Lecture 2 & 3: Speech Analysis and Linguistics
Speech Basics

Instructor: Dr Chiranjeevi Yarra

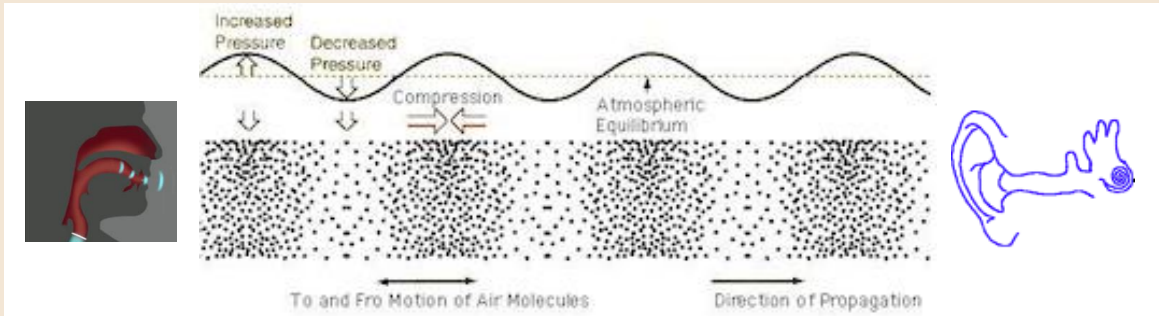
Speech lab, LTRC

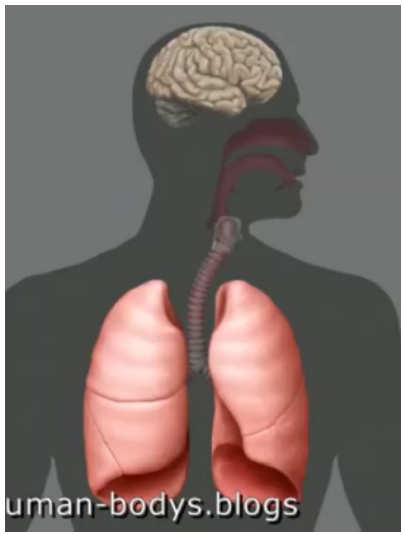
Aug 03& 07, 2023

Outline

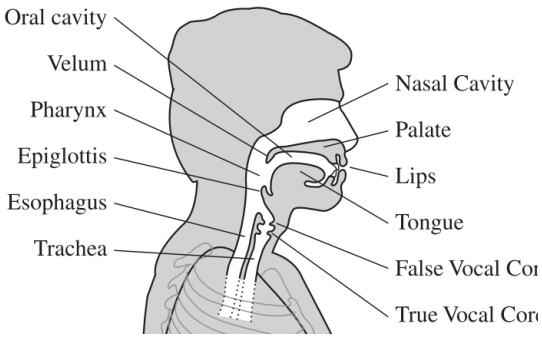
Speech Production Mechanism

Speech is a natural mode of communication

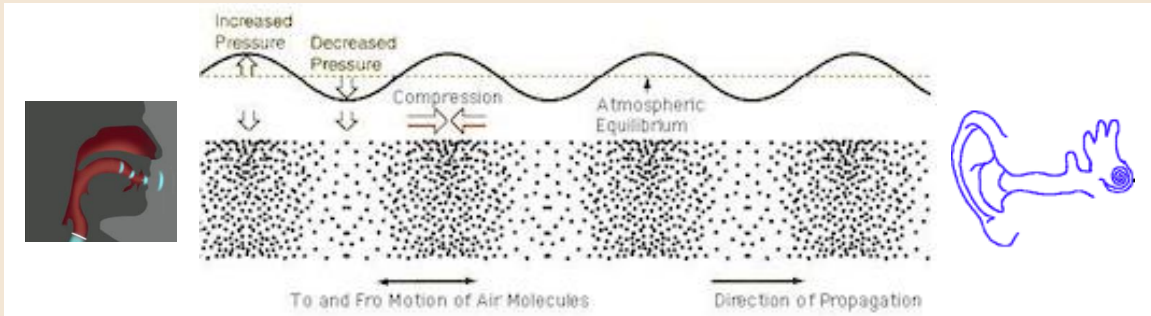


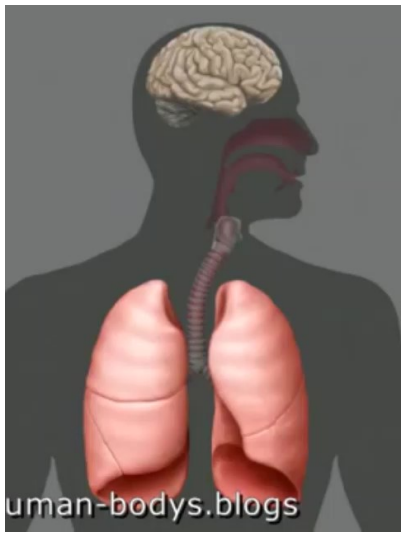


Speech Production Mechanism



Speech is a natural mode of communication

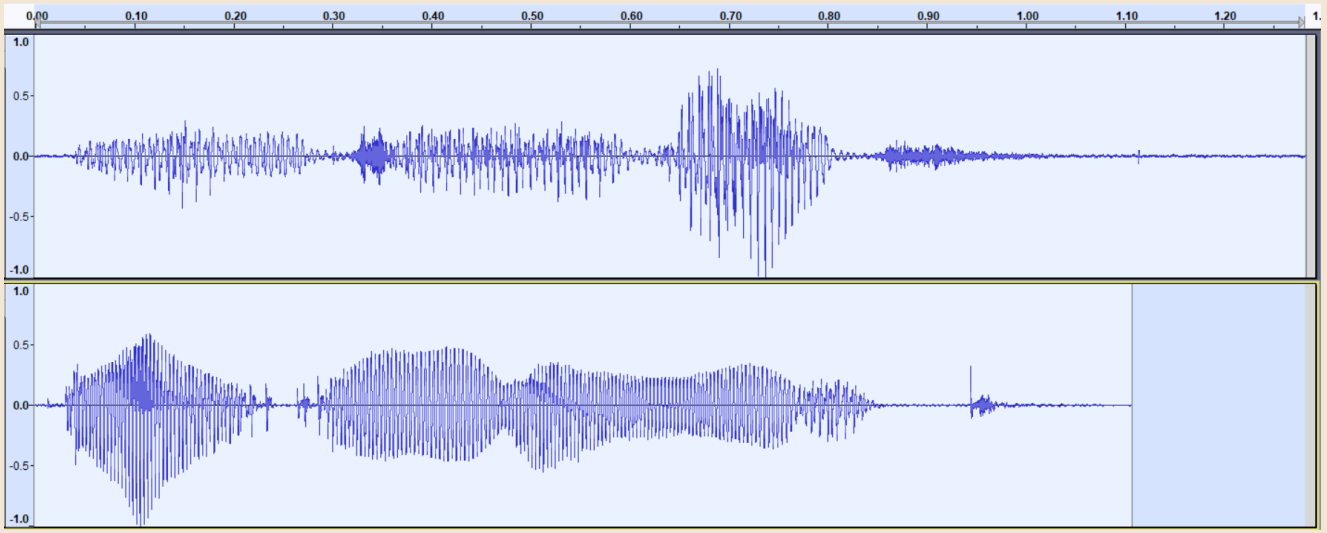




Non-stationary nature

Illustrative example

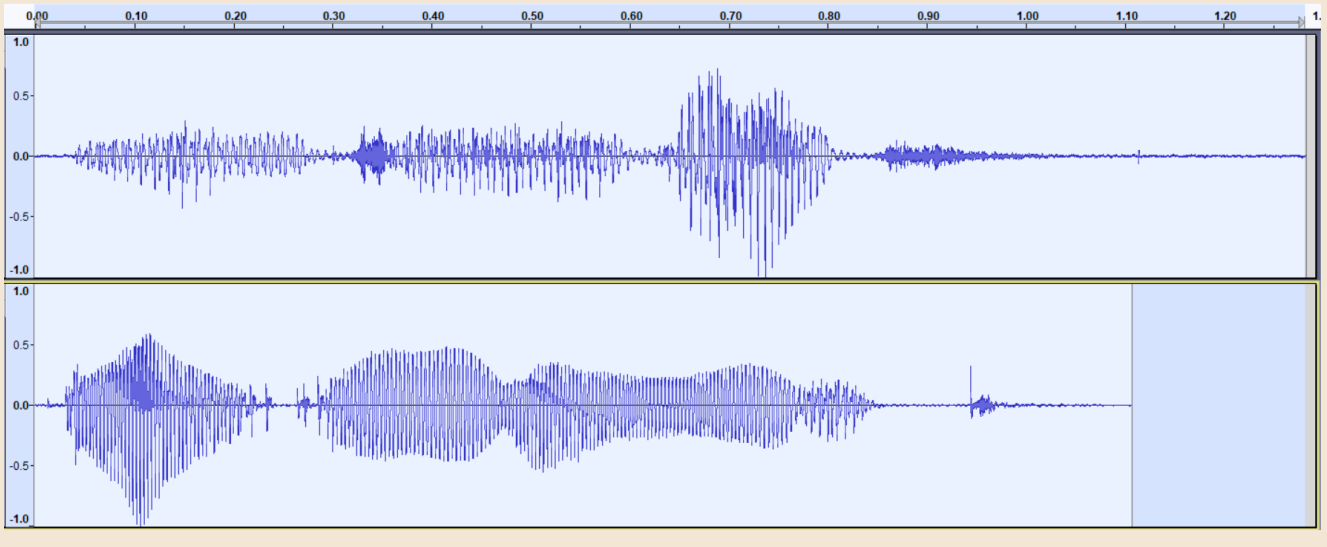
Sentence: "You enjoyed it"; Signal #1: , Signal #2:

