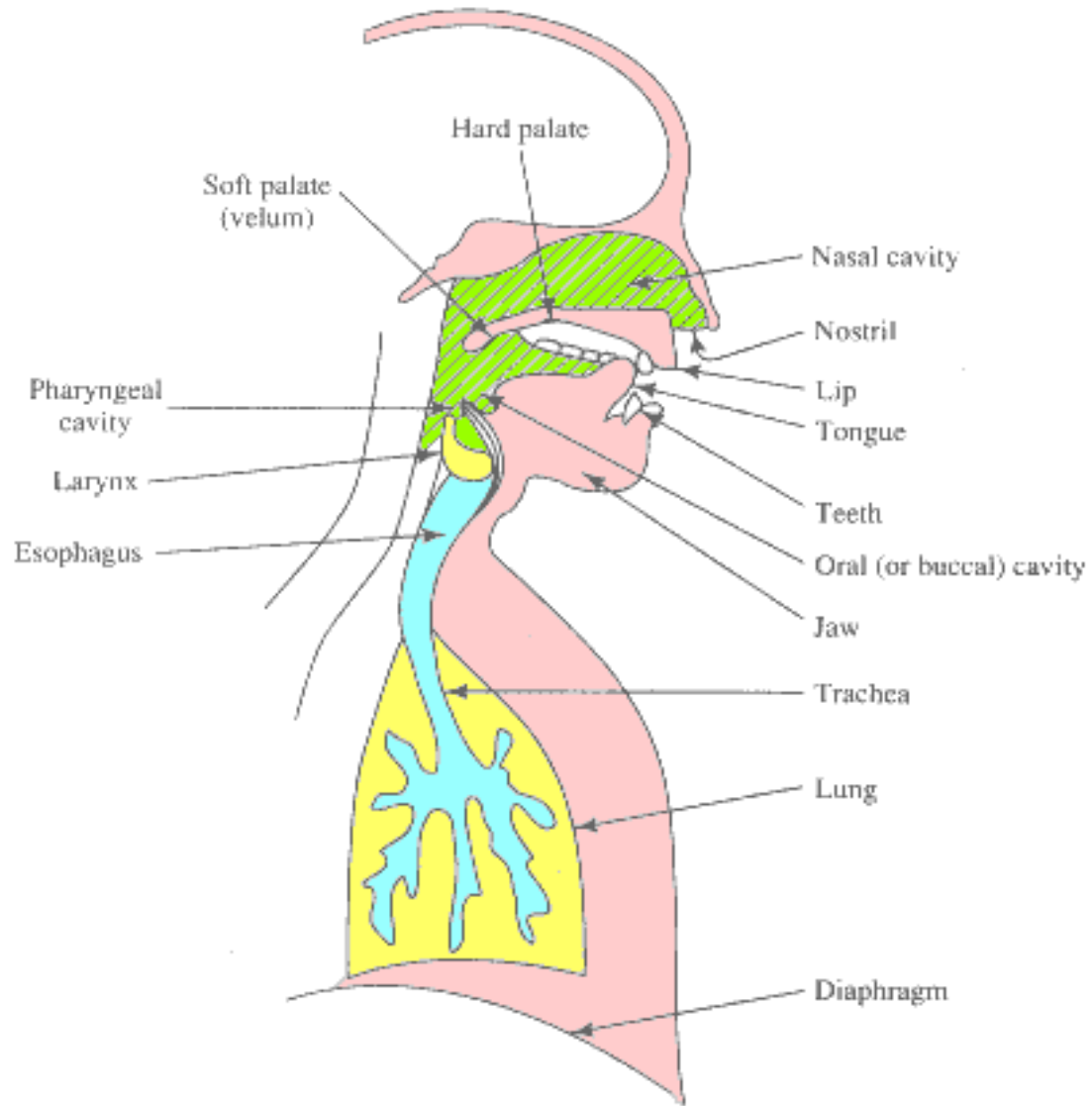
- Today voice identification analysis has matured into a sophisticated identification technique, using the latest technology.
- The research, which is still continuing today, demonstrates the validity and reliability of the process when performed by a trained examiner using established, standardized procedures.

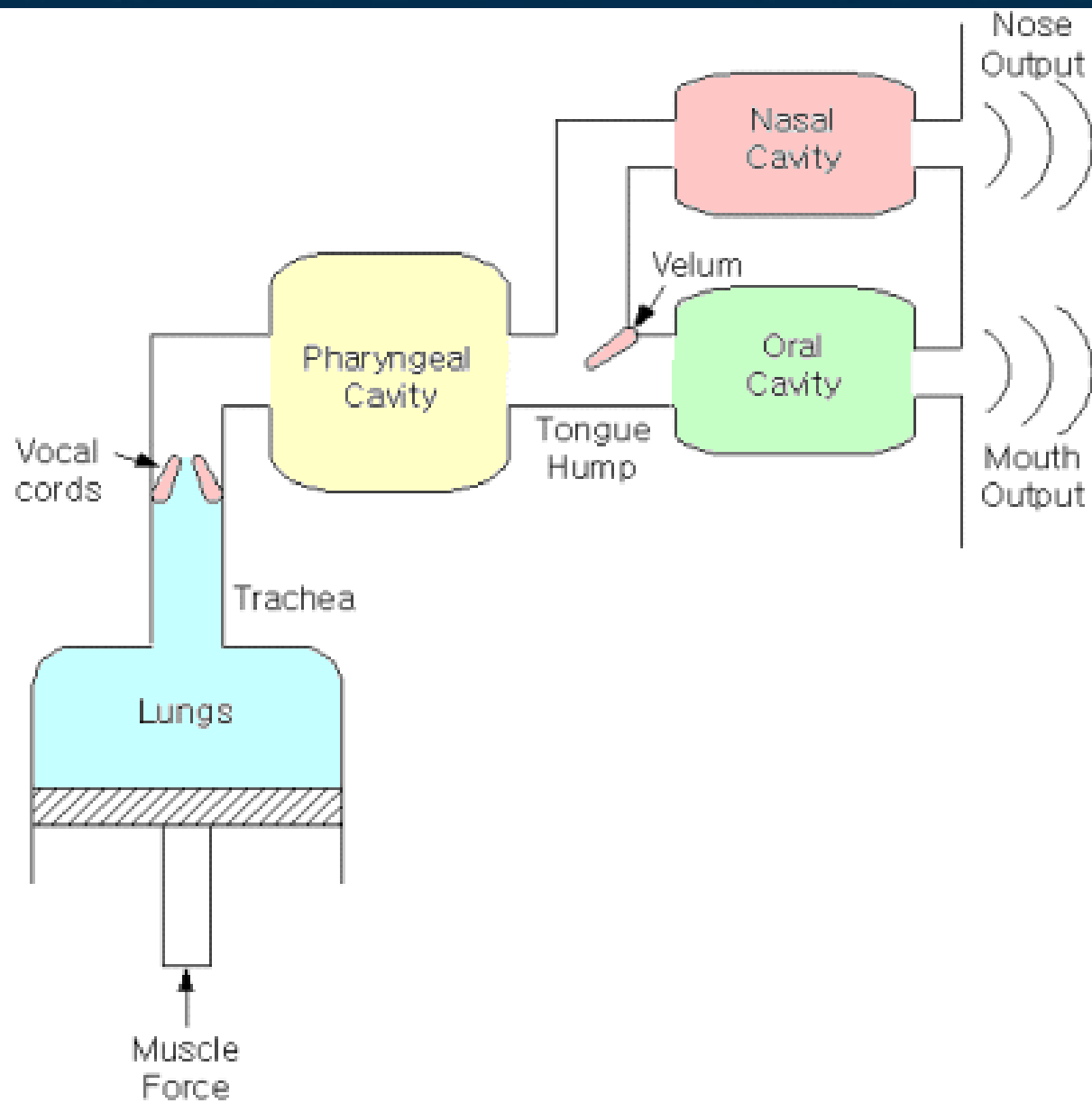
- The comparison of human voices now focuses on every aspect of the words spoken; the words themselves, the way the words flow together, and the pauses between them.
- Both aural and spectrographic analysis are combined to form the conclusion about the identity of the voices in questioned.

Basic Elements of Speech Production

Human organs that combine to produce intelligible speech are:

- The lungs & windpipe
- The larynx
- Upper throat
- The mouth
- The nasal cavities





- The lungs and windpipe provide the study stream of air, which is the energy base of speech.
- The larynx is a complex set of muscles, including the vocal chords, which controls the airflow from the lungs.
- When a person speaks, the muscles of the vocal chords close and the pressure of air from the lungs pushes through them causes to flutter vocal chords rapidly. This fluttering per second is fundamental frequency or first formant.

- This can be anywhere from 60 Hz to 500 Hz. As this basic vocal tone progresses up through the throat, mouth and nasal cavities, it is modulated by an extremely complex combination of muscular actions, that produce the resonance, harmonic overtones and stops & starts the characterized intelligible speech.
- The harmonic frequencies superimposed on the fundamental frequency can go up to as high as 5000 Hz.

Human Voice Frequency Ranges

- Voice frequency range 20 Hz – 14,000 Hz
- Human ear perception range 20 Hz – 20,000 Hz
- Generally vowel sounds below 1000 Hz
- Consonants sounds range above 1000 Hz
- Most significant speech of human beings is 100 Hz – 4000 Hz

Human Voice Frequency Ranges

- Fundamental frequency range of adult male is around 120 Hz (ranges from 80 Hz – 200 Hz)
- Fundamental frequency range of adult female is around 220 Hz (ranges from 165 Hz – 255 Hz)
- Fundamental frequency range of Children is around 250 Hz – 500 Hz
- Usable telecommunication range 100 Hz – 5000 Hz

Voice identification is based on the theory that:

- Human voice is individual
- The fundamental frequency of each individual is unique
- It is impossible that any two persons have the same configuration of vocal organs or use them exactly the same way.
- the distinctive environment influences in the learning process of speech development.
- the unique development of neurological faculties are also responsible for the production of speech.

