

Speech Science

WiSe 2024

Exercise 8: Prosodic Analyses

Jan 15, 2024

Bernd Möbius & Omnia Ibrahim

Language Science and Technology
Saarland University

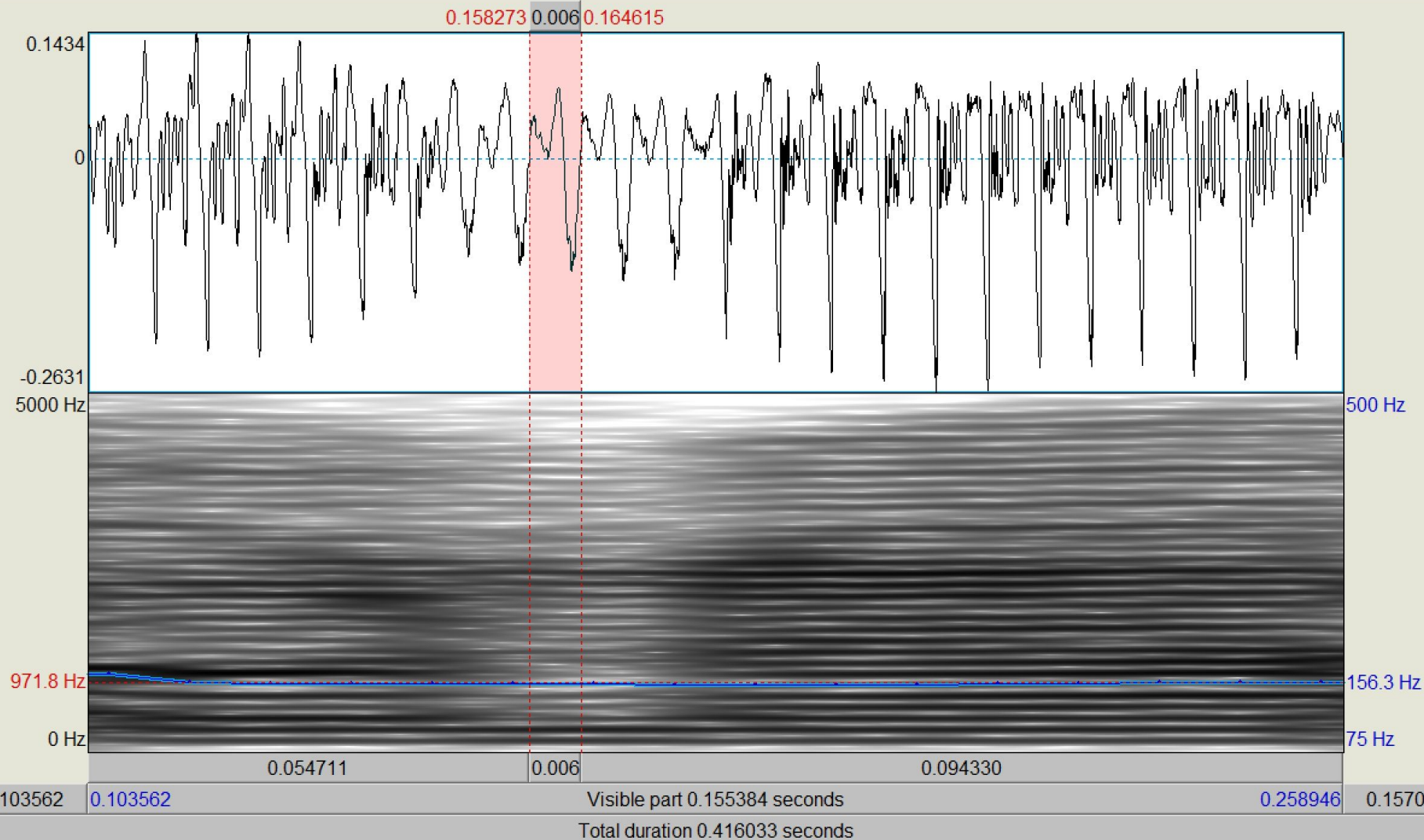


Demonstration – Manual Calculation of F0

16. Sound from ManipulationEditor