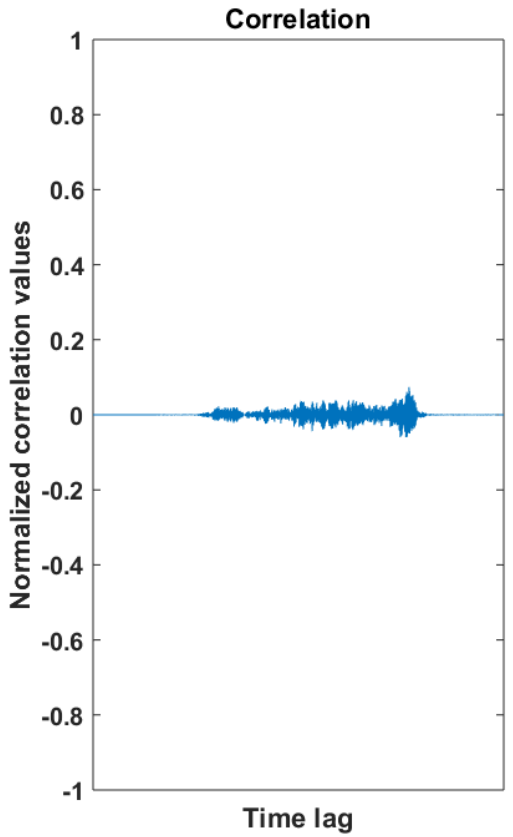


Non-stationary nature

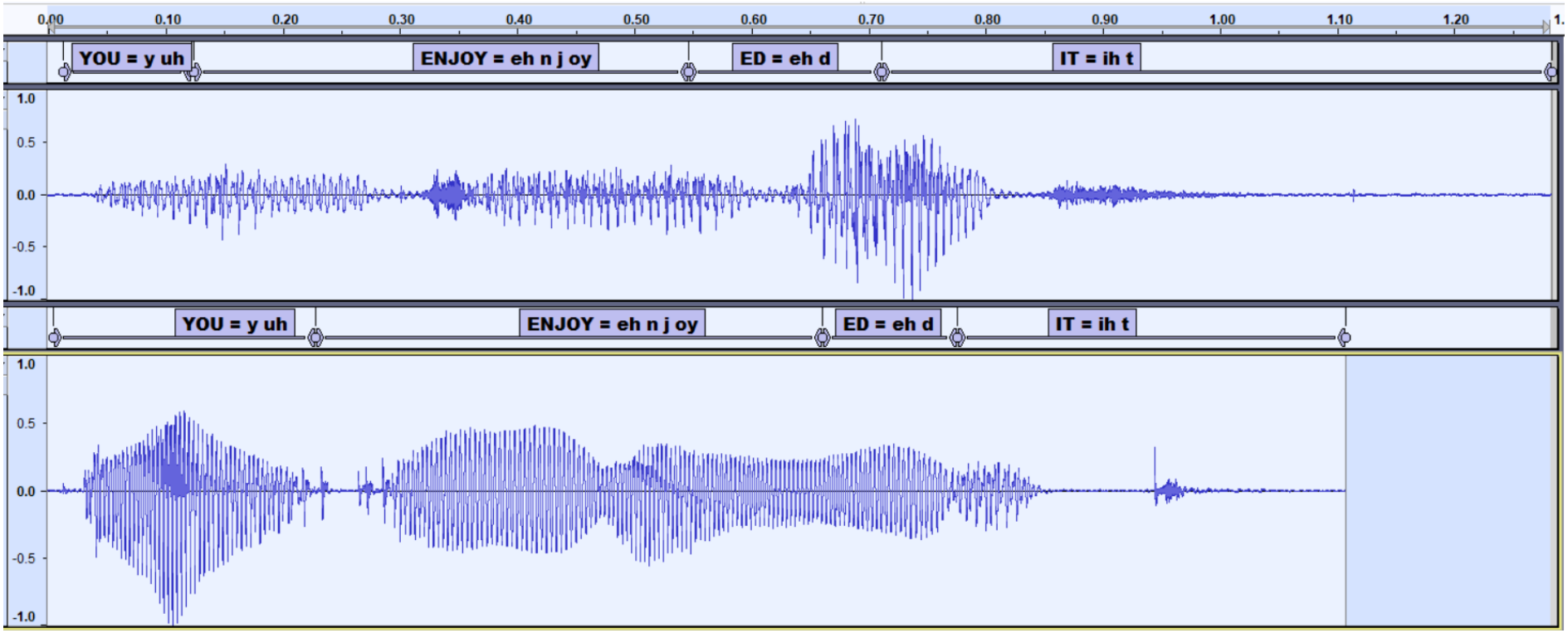
Illustrative example

Sentence: "You enjoyed it"; Signal #1: , Signal #2:

