SSMT Take Home Exam

Ananya Mukherjee (2018801009)

Video Details

- Name : (Programming, Data Structures and Algorithms -mod01lec49)
- Link :- https://www.youtube.com/watch?v=YOtXKE1Kjdk
- Duration : (16 mins)
- Speaker : Female (Prof. Hema Murthy)

Experiment 1 : Automatic Transcription

- **Changed mp3 to .wav format.**
- **Used Pydub AudioSegment, Split_on_Silence**
- **Min silence length = 1.5**
- **Silence Threshold = -36 dBfs**
- **Processed individual chunks by feeding it to Google ASR**
- **Captured the transcribed output in a text file.**
- **Problem : Non Spoken Noise (NSN) which was due to the sound made by chalk while writing on board.**

Results :

https://iiitaphyd-my.sharepoint.com/:w:/g/personal/ananya_mukherjee_research_iiit_ac_in/EevGsP2TCxpDrlS24CX_pmYBYERVJHuW12p7gDeFoV-sUw?e=DxW3KX

Experiment 2 : Automatic Transcription

Made the following changes to the sample wave file.

1. Down-sampled from 44.1 K Hz to 16 K Hz

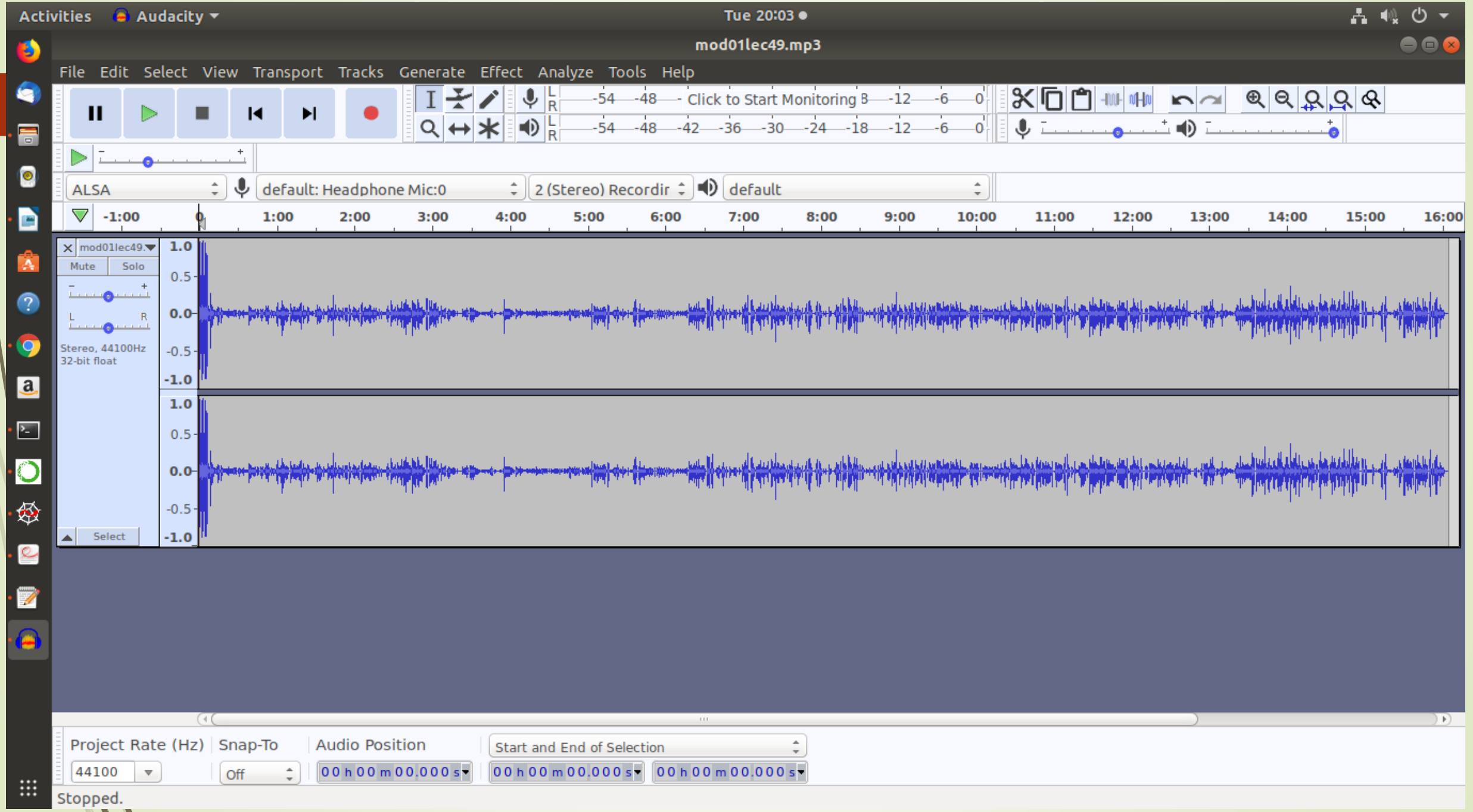
- As per study, Google ASR is trained on 16 K Hz sampling frequency but the downloaded youtube audio wave file was of 44.1 K Hz.