File Edit Query View Select Spectrum Pitch Intensity Formant Pulses

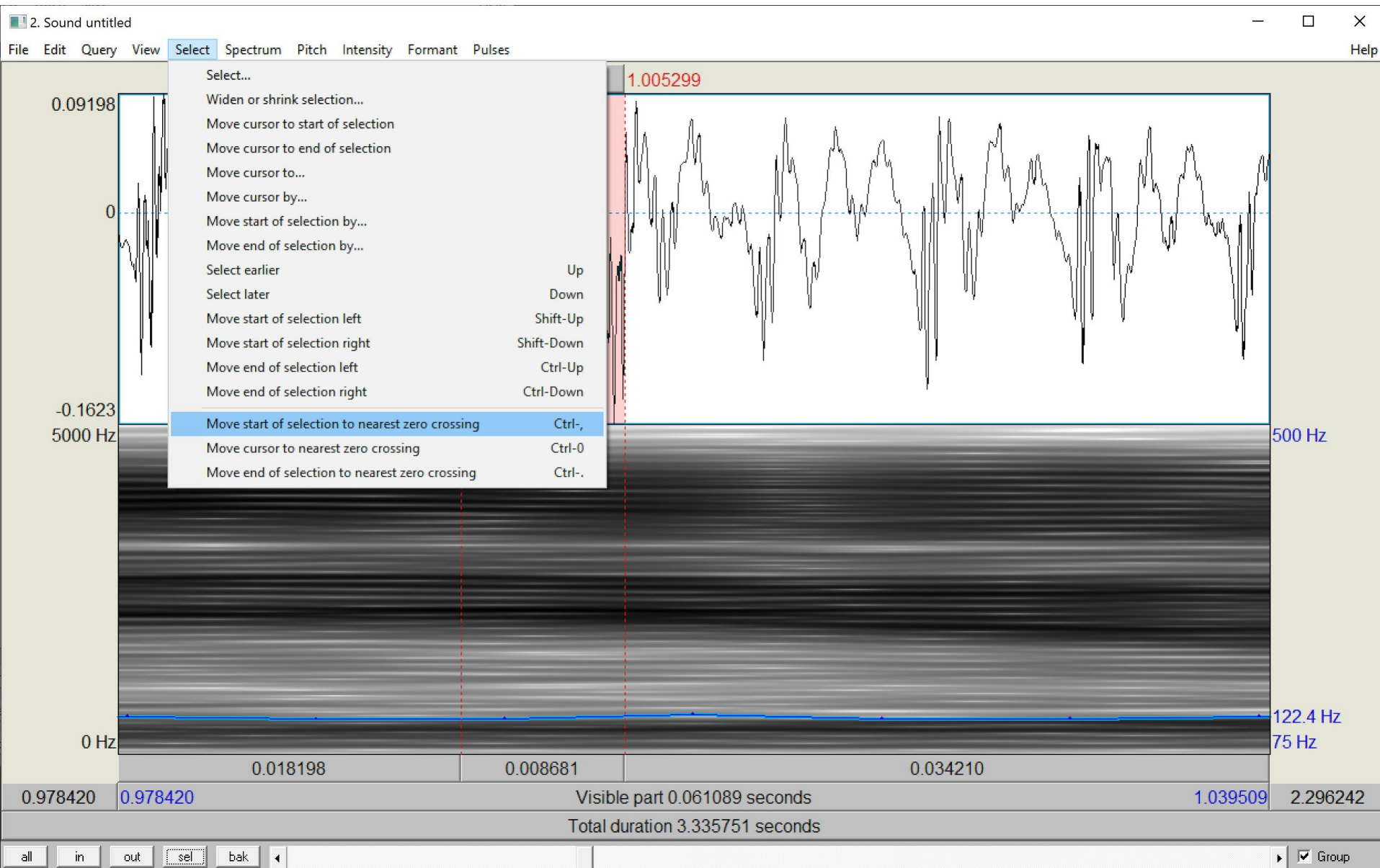
Help



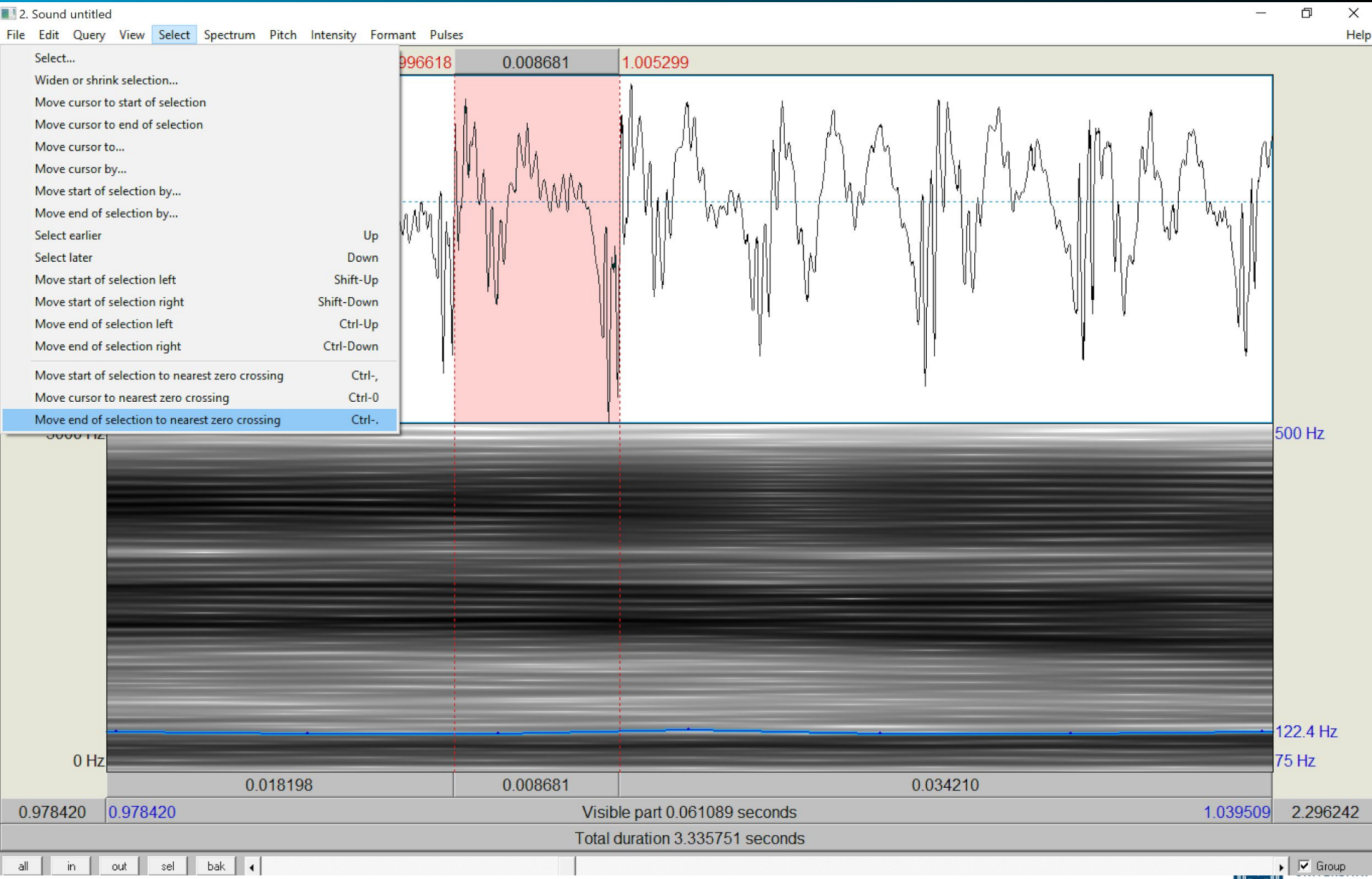
all in out sel bak

Group

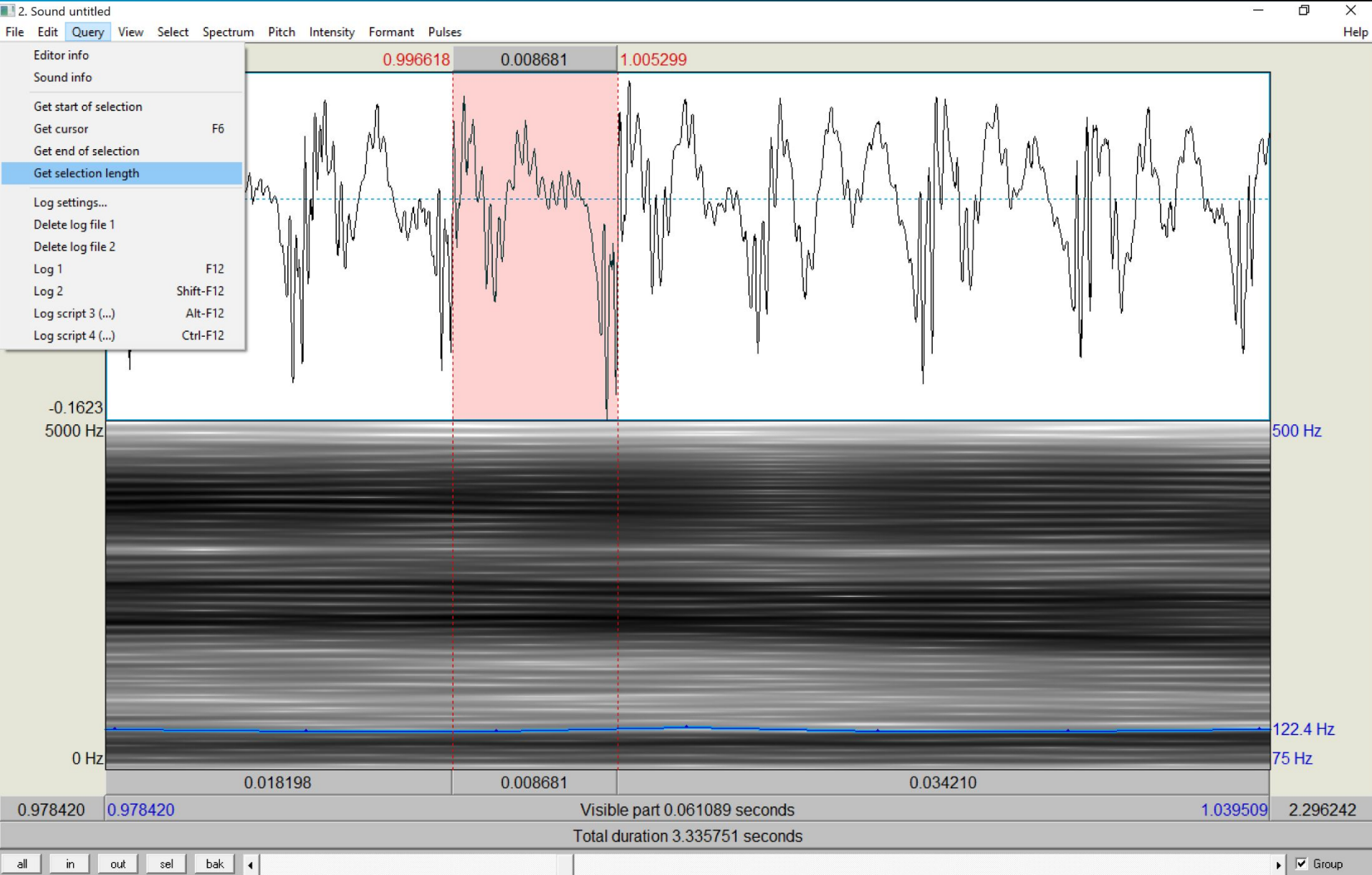
Demonstration – Manual Calculation of F0