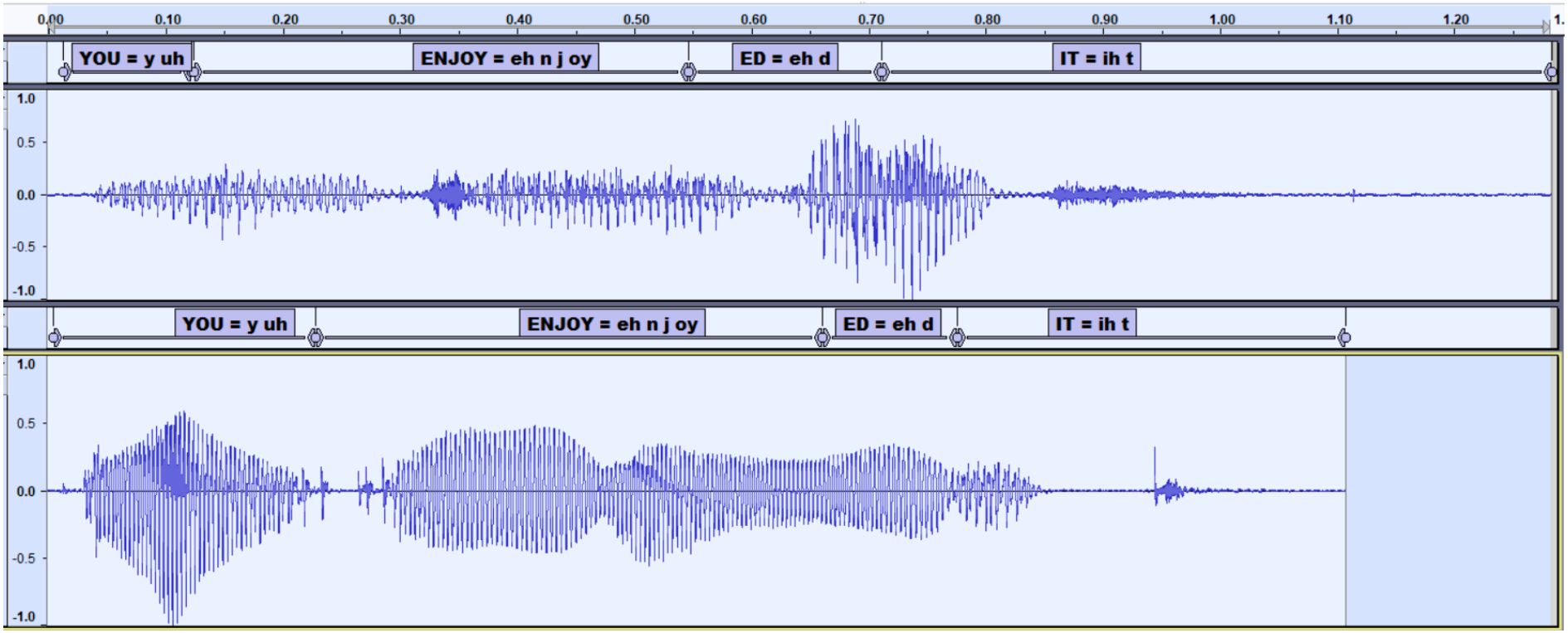




Time-varying Nature



Unknown Segment Boundaries



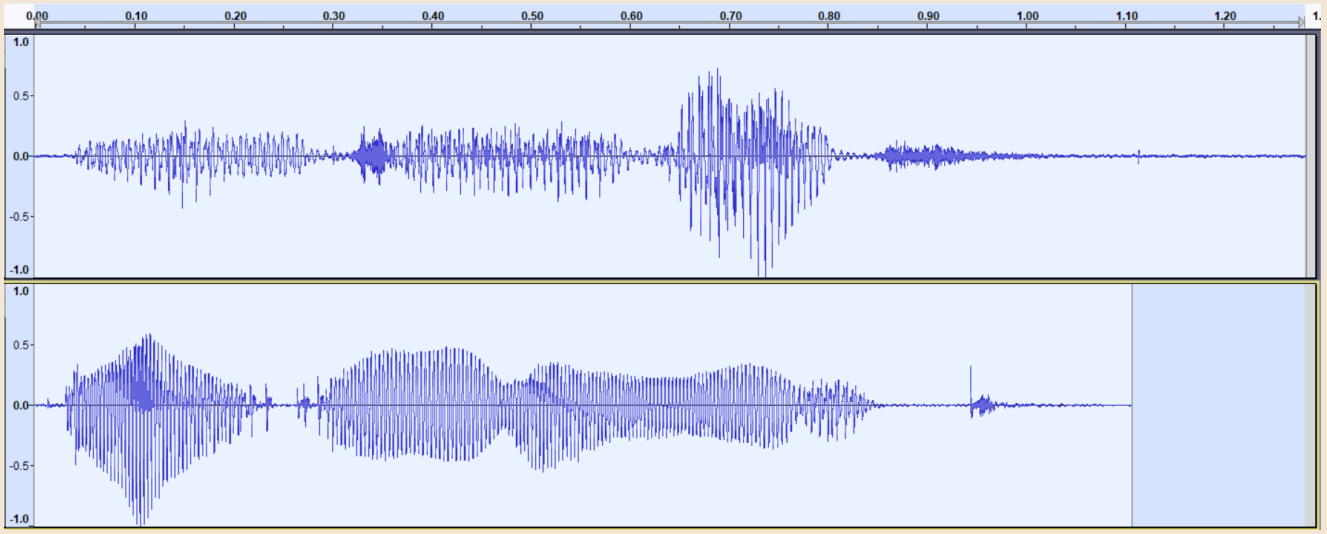
Summary

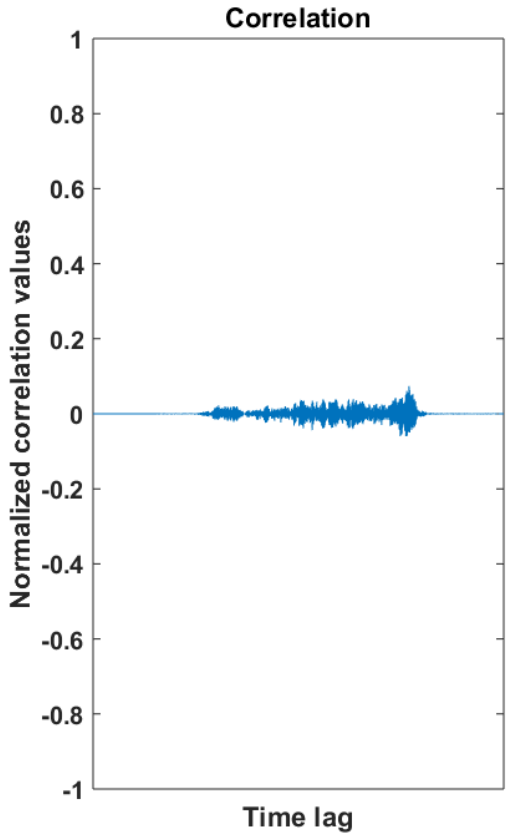
1. Non-stationary
2. Time-varying nature
3. Unknown Segment Boundaries