METHOD OF SPEAKER IDENTIFICATION

- ❑ SEMI AUTOMATIC METHOD
 - a. Auditory examination (listening)
 - b. Spectrographic examination
- ❑ FULLY AUTOMATIC SPEAKER IDENTIFICATION

Speaker Identification

- Original Questioned audio file
- Transcription of Questioned conversation
- Specimen voice in similar word combination
- Original voice recorder (If available)

Role of I.O

- The technique of speaker identification requires two types of samples namely questioned sample (crime sample) and specimen samples (control samples of suspect / suspects).
- The following are the guidelines for recording specimen voice samples for the purpose of speaker identification.

- An accurate transcript of the text of the questioned call / conversation should first be made.
- I.O should become familiar with the transcript (the rate of speaking and oral characteristics of the criminal's voice) prior to record control samples from suspects.
- If the questioned recording of the criminal was obtained through a telephone then the specimen recording should also be made through telephone (Direct recording can also be made).

- The recording of the specimen sample should be performed in presence of two independent witnesses. The witnesses should speak about their particulars like name, father's name, residential locality, occupation etc., and he also should speak that "I Mr. Y, Son of Mr. A, resident of B locality, in my presence Mr. X is giving his sample voice voluntarily.
- In case, suspect is not educated and cannot read the prepared text / transcript, the sample can be recorded by making the conversation with the suspect. The conversation should be prolonged in such a manner that the relevant words are repeated number of times. The similar text in the same language should be prepared for recording the specimen samples.

- The recording should be repeated as many times as felt necessary.
- A reasonably quiet environment should be maintained.
- Make the suspect conversant with the specimen speech before the recording starts.
- The speaker should be at a distance of about one foot from the microphone and should speak with normal speed and loudness. It is an essential requirement to stabilize the speech of the speaker for which he may be asked to speak continuously for two minutes before he reads the prepared text. The speaker should be directed not to speak too fast or too slow but he should read/speak normally.

- The I.O should make all efforts to eliminate background noise as much as possible by not playing radio, TV and Air conditioners and fans. Also should avoid overlapping conversations.
- In the event of suspect disguising his voice, the I.O. should ask for the repetition of clue words, until he satisfies.
- Recording should be played back before the suspect leaves; so that any deficiency of the sample can be corrected.

■ **Precautions:**

- The recording system should be tested prior to the actual recording of the samples.
- Original recordings should be properly labeled.
- Questioned digital storage media should be properly labeled.
- Questioned and specimen samples should be on separate digital storage media with proper marking and packing.
- Digital storage media should be kept away from "Magnets".
- Irrelevant portion of the recording should not be deleted by the I.O rather than it should be pointed out.

I. Auditory Examination

- The auditory analysis is carried out through repeated careful listening using good quality audio equipments.
- The expert audio-listener should have sensitive ear perception and knowledge of speech language.
- The audio equipments should be played in the normal speed in which the voice have been recorded.

- The repeated listening helps in study of the anatomical linguistic features, phonetic and physiological features of the speaker and also of emotional, behavioural and the intentional approach of the speaker.

- The inter speakers are differentiated by the attributes of voice production machine, language of speech and manner of speaking.
- The auditory analysis reveals dynamics of speech such as dialect, pitch, articulation, general voice quality, intonation, stress/emphasis disguise, speech defects etc.,

- In some situations, the speaker makes an attempt to disguise the normal voice during recording of the specimen voice sample.
- The presence of emotional and psychological tensions at the time of recording likely to change/shift some of the spectrographic features from the normal.

- The use of good quality listening instrument helps the examiner to assess the apparent age of speaker, dynamics of loudness, degree of nasality and pathology of voice (whispery, creaky, breath of hoarse).

- The level of stylistic modes in speech, viz. Formal, normal, informal, casual, vulgar utterances etc., can lead to assess overall personality impression of speaker typical to the emotion viz., expressive, pathetic, intellectual, polite, sarcastic, cynic, logical, illogical, chaotic behavioural features etc.,

- After listening to the questioned and standard the auditory features i.e. linguistic and phonetic features (dialect, accent, articulation, vocal quality, stress etc.,) are compared.
- Peculiar speech habits are also scrutinized.

II. Spectrographic Examination

- The sound spectrograph is an automatic sound wave analyzer with a high quality functional recorder.
- The speech samples to be analyzed are recorded onto the sound spectrograph.
- The sound spectrograph gives a 3 – D plot for each word i.e., time on x-axis, frequency on y-axis and darkness of the plot depicting the intensity.

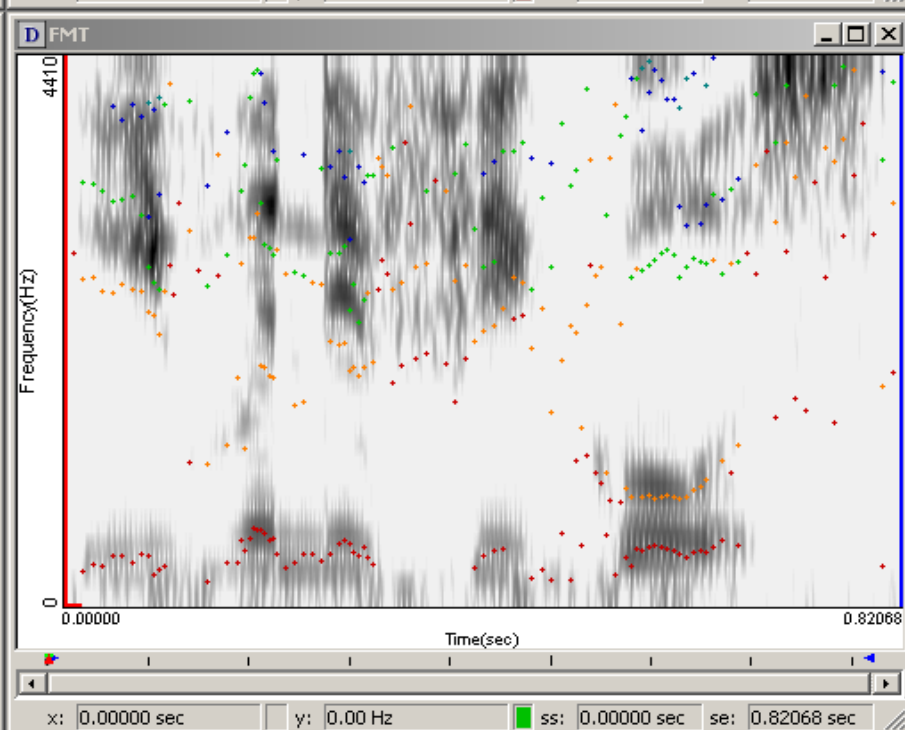
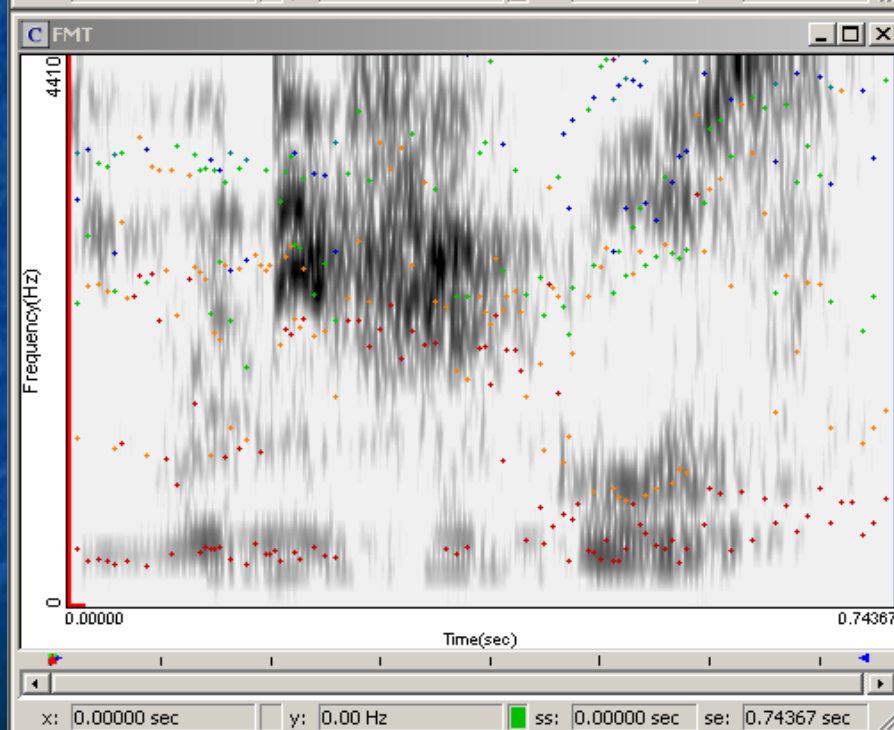
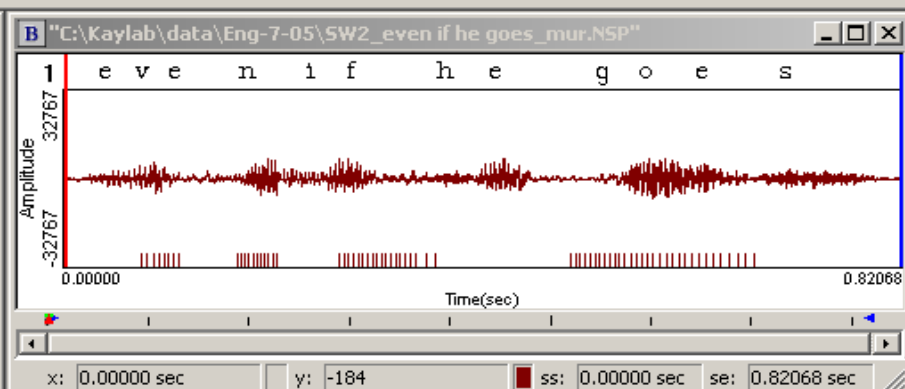
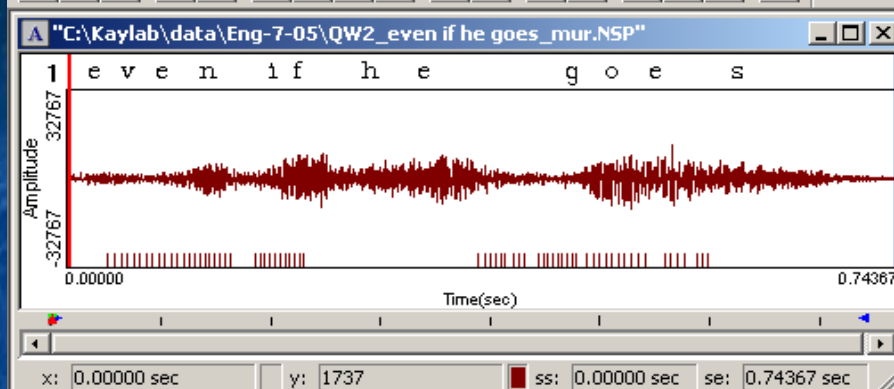
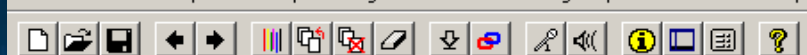
Tehalca.com