Demonstration – Manipulation of duration



Demonstration – Manual Calculation of F0

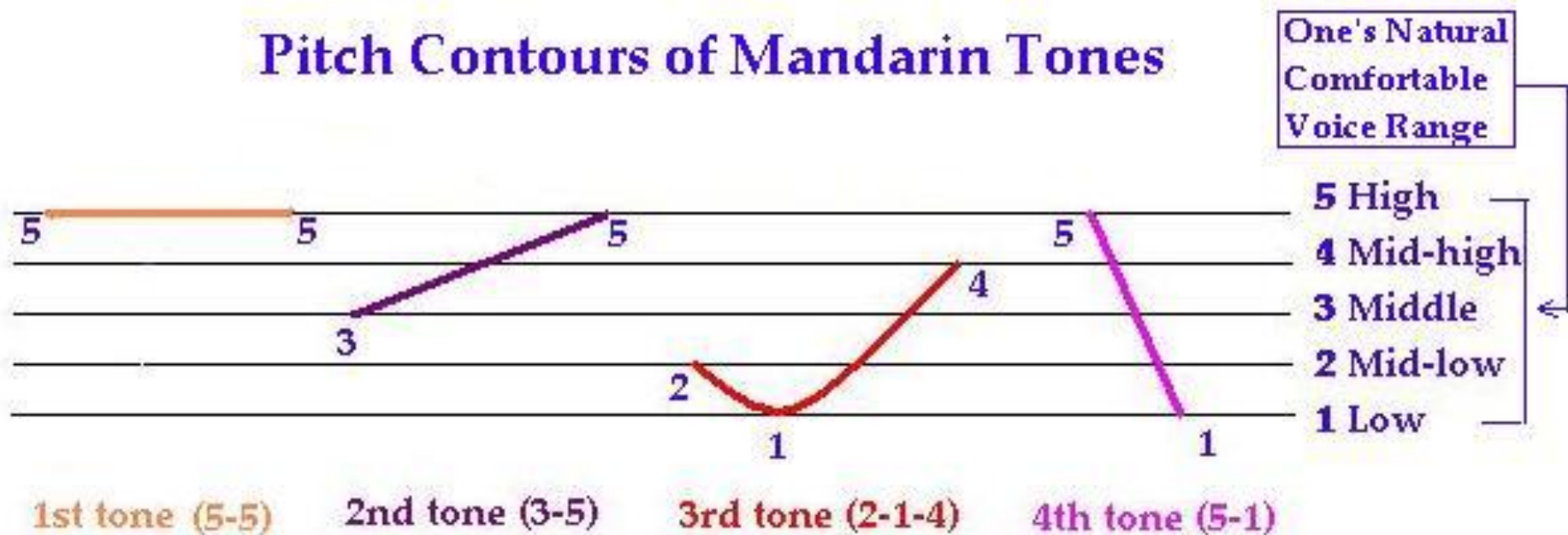


$$\text{Period} = 1 / \text{Frequency}$$

$$\text{Frequency} = 1 / \text{Period}$$

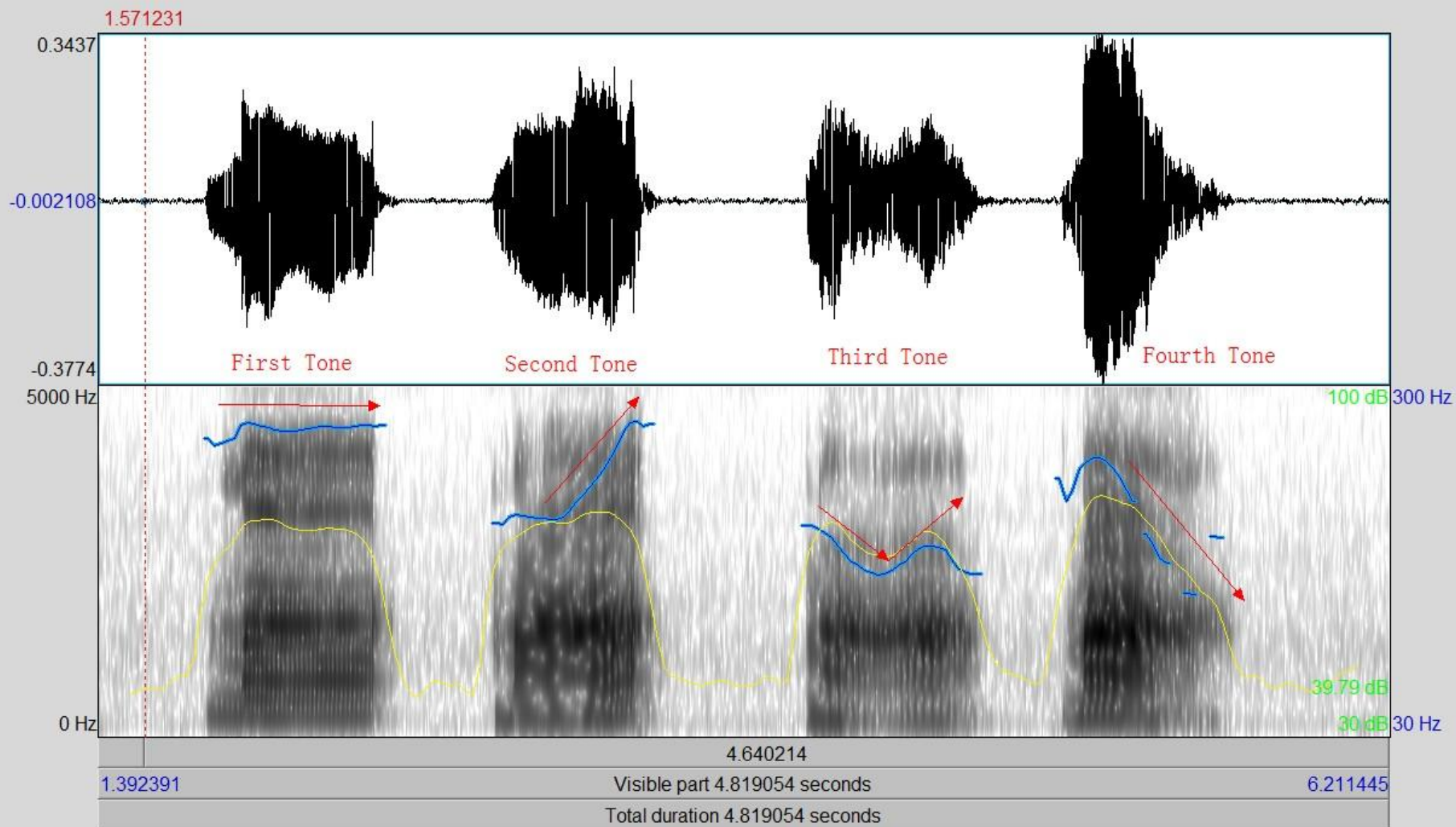
Mandarin Tones

Pitch Contours of Mandarin Tones



From: <http://web.mit.edu/jinzhang/www/pinyin/tones/>

Mandarin Tones Praat Example



From:

https://corpus.eduhk.hk/mandarin_pronunciation/?page_id=664

Research question

Test the influence of fundamental frequency (f_0) on human and machine speaker recognition performance in vocalic test utterances.