Non-stationary nature

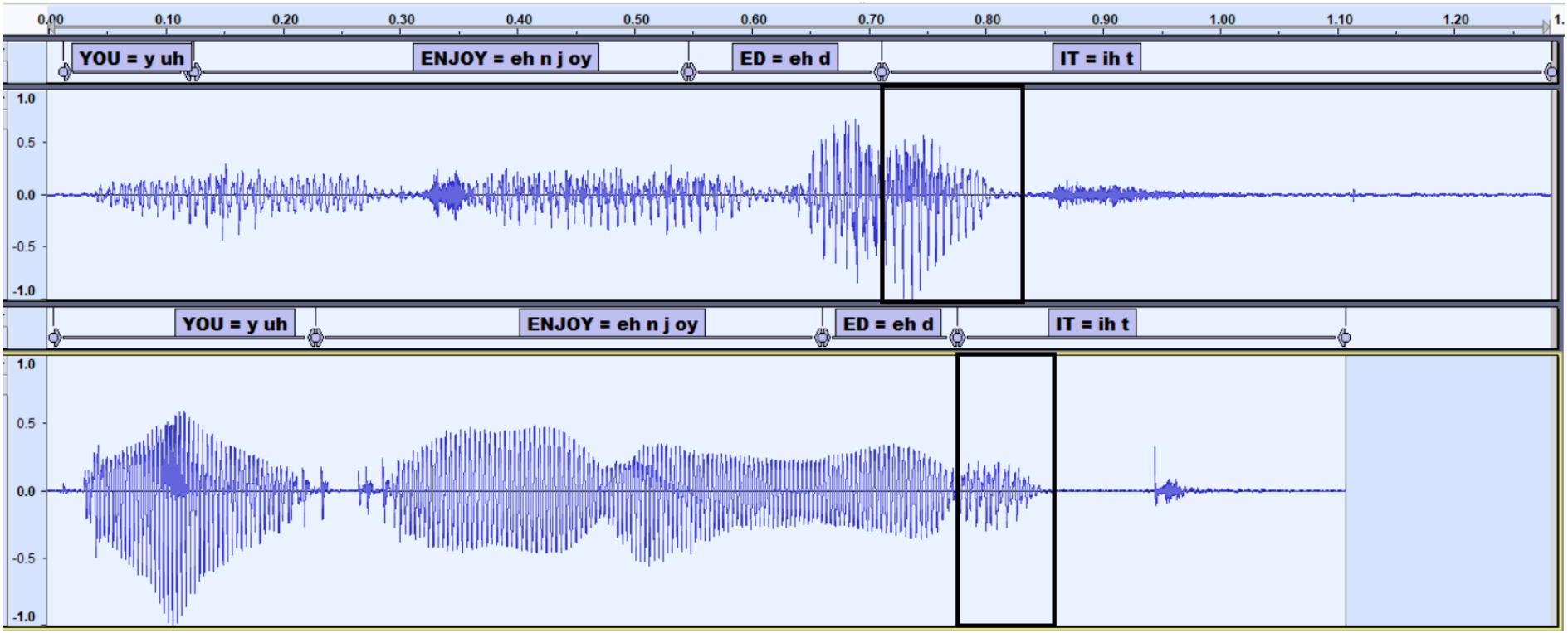
Illustrative example

Sentence: "You enjoyed it"; Signal #1: , Signal #2:

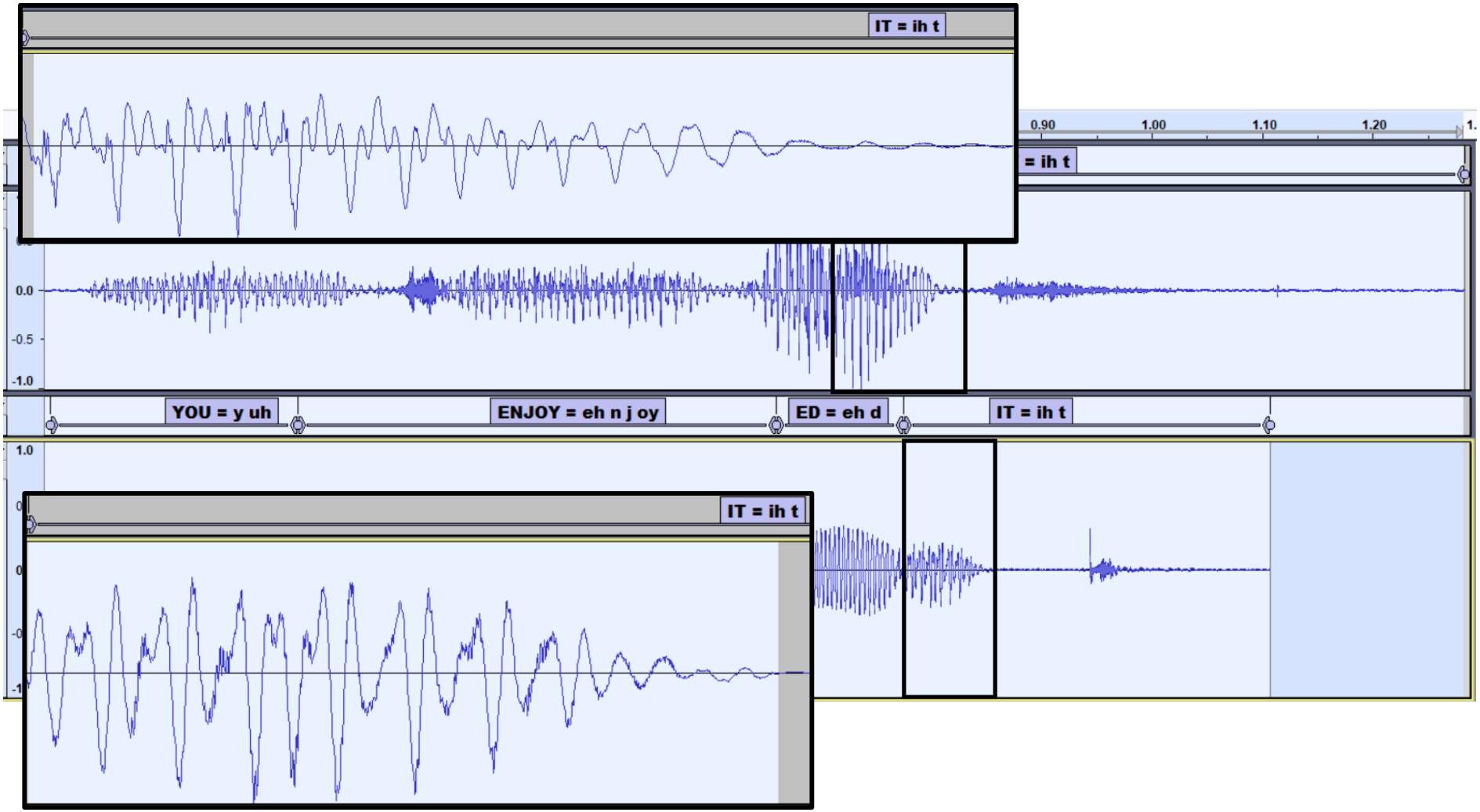




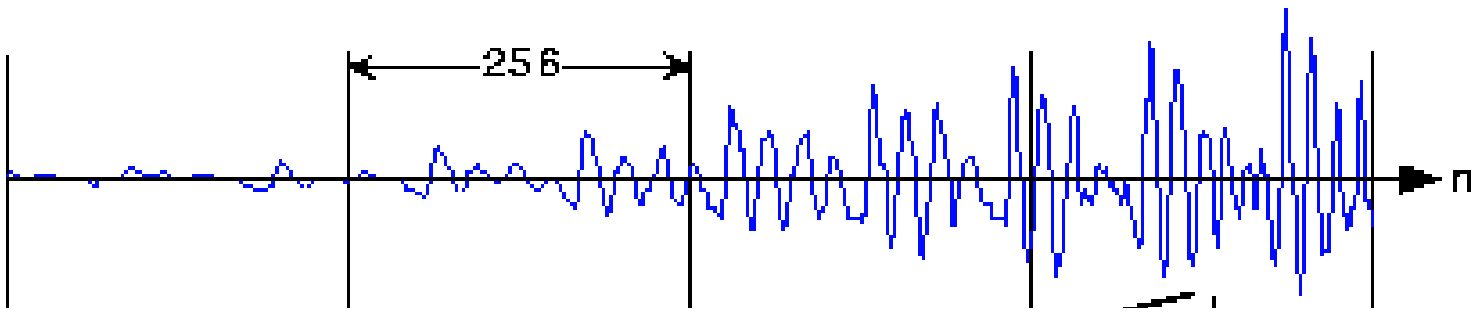
Non-stationary, but there is a structure in small segment



Non-stationary, but there is a structure in small segment



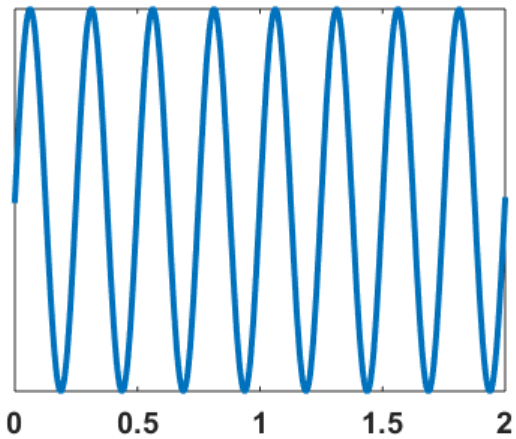
Stationary within the segment