Questioned



Specimen

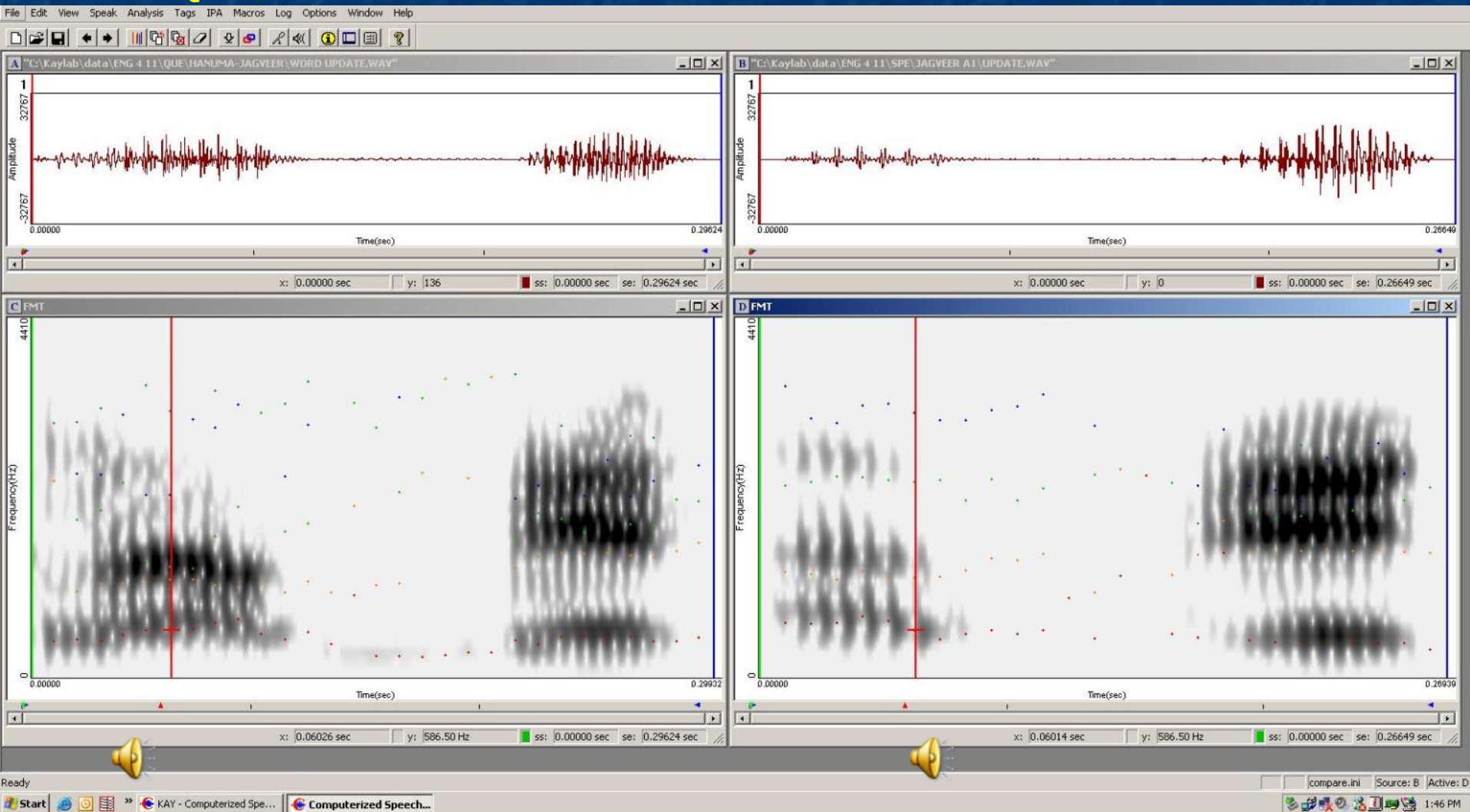




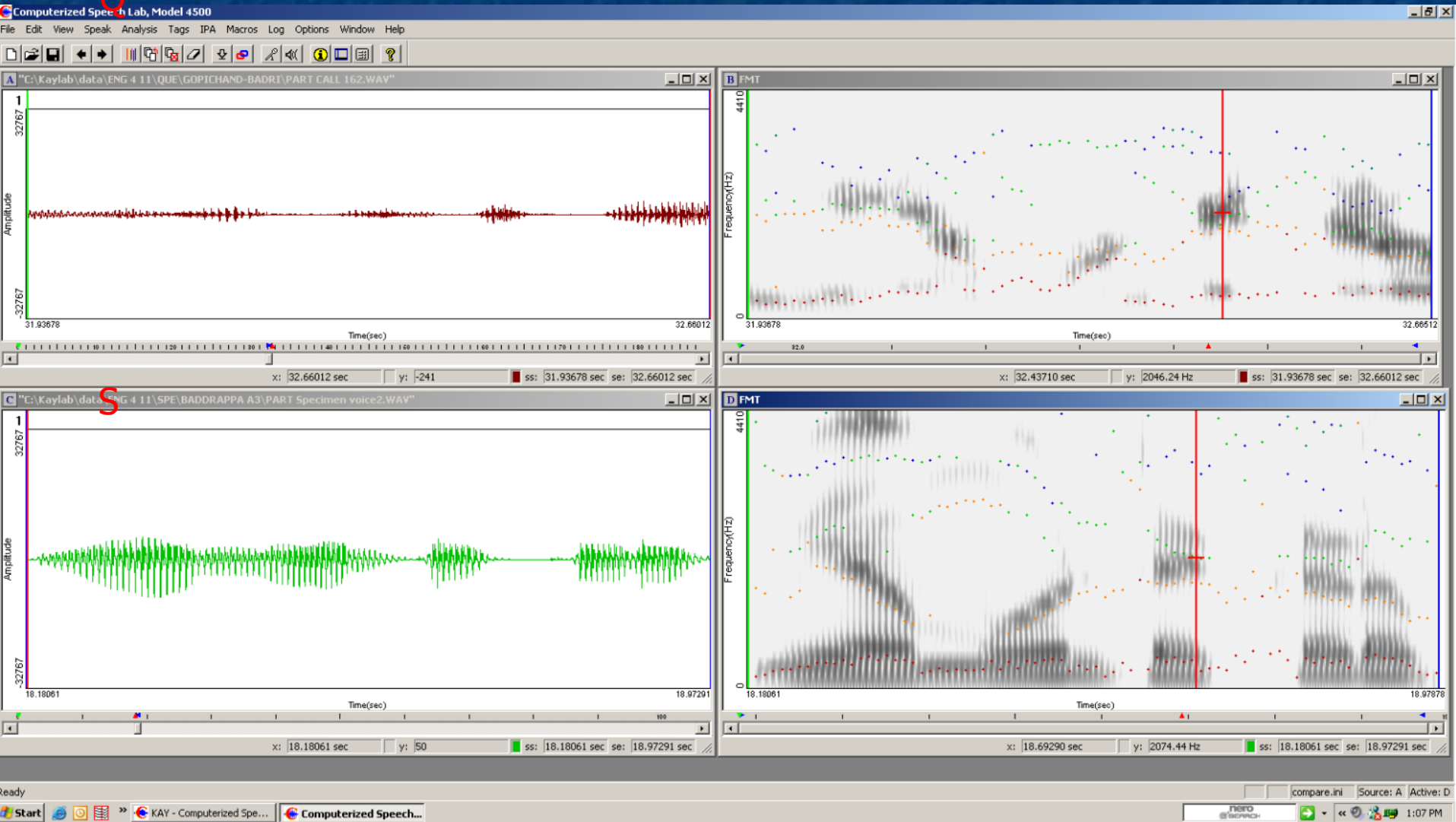
Q

UPDATE

S



AMOUNT – A2



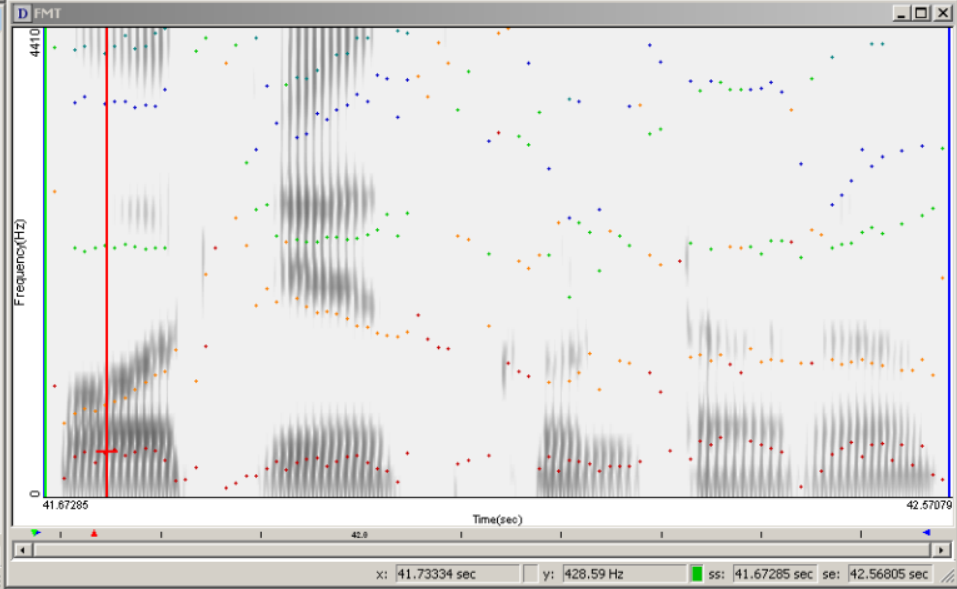
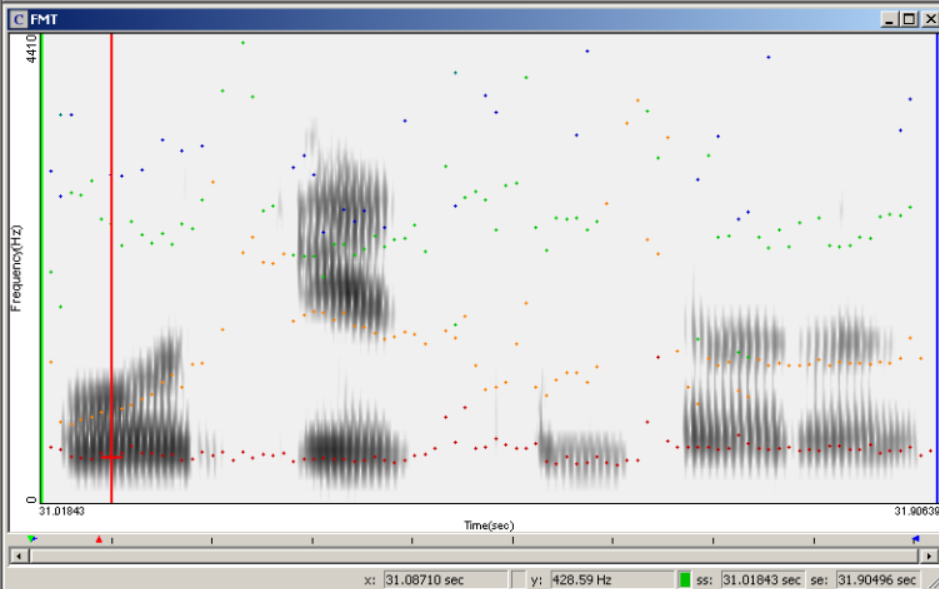
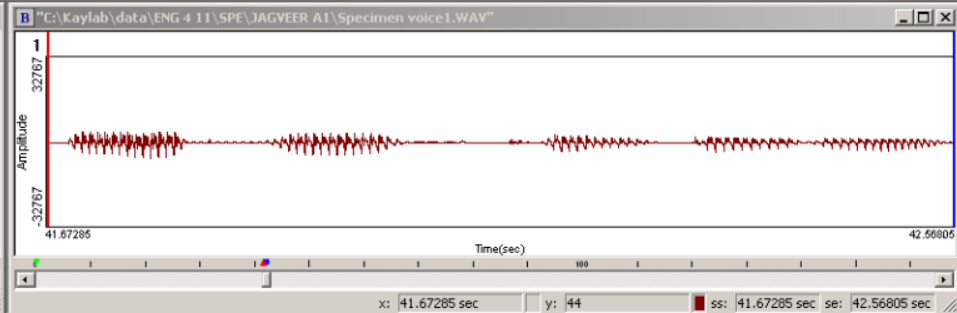
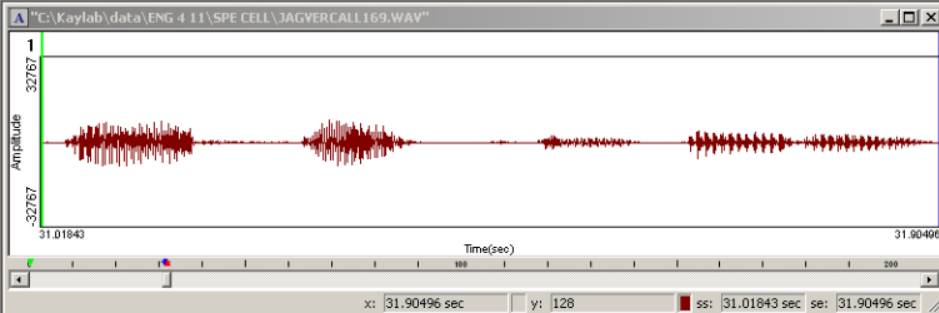
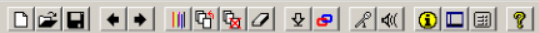
Q

NALUGU CHESUKONDI – A1

S

Computerized Speech Lab, Model 4500

File Edit View Speak Analysis Tags IPA Macros Log Options Window Help



Ready

Start D VENKAT (G:) KAY - Computerized Spe... Computerized Speech...