Expectations:

- With **increasing f_0** in steady state vowels we expect that speaker recognition performance will decrease since speaker specific vocal tract information decreases (spectral undersampling).
- With **sweeping f_0** in vowels we expect higher speaker recognition performance compared to steady state vowels at any f_0 because more information about the vocal tract is revealed.
- **Humans should rely more on f_0 compared to ASR** based on MFCC representations as these representation are more influenced by the vocal tract transfer function.

General method

Training material:

80 sentences (9 to 37 syllables):

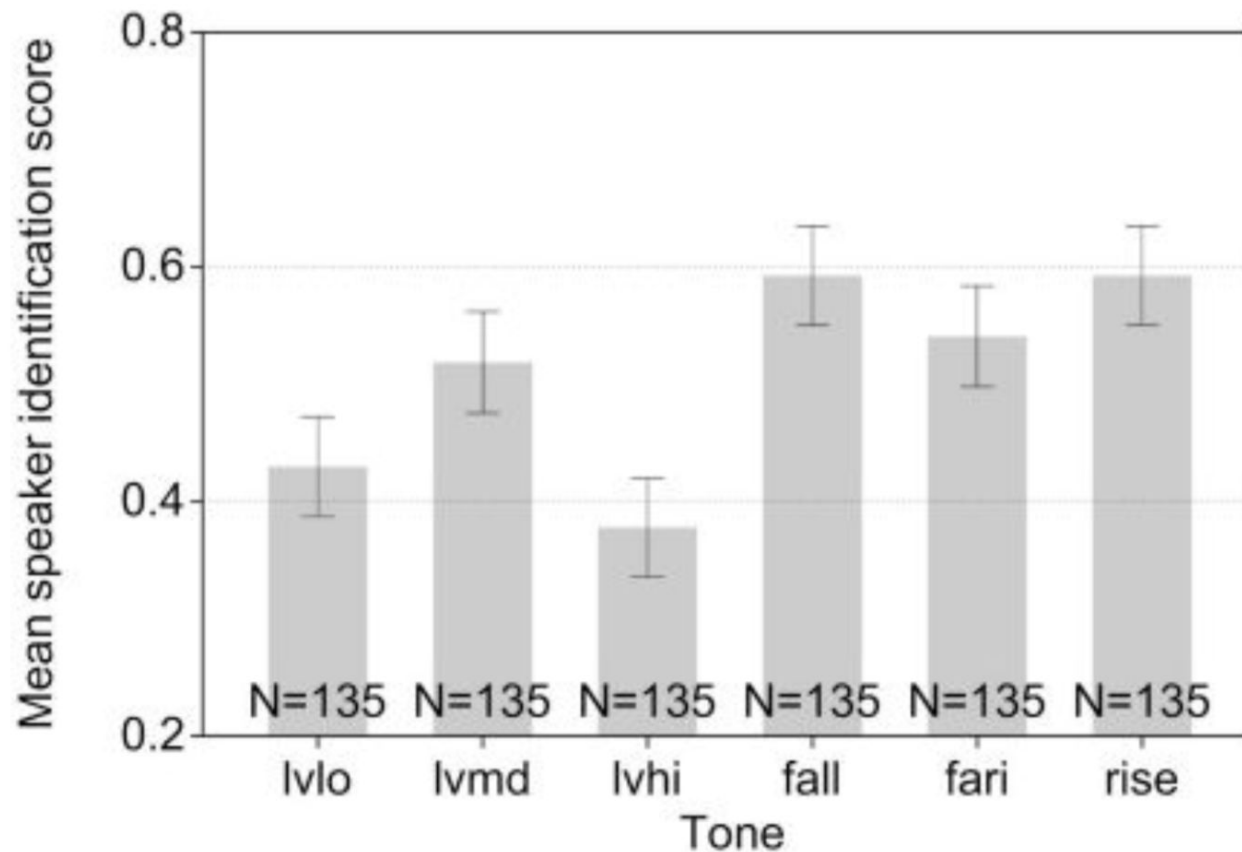
N=1200 (15 speakers * 80 sentences)

Test material:

- **level low (lvlo):** low steady-state pitch
- **level mid (lvmd):** mid steady-state pitch
- **level high (lvhi):** high steady-state pitch
- **contour falling (fall):** falling pitch contour
- **contour fall-rise (fari):** fall followed by rise
- **contour rising (rise):** rising pitch contour

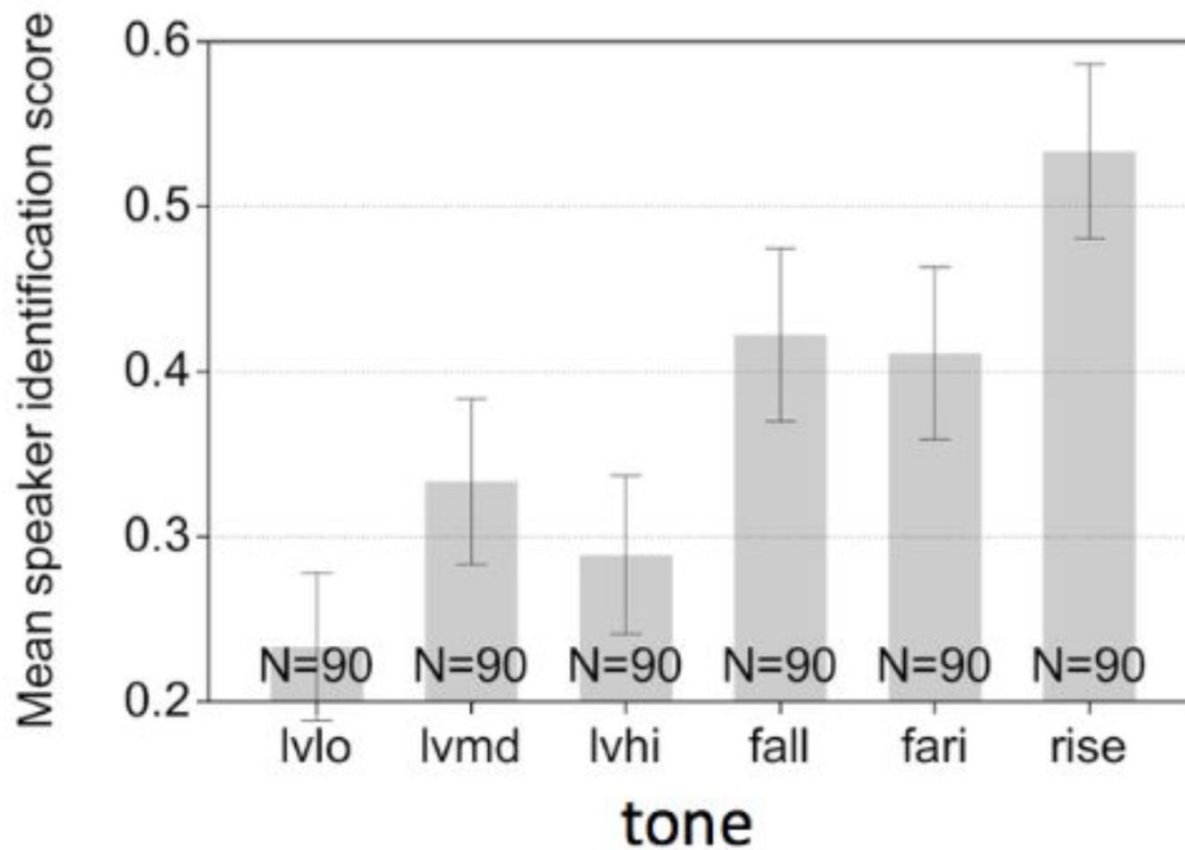
N=810: 15 speakers * 3 vowels (a, i, u) * 6 tones * 3 repetitions

Results: Computer recognition



- SID was worse under level tones compared to contour tones.
- SID was best for mid-level tones compared to low and high.