- Signal analysis within the short segments
- Short-time energy, Short-time zero crossing rate, Short-time Fourier Transform etc..

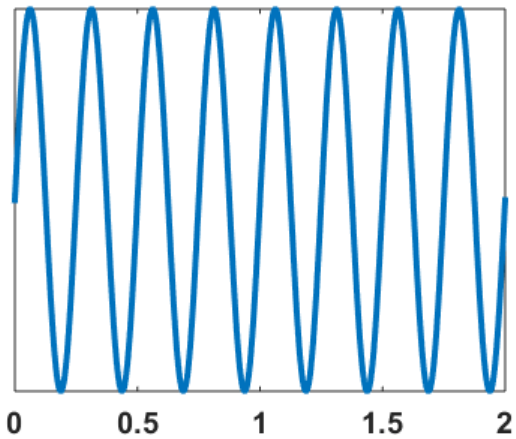
Fourier Transform (FT)

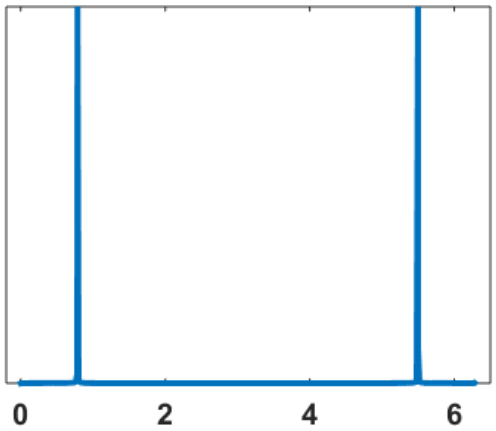
Sine signal:



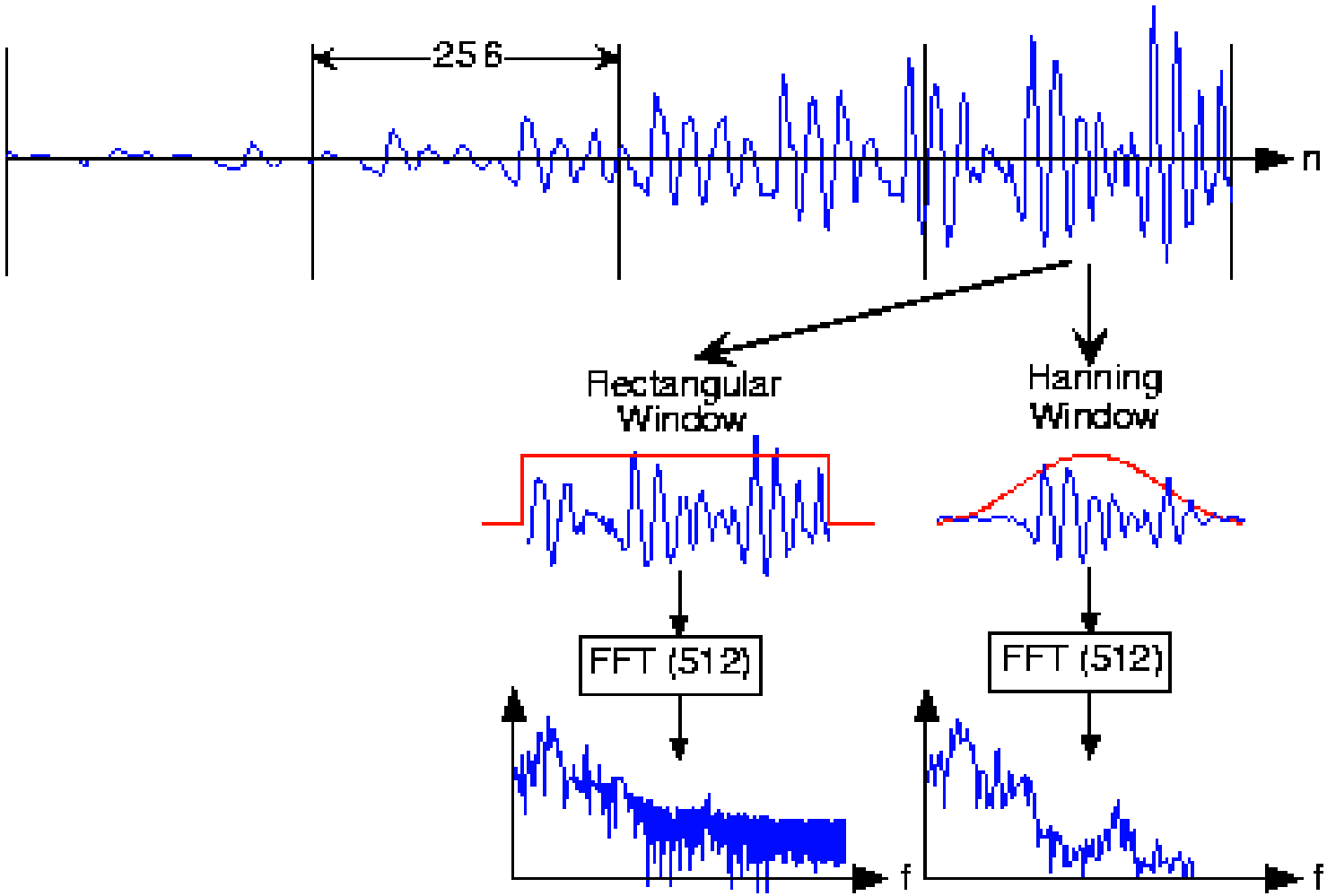
Fourier Transform (FT)