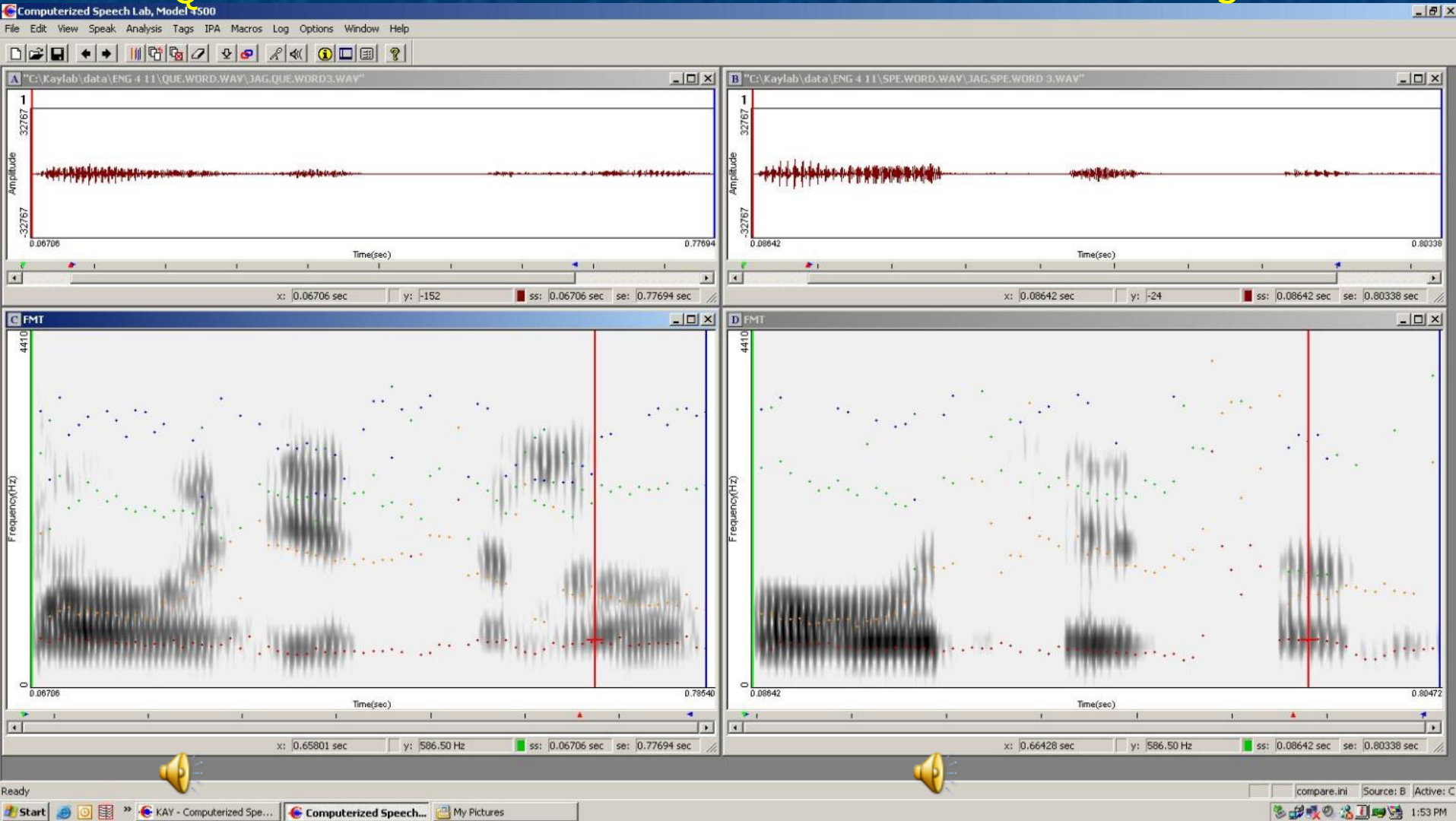
compare.ini Source: B Active: C

3:28 PM

FORWARD CHESTHANU

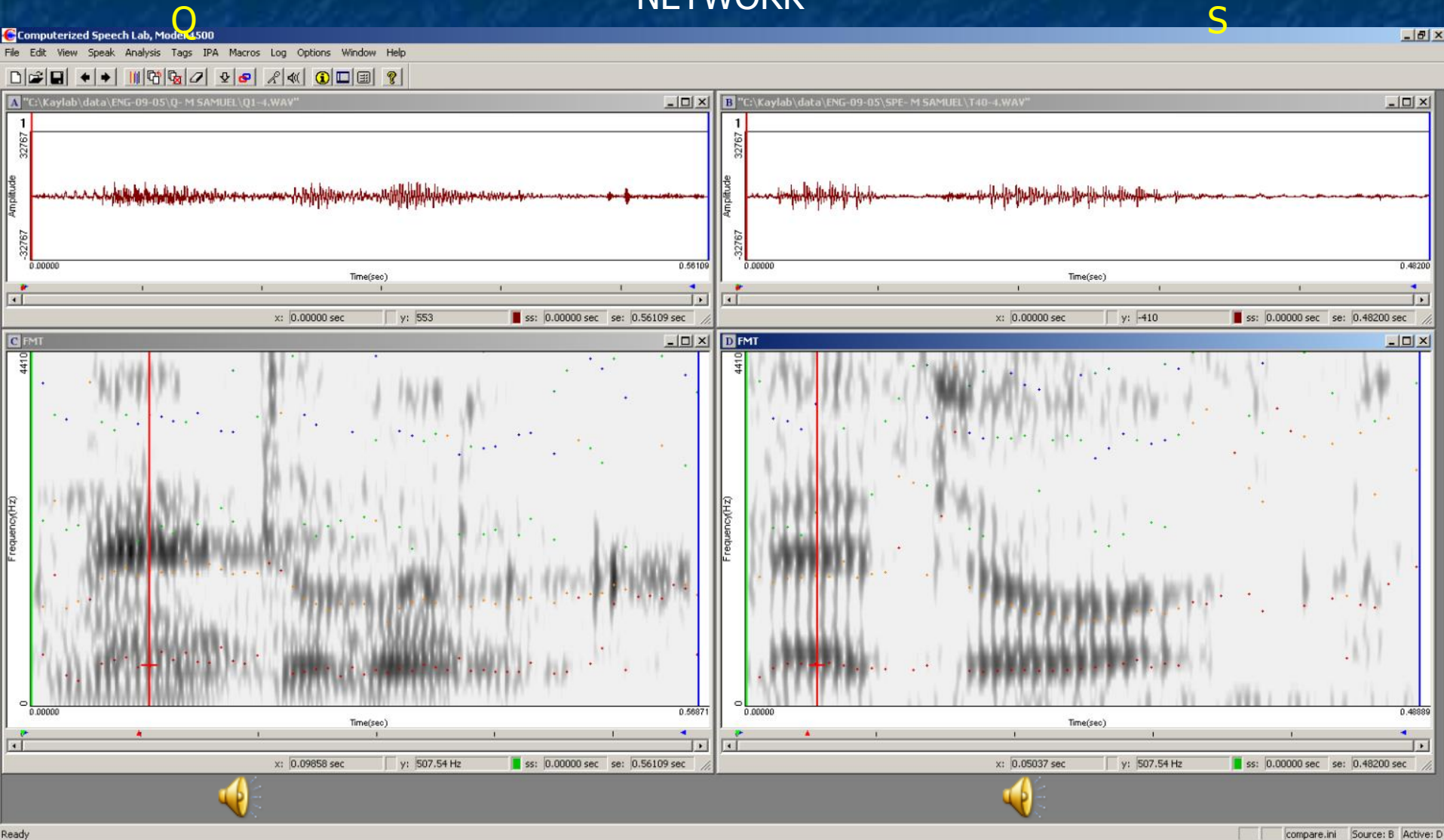
Q

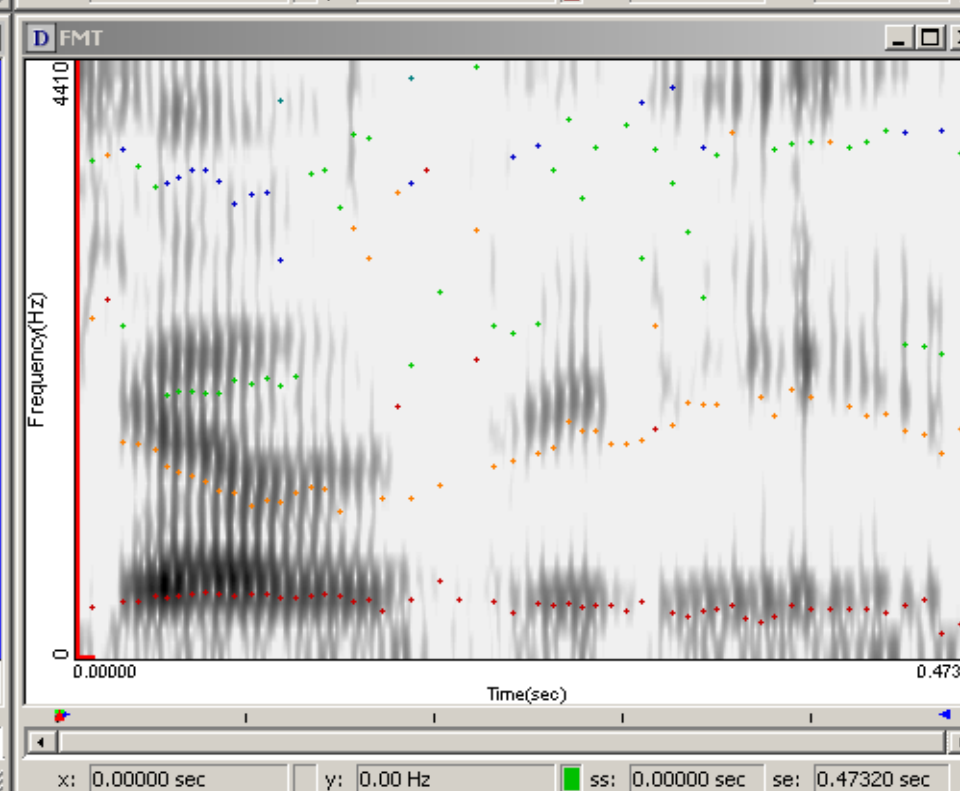