Results: Human recognition



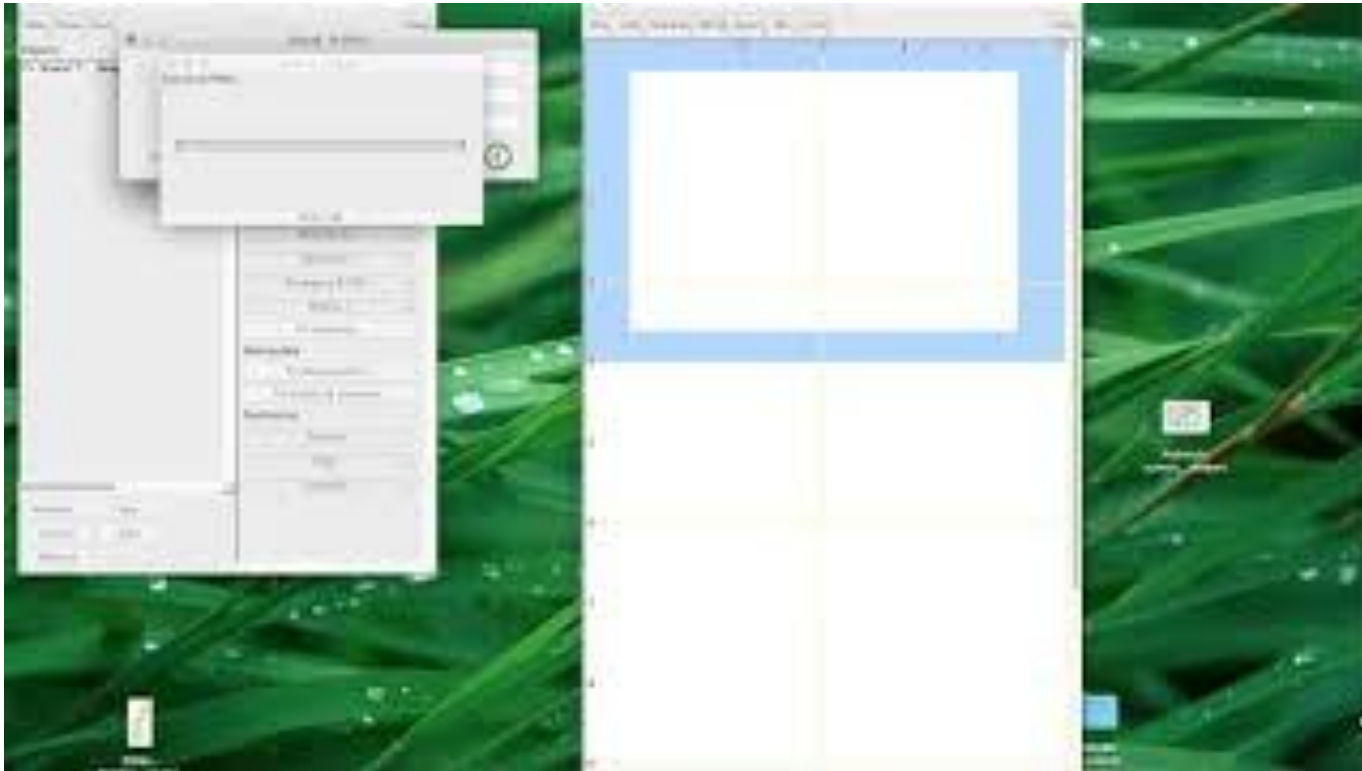
Can humans control their voice identity?

- Surely: known for voice-disguise, humans can hide their identity (more or less well).
 - Can humans expose their identity?
 - If yes, why would they do that?
 - How is that related to communicative goals? I.e., transfer of linguistic information?
- **Look at the development of voice recognition....**



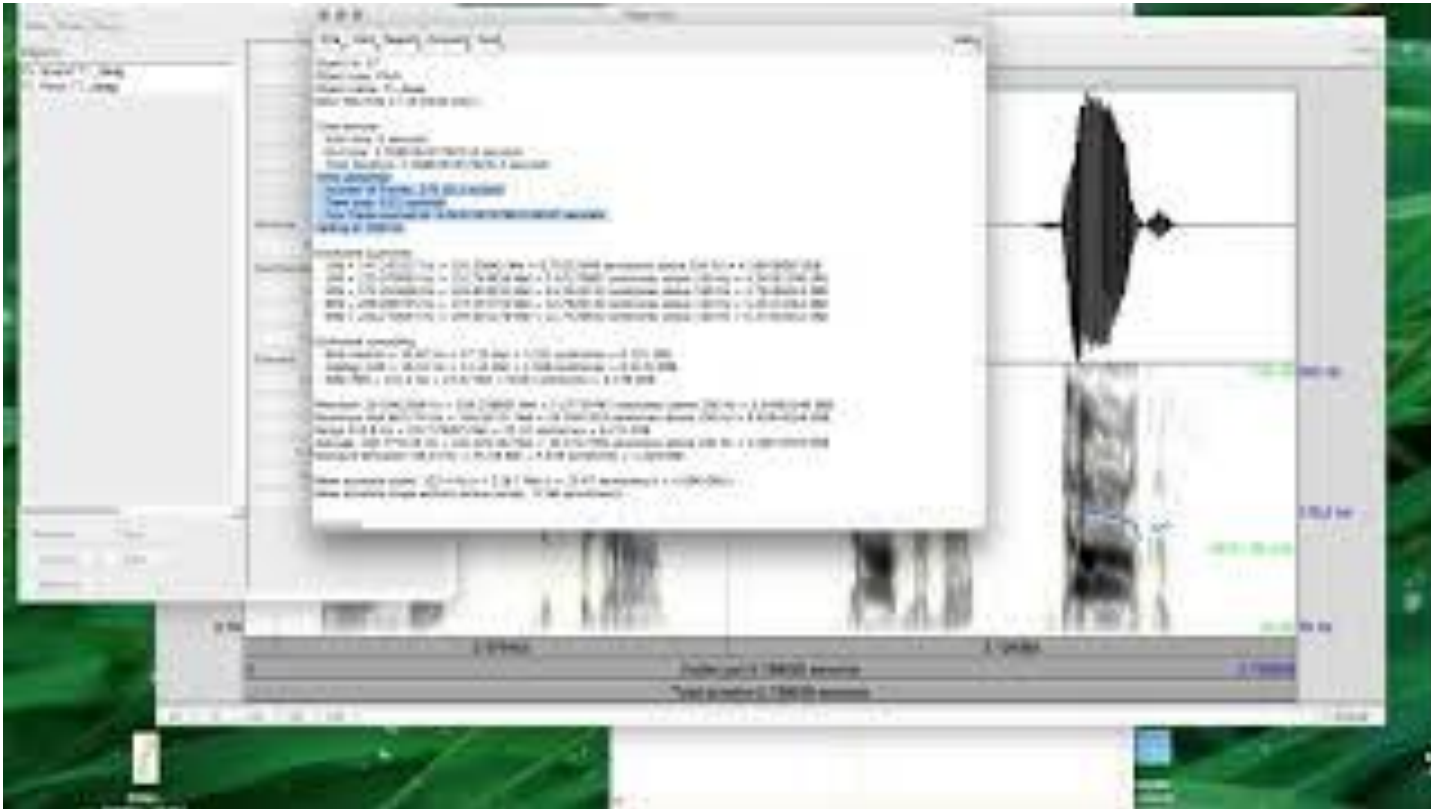
- **Clear Speech**
- **Deceptive Speech**
- **Infant Directed Speech**