2. Changed from Stereo to Mono

- Youtube audio wave file is stereo by default. But Google ASR is trained on mono.

3. The audio wave file was split based on pause.

- Consider audio below -36 dB as silence. Consider duration of silence as .8 seconds.


Activities
Audacity
Tue 20:03
mod01lec49.mp3
File Edit Select View Transport Tracks Generate Effect Analyze Tools Help
Click to Start Monitoring
ALSA
default: Headphone Mic:0
2 (Stereo) Recordir
default
mod01lec49.
Mute
Solo
Stereo, 44100Hz
32-bit float
Select
Project Rate (Hz)
Snap-To
Audio Position
Start and End of Selection
44100
Off
00 h 00 m 00.000 s
Stopped.

Results of Experiement 2:

Experiment 2A:

(r.adjust_for_ambient_noise(source))

Adjusts the energy threshold dynamically using audio from source (an AudioSource instance) to account for **ambient noise**.

Experiment 2B:

Without adjusting the energy threshold.

Comparision of 2A & 2B

https://iiitaphyd-my.sharepoint.com/:x:/g/personal/ananya_mukherjee_research_iiit_ac_in/ETsMeP1KGipPjzMfo6Cys1oB8PfCeyKoA-b6W6s6ub6V_g?e=KVxeZA

Code:

https://iiitaphyd-my.sharepoint.com/:u:/g/personal/ananya_mukherjee_research_iiit_ac_in/EXvhEXMN0SJAqaH7mGlPt3QBSMv4GrQF1cKlbe1QR_oUQQ?e=BIRZcM

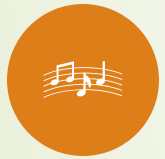
Manual Transcription

- https://iiitaphyd-my.sharepoint.com/:w:/g/personal/ananya_mukherjee_research_iiit_ac_in/EbzFLUh40lJFhT7KrTPuyhYBCRemiENTz0w_bcTbSihHhw?e=3AMLC8
- Manually transcribed the previous video and marked the dysfluencies and prosodic features in it.

Dysfluencies

- **Speech Dysfluencies**
  - Clearing Throat
  - Touch wood
  - Background noises
- **Text Dysfluencies**
- Filler Words (also include pause fillers)
- Filler Sounds/hesitations (uh, aa, aahm)
- Pet Phrases
- Repair

Prosodic Feature



Intonation

Energy

Duration

Topic of the video	Speaker	Link to transcripts	Dysfluencies	Energy	Intonation
Fundamentals of Nuclear Power Generation-mod10-lec26	Dr. Dipankar N Basu (Male)	link	Frequent repairs and 'ah' **(hesitations)**	Low - Normal	Very minor changes in the pitch. There are gulping sounds throughout the video. Sometimes he rushes through his speech.
CHARLES DICKENS 'THE SIGNAL - MAN' - Close reading and the anxieties of the age	Female Speaker	link	Spoke very fluently with frequent **'ah'** and '**again**' as petword	Always High	Always maintained a high pitch. Pitch increased even more while stressing on several words.
Lecture - 1 Introduction to programming languages	Dr. S. Arun Kumar (Male)	link	Too many filler words 'ah' , pet words like 'I mean', okay, right	Varies	Raises his pitch at the start and maintains it and then drops it at the last word. Also sometimes ends by stressing the last word. (5:30)
Chemical and Biological Thermodynamics	Prof. Nand Kishore	link	Fluent video with'ah' & 'so' and very repairs.	High	Constant with minor variations (high at beginning, low at the end. Stressing on few words like "word", "weight" etc topic words)

THANK YOU

Questions?