Sine signal:

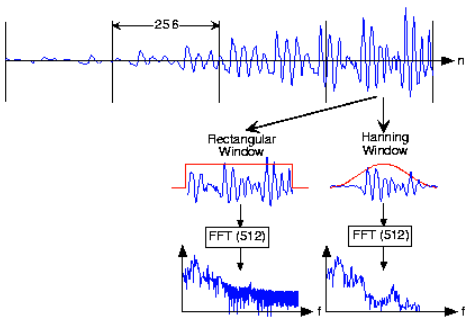


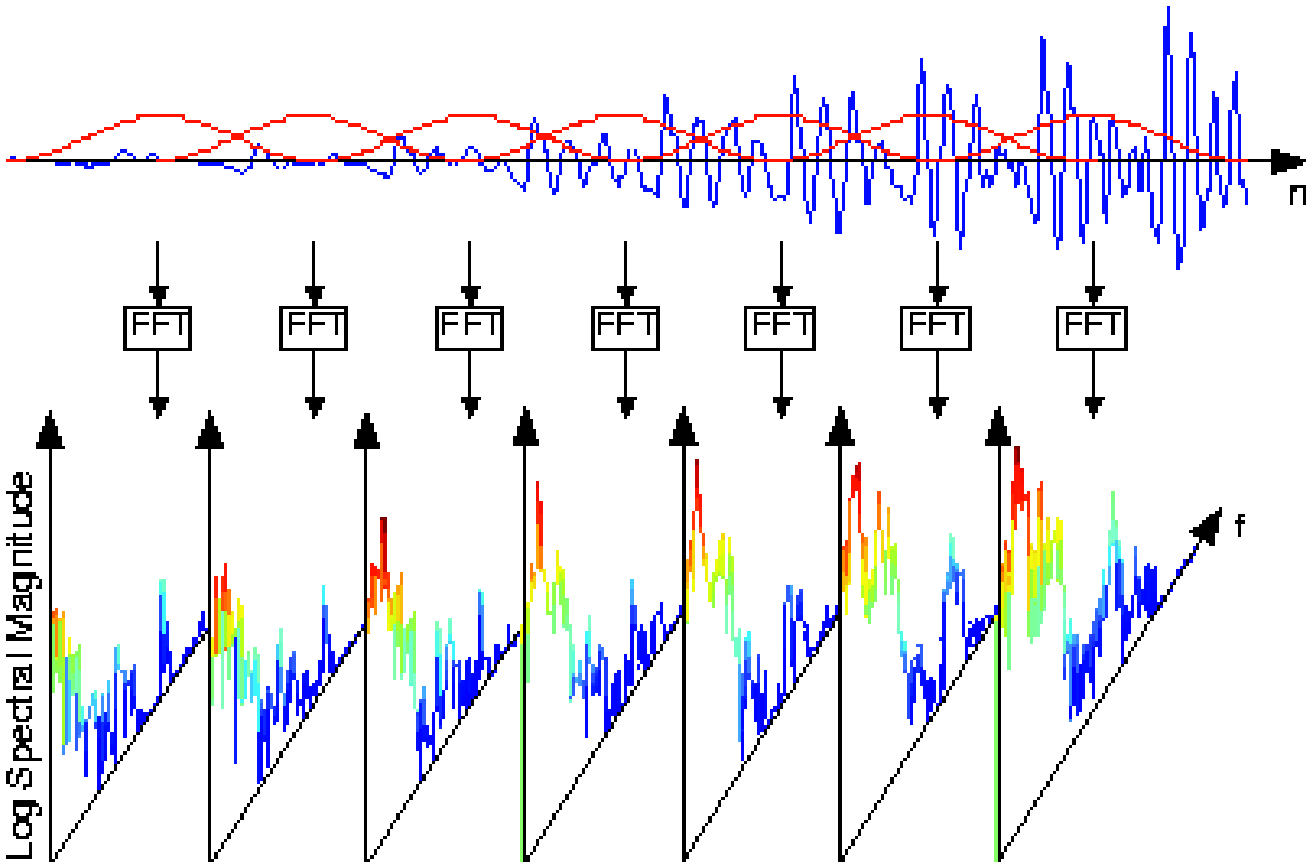


Short-time Fourier Transform (STFT)



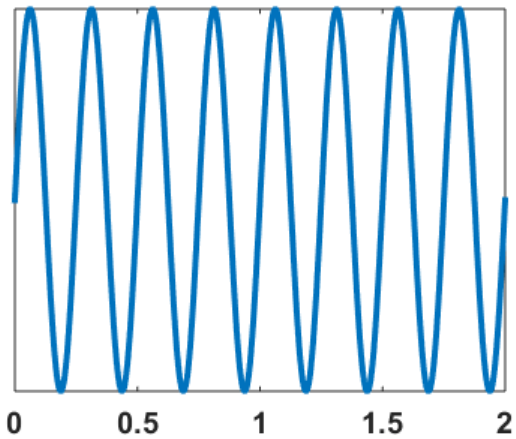
Short-time Fourier Transform (STFT)





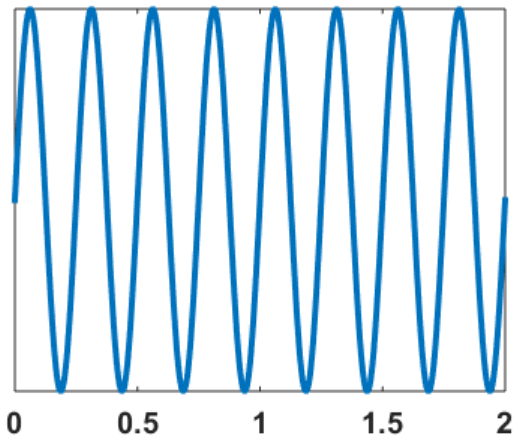
Short-time Fourier Transform (STFT)

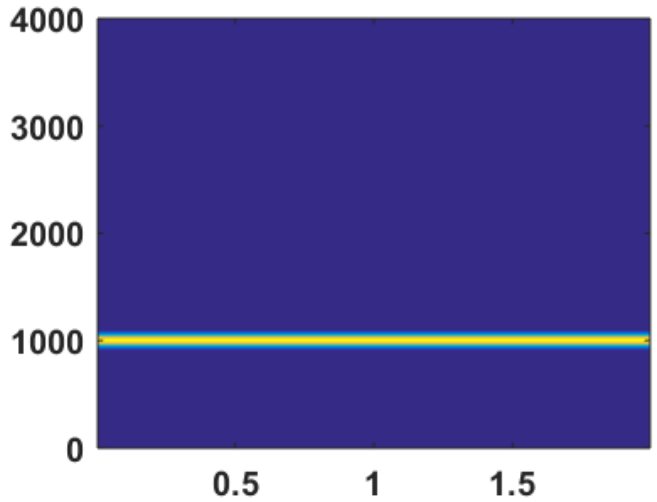
Sine signal:



Short-time Fourier Transform (STFT)

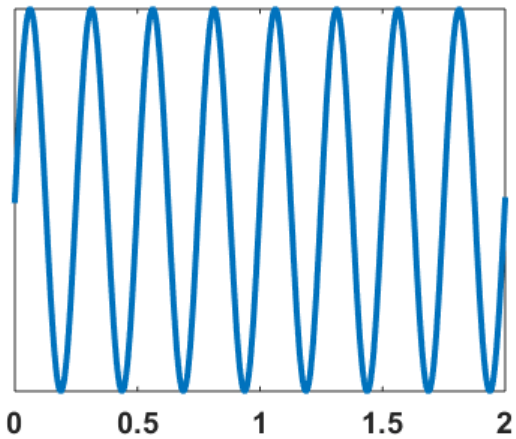
Sine signal:

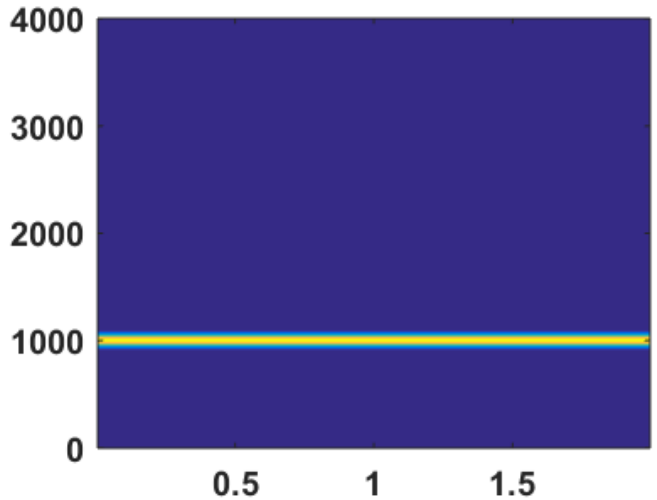




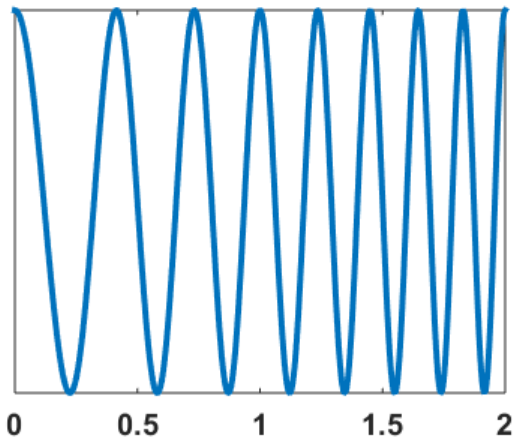
Short-time Fourier Transform (STFT)

Sine signal:



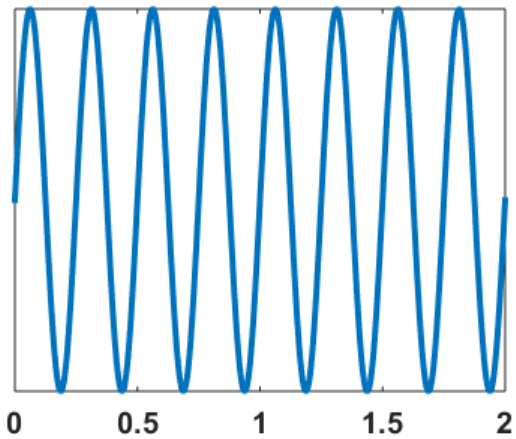