Resources: Praat Extract Pitch



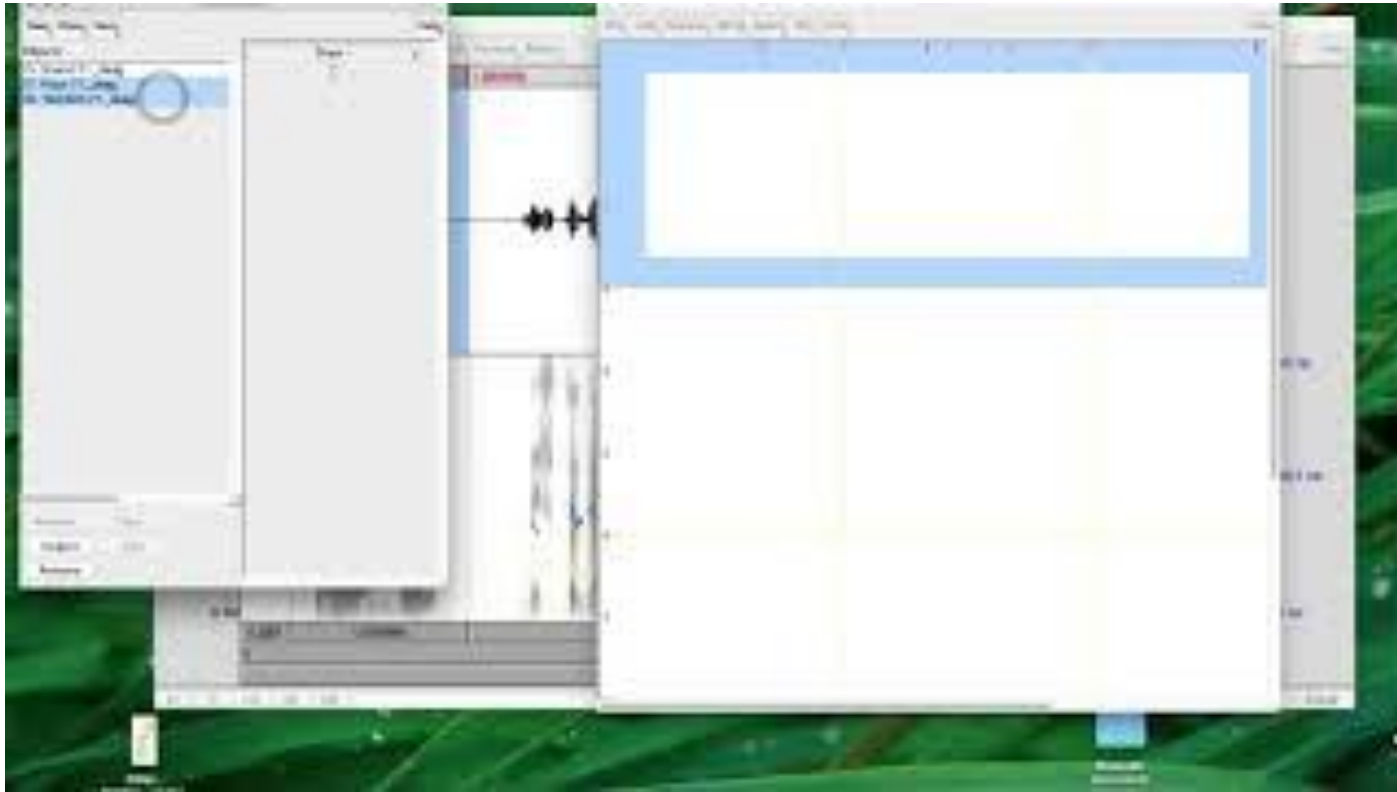
<https://www.youtube.com/watch?v=p8vgEu9p7v8>

Resources: Praat Pitch Info



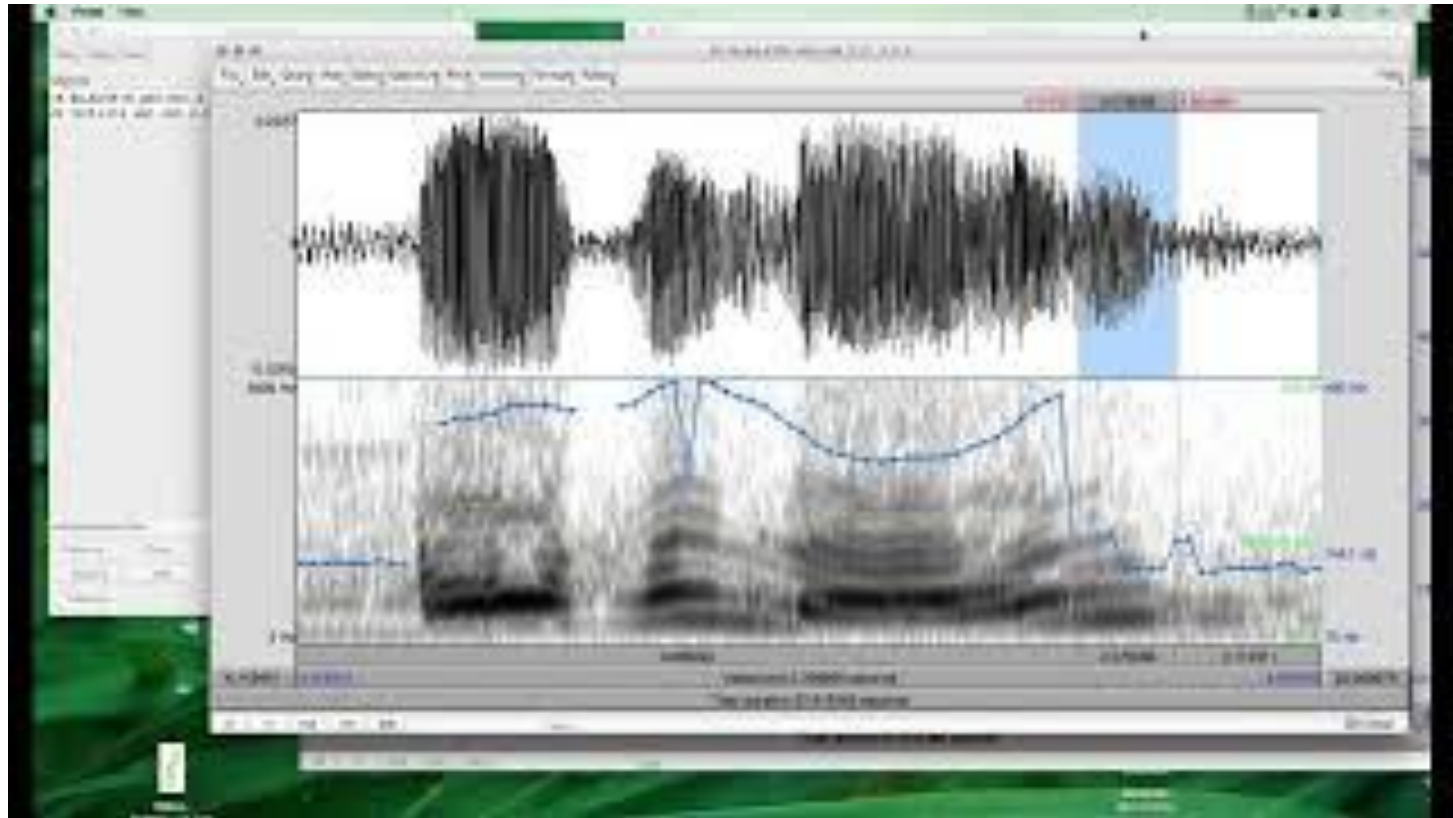
[https://www.youtube.com/watch?v= JOclvMXzCQ](https://www.youtube.com/watch?v=JOclvMXzCQ)

Resources: Praat Pitch Picture (Draw)