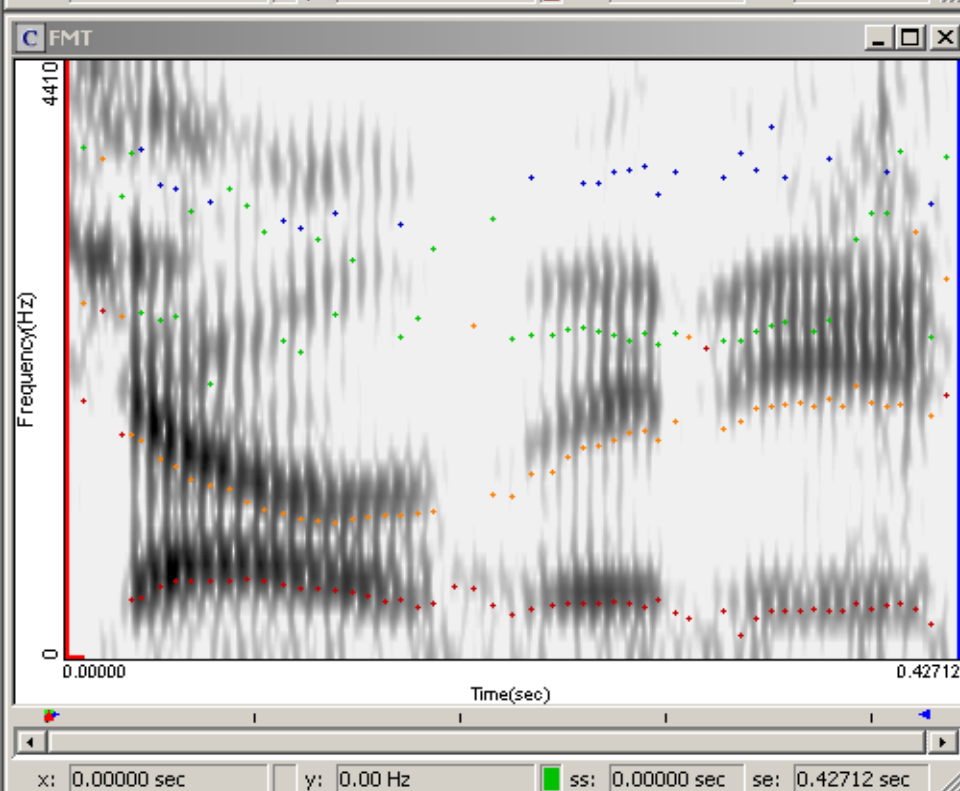
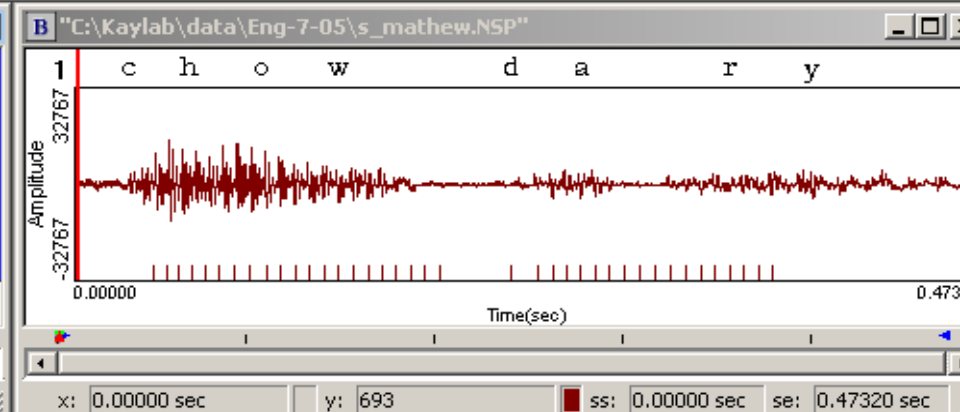
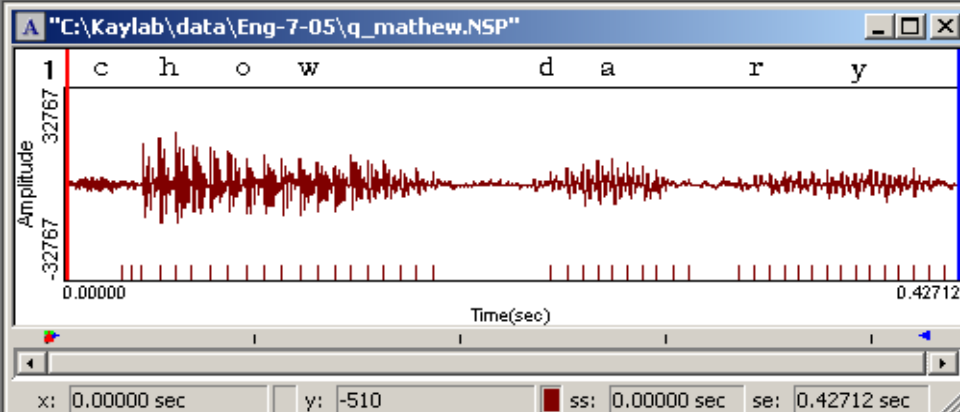
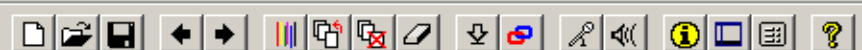
S