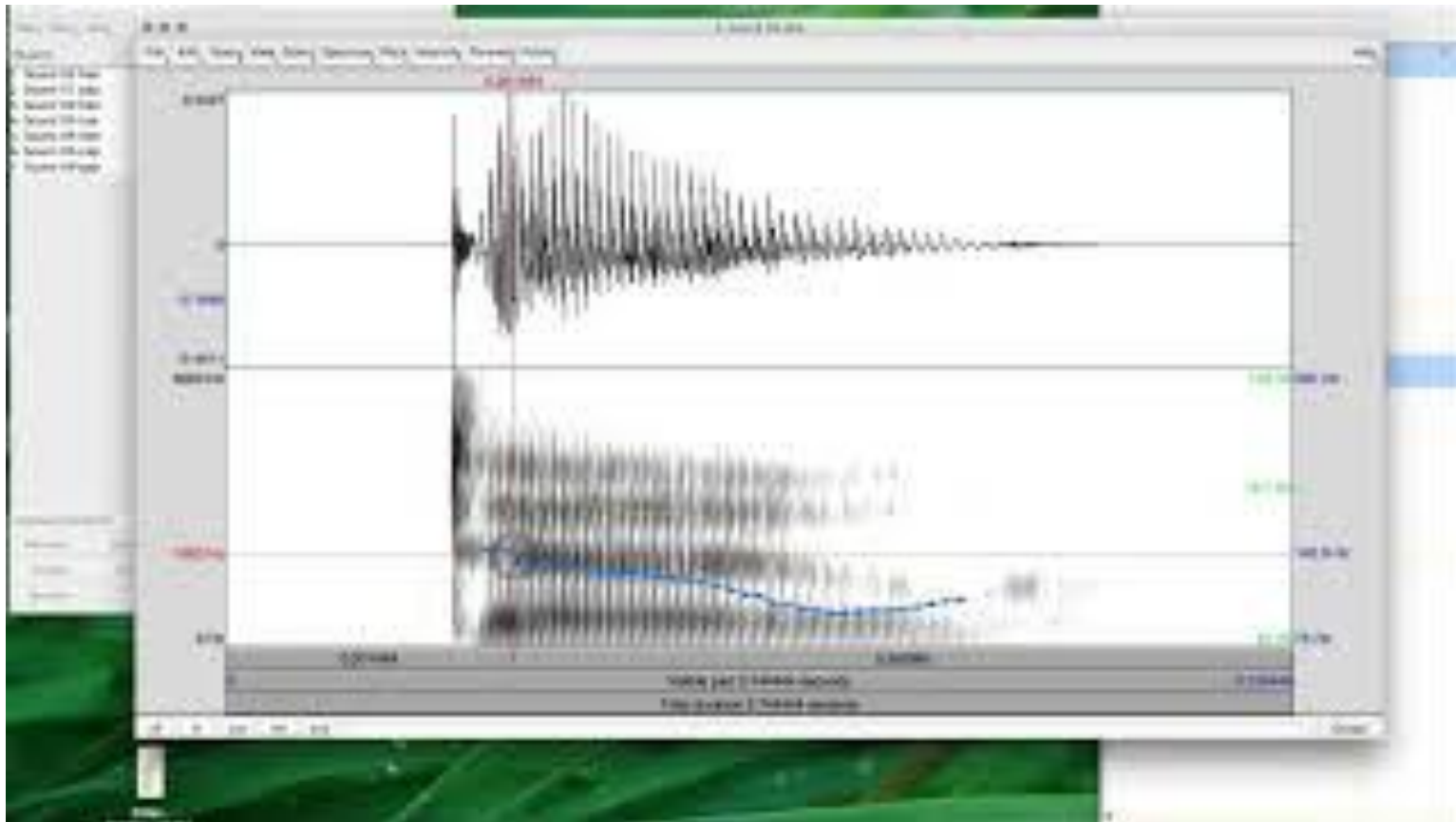
<https://www.youtube.com/watch?v=IElhdt18oEs>

Resources: Praat Cleaning Up Pitch



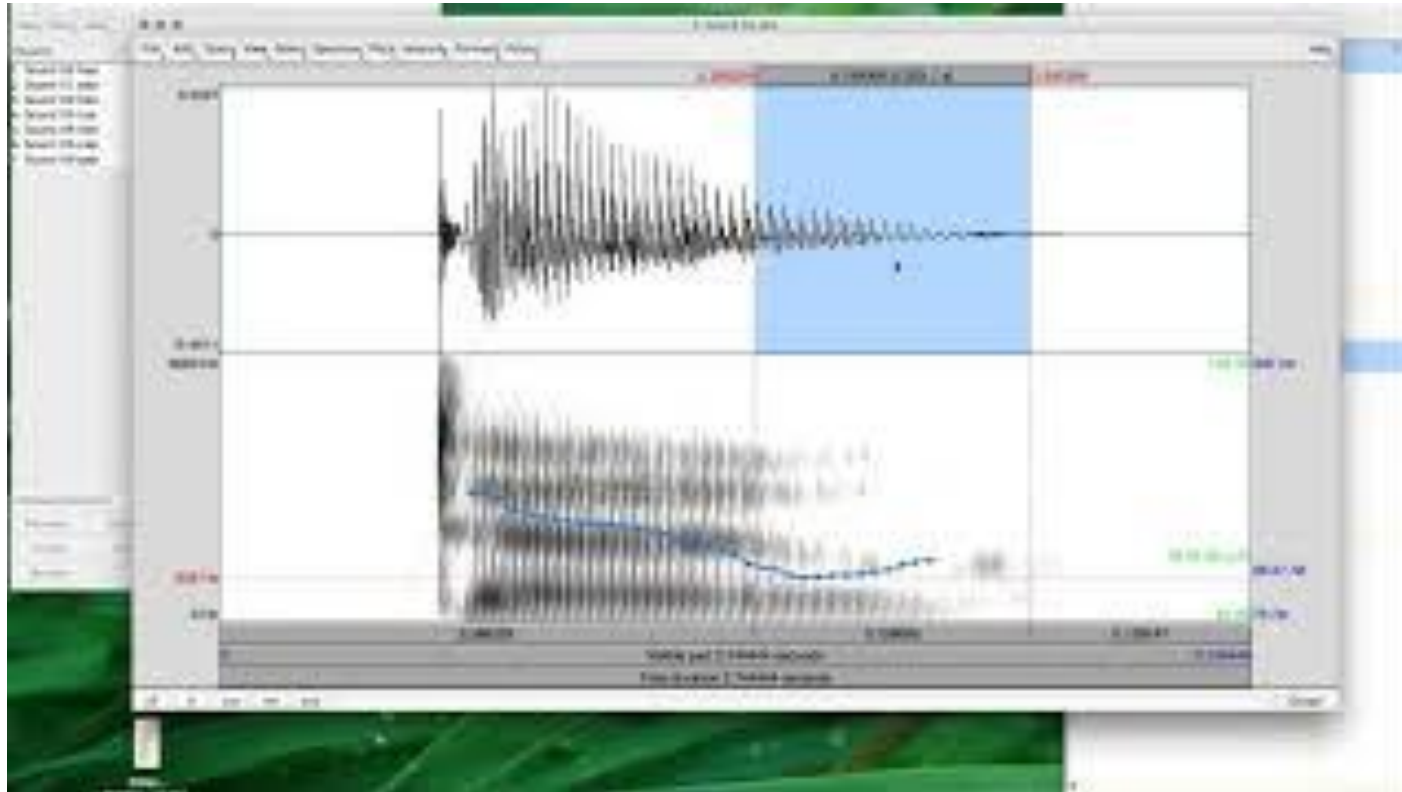
<https://www.youtube.com/watch?v=JGdWFW-rFlg>