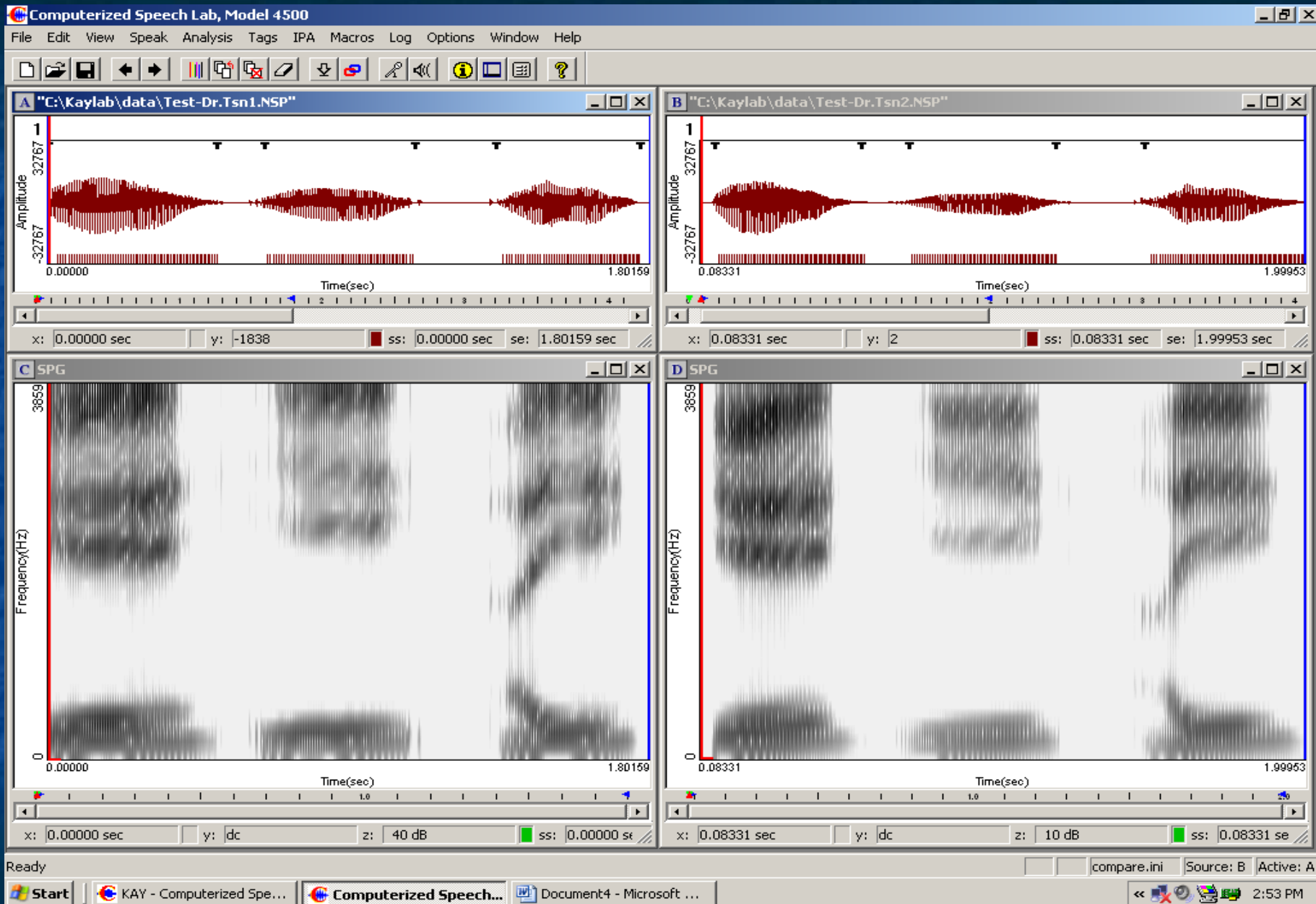


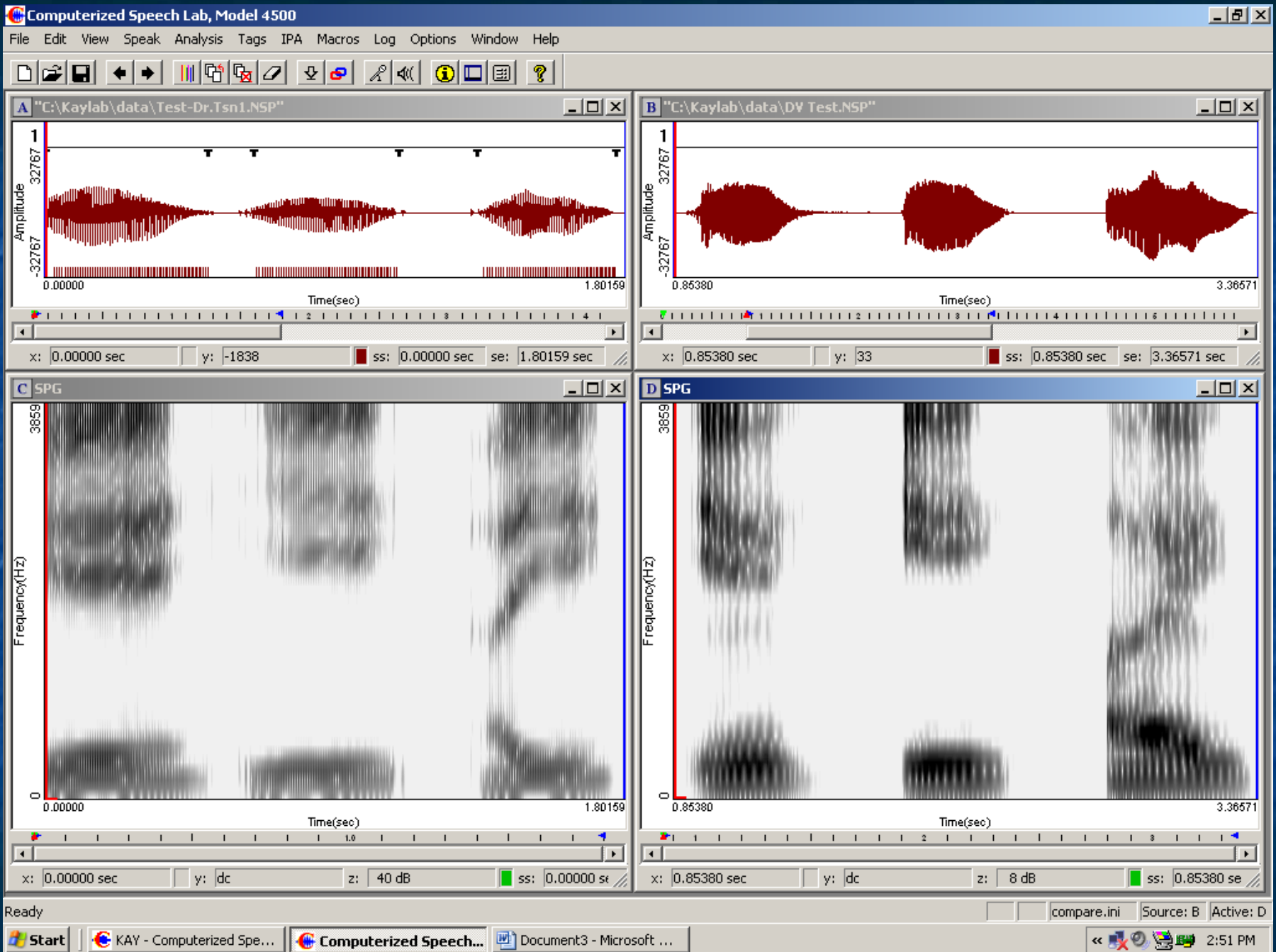
ENG/09/05

NETWORK

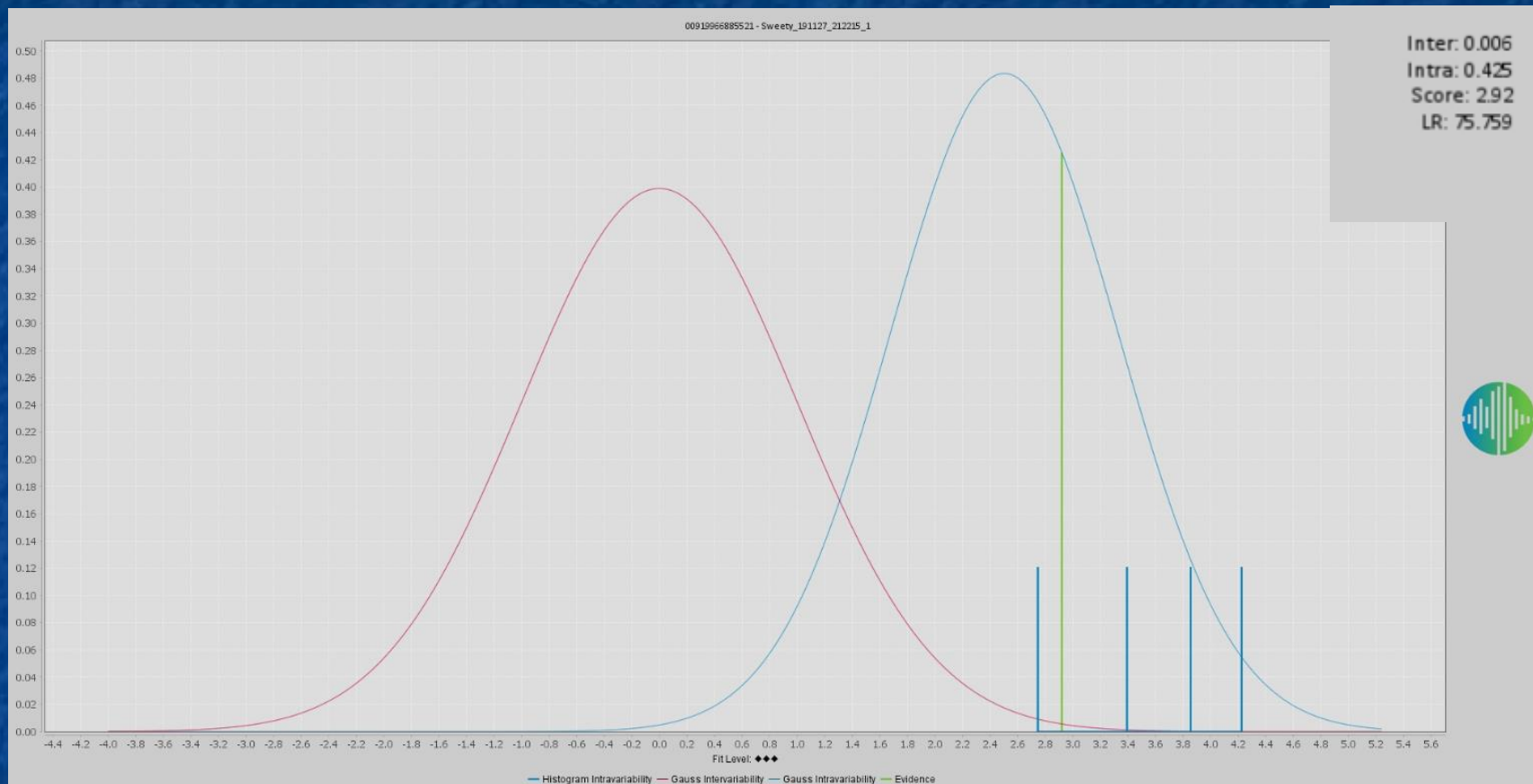




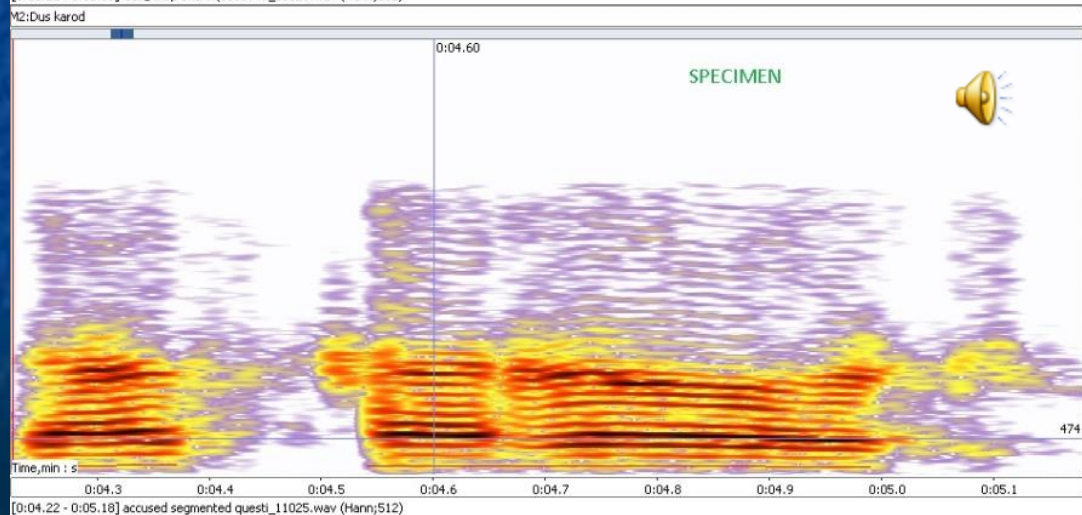
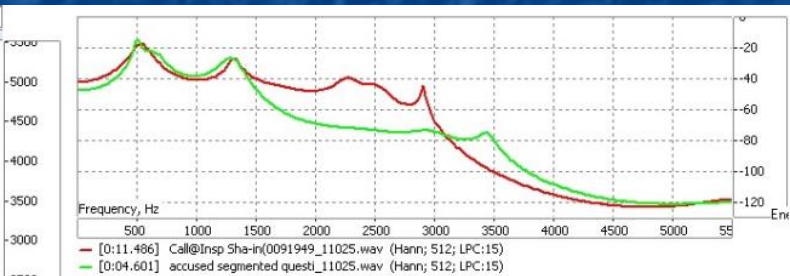
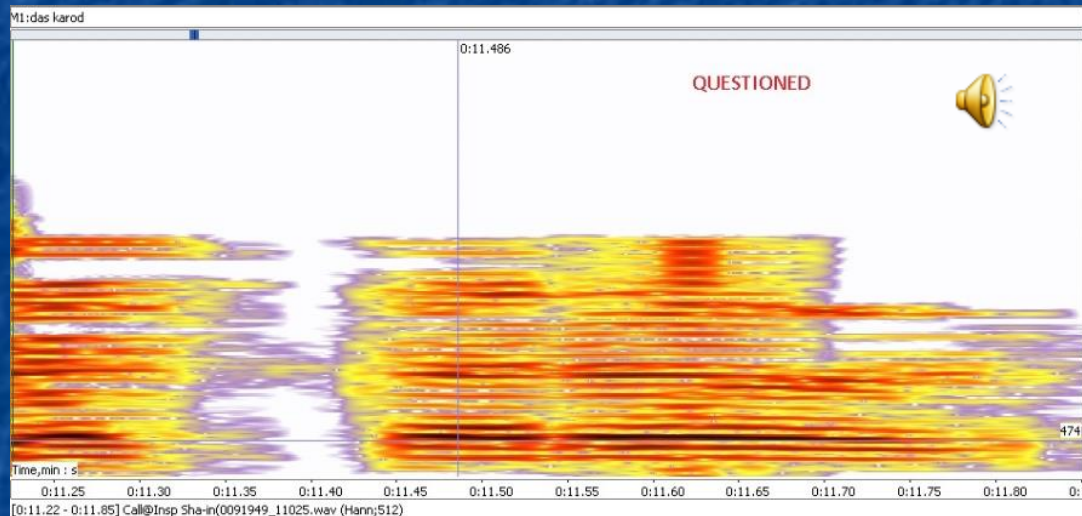




Latest Advancement is Automatic Speaker Identification



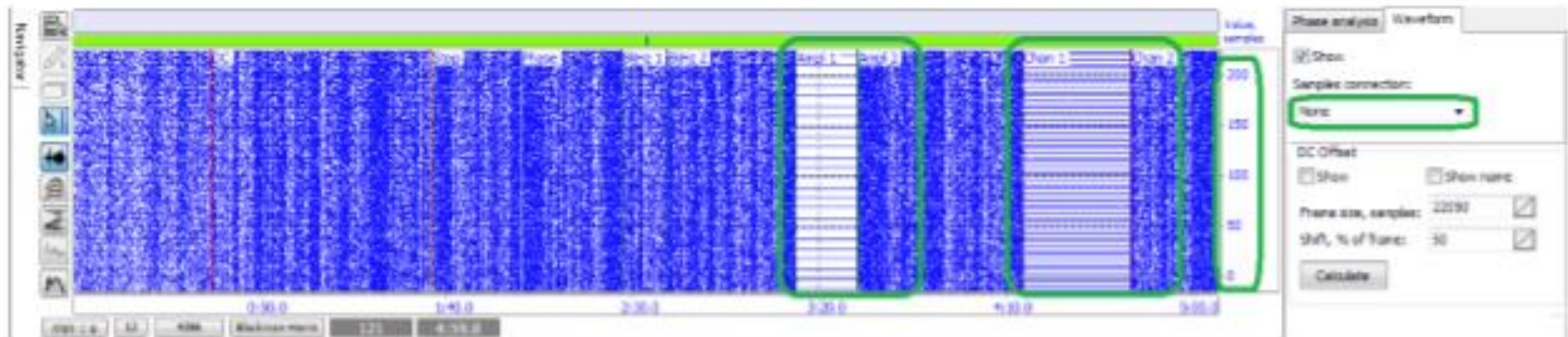
SPECTROGRAMS FOR THE WORD "DUS KAROD"



1. Waveform

Section 1. *Traces of amplitude attenuation or compression law change.*

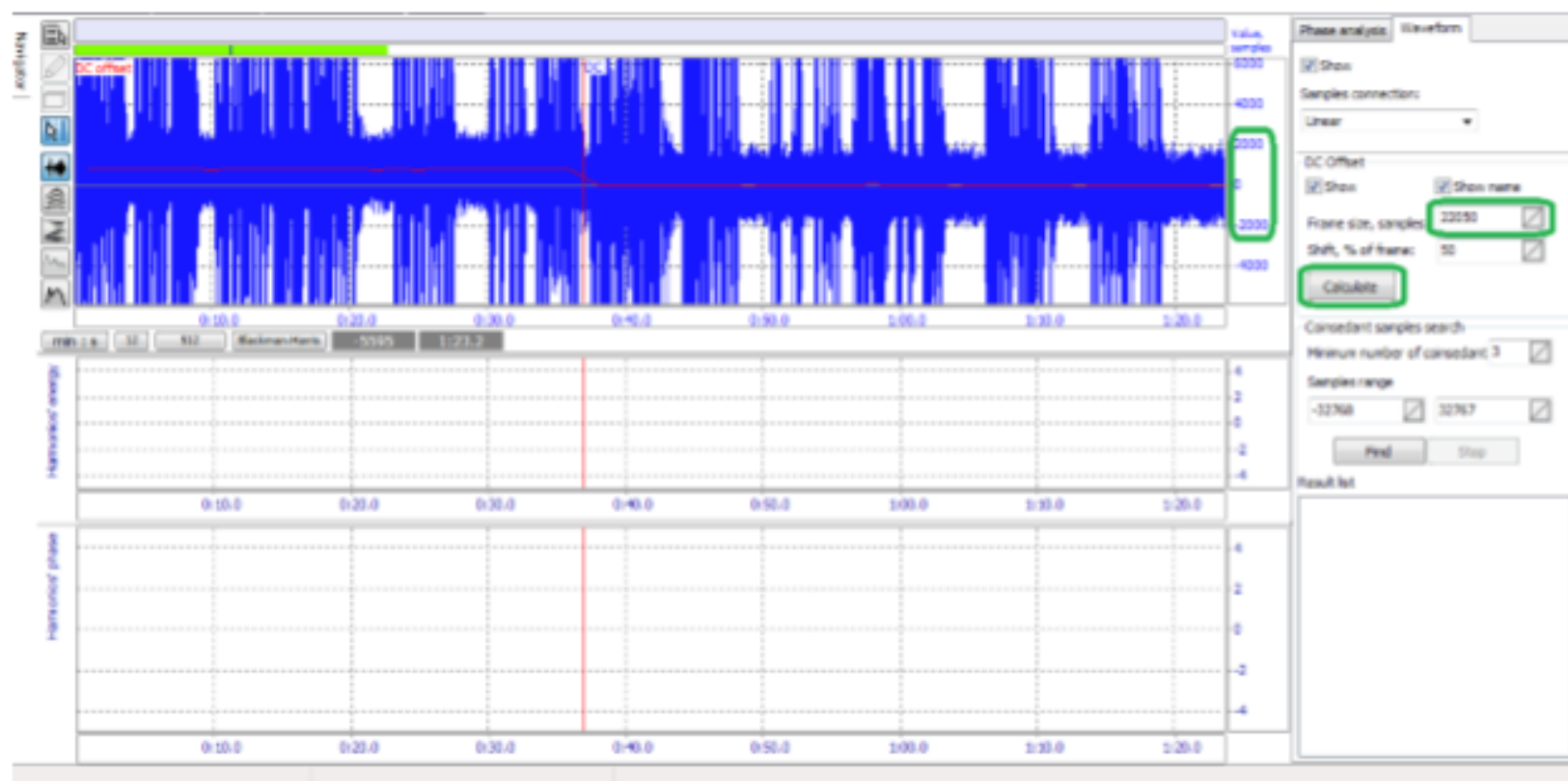
Choose "None" in Samples connection field. Zoom in vertical axis (MW¹ over axis) to range [0, 200].



Amplitude of two fragments was increased during editing.

Section 2. DC Offset break.

DC Offset correspond to center of samples distribution, weighted on a frame. Take 2 seconds for frame length (22050 samples). Zoom in vertical axis to range [-4000, 4000]. Click “Calculate” button.



There is DC Offset break at 00:37, which is a trace of editing. Usually this happens when different parts of audio were put together at this point.

Section 3. Repeating samples search.

Some digital recorders put several repeating zero samples during recording pause.

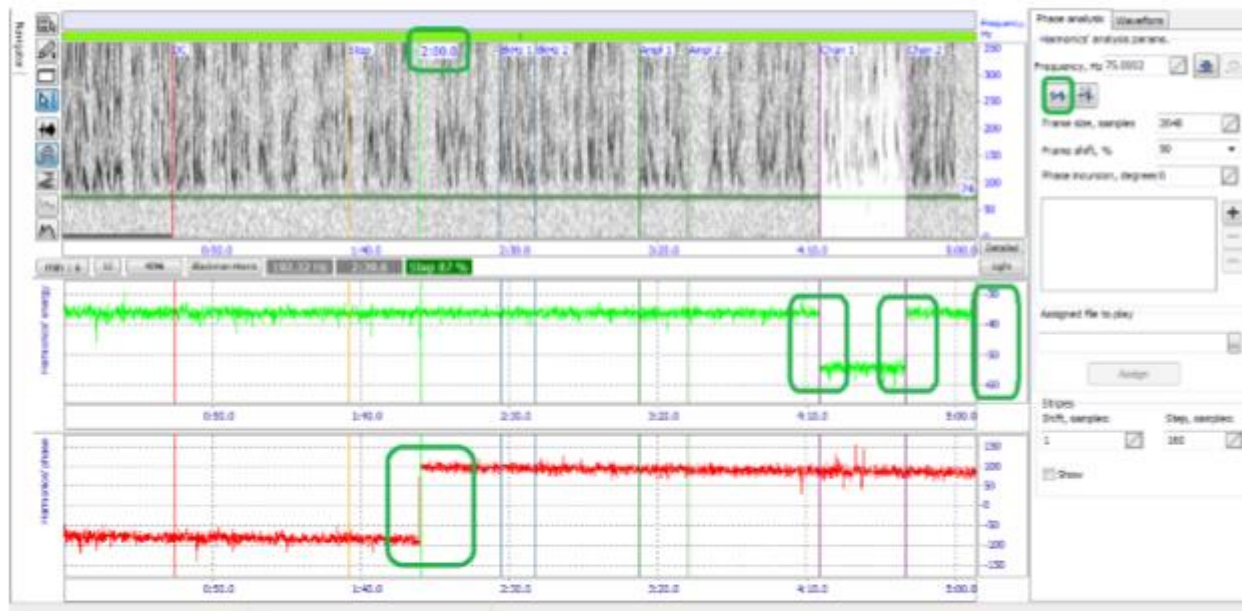
Set Minimum number 10, Samples range [-2000, 2000], click "Find" button.



Choose 1:36 from the Result list. The editing (or pausing) point cannot be heard.

2. Phase analysis

Narrowband power supply harmonics should not have phase continuity breaks.

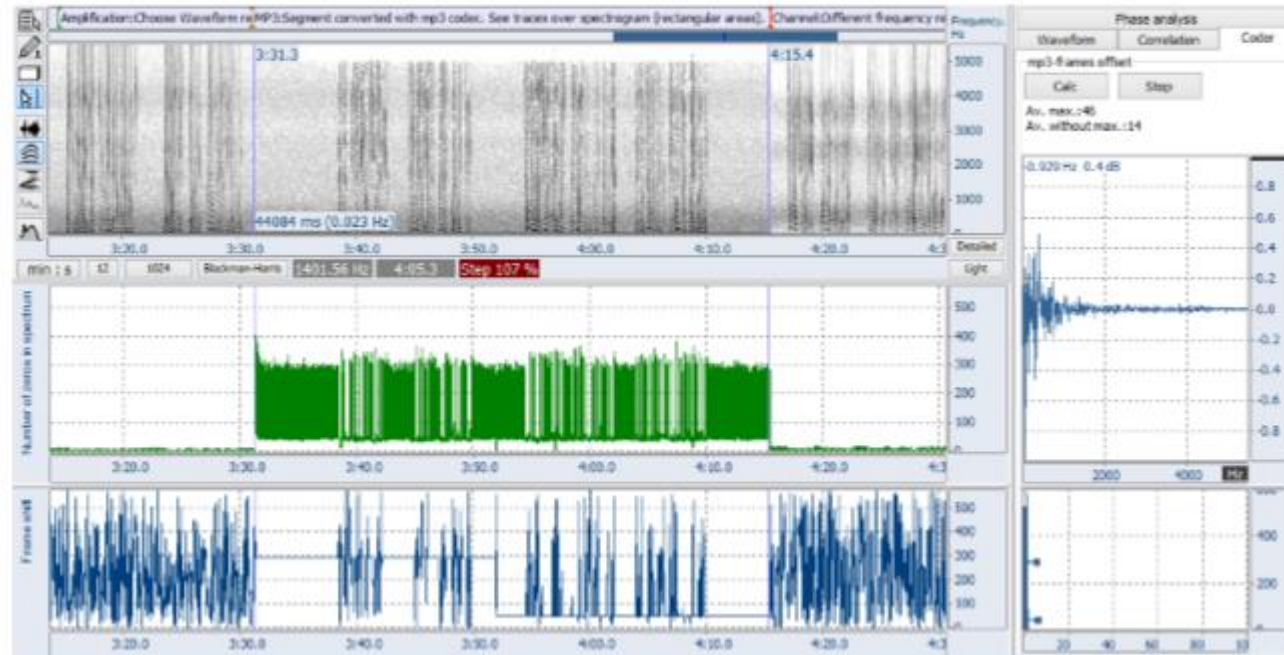


Phase break at 02:00 cannot be heard, but this is the trace of the recording editing. There is a break of harmonics' energy too (see channel amplitude-frequency characteristic changing section).

3. Coder

Extraction and periodicity analysis of MP3 frames borders is the unique method that reveals editing of digital recordings, such as conversion to MP3 file format (even if it is WAV already) and editing of MP3 files.

Click Calc. button to extract the MP3 frames borders and frame shift



Based on the aural & spectrographic examinations the expert integrates his findings into one of the following five conclusions:

- A positive identification
- Probable identification
- Positive elimination
- Probable elimination
- No decision

SOME IMPORTANT CASE STUDIES

CASE STUDY-I

- In an Internet Banking Transaction of ICICI bank, the culprit could manage to get the personal details such as name, date of birth, address of dormant accounts of four NRIs in connivance with a bank officer. For securing Internet Banking passwords of the accounts, the culprit forged the signatures of NRI's (again with the help of bank officer) and got the addresses changed.

- Then the accused telephoned the call centre of ICICI bank and requested for issue of new Internet Banking passwords, posing as the real account holder. As per the protocol of the bank, for identification of account holder the personal details mentioned above were asked and on the receipt of satisfactory replies, the phone banking officer issued fresh passwords which were sent by post to the changed new address of the accused. Then the accused withdrawn Rs. 40.35 Lakhs from these accounts.

- During investigation on the complaint of NRI's the conversation between the accused & phone banking officers (which will be recorded by default) were copied onto CD's and sent to APFSL along with control voices of the suspects for comparison.

- The questioned & specimen voices were analyzed by means of auditory and spectrographic examinations using computerized speech Lab (CSL 4500) and opined that questioned & specimen voices belong to the same individual.

CASE STUDY-II

- In a case of black mail (in Guntur District) the accused who is an advocate, demanded 15 Lakhs from Managing Director of a finance company, otherwise threatened to launch an agitation on the allegation that the finance company was claiming high interest rates. The entire conversation was recorded and referred to APFSL along with its control.

- The auditory and spectrographic examinations made on Computerized Speech Lab revealed that the questioned and control voices belong to the same individual.

CASE STUDY-III

- In a case referred from Ratnagiri District of Maharashtra state, threatening calls demanding huge ransom received by a builder & land developer through cell phone were recorded. During investigation it was found that the cell phone was used by a prisoner (convicted in a bank robbery & lodged in the local jail) posing himself as member of Chotarajan's gang.

The recorded questioned and specimen voices which were sent to APFSL were analyzed using Computerized Speech Lab and it was opined that they belong to the same individual.

- ❑ Assistant Public Prosecutor murder case
- ❑ Railway linemen (Bihar) murder case
- ❑ Hatred speech (Public meeting)
- ❑ Abusive calls

THANK YOU