Resources: Praat Microprosody



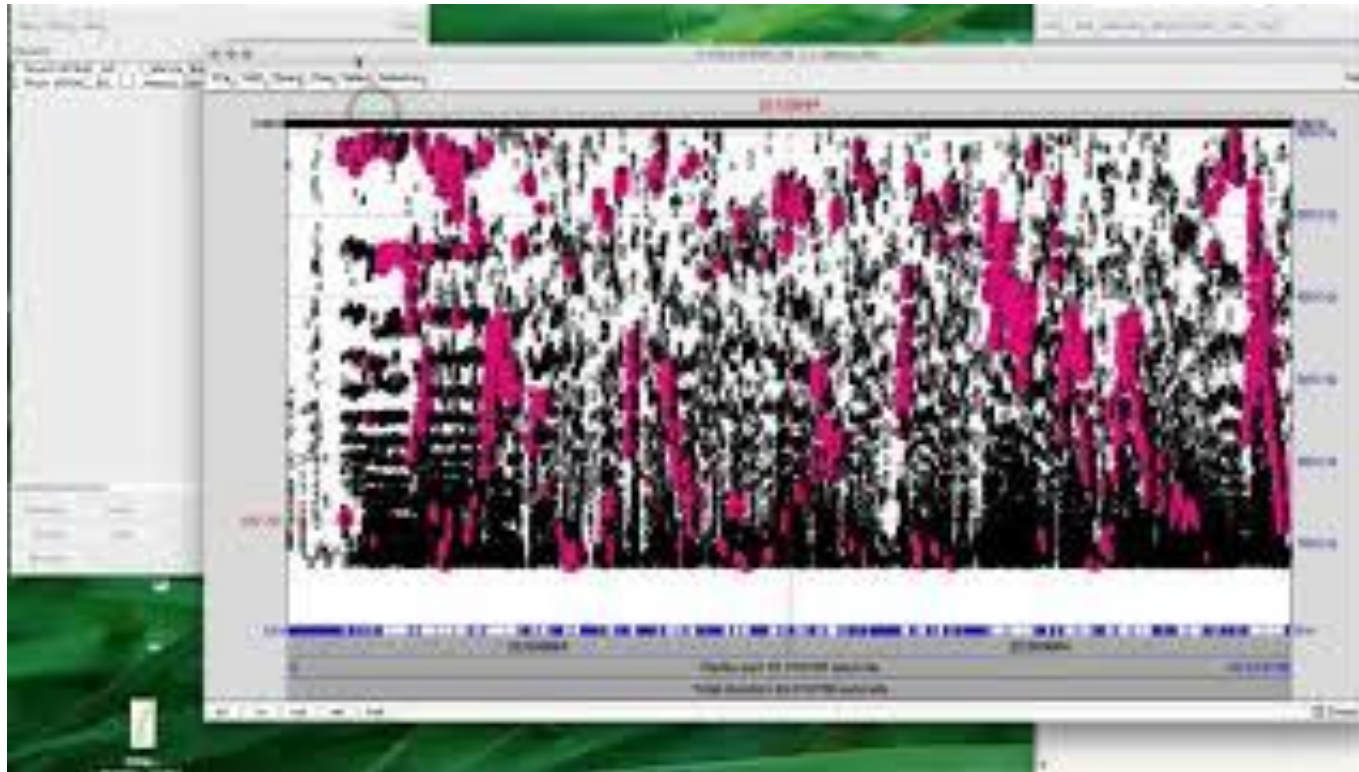
<https://www.youtube.com/watch?v=kteTGERw5Jg>

Resources: Praat Pitch and Amplitude



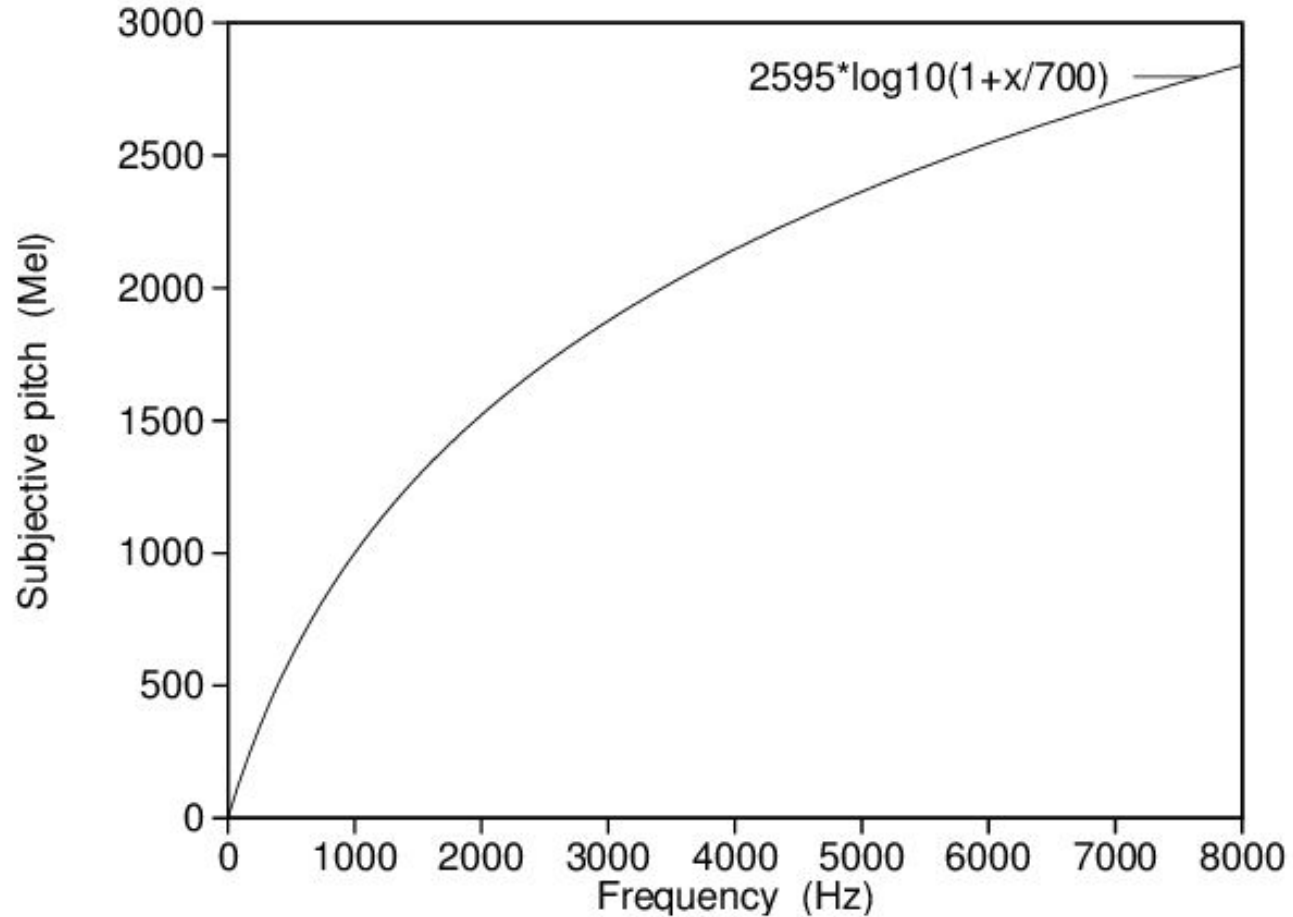
<https://www.youtube.com/watch?v=IbUGq-MLj8>

Resources: Praat Pitch Real-World Example



<https://www.youtube.com/watch?v=xBTV5zRZX4g>

Mel Scale



From:

https://www.researchgate.net/figure/The-relation-between-the-mel-frequency-scale-and-the-linear-frequency-scale_fig4_2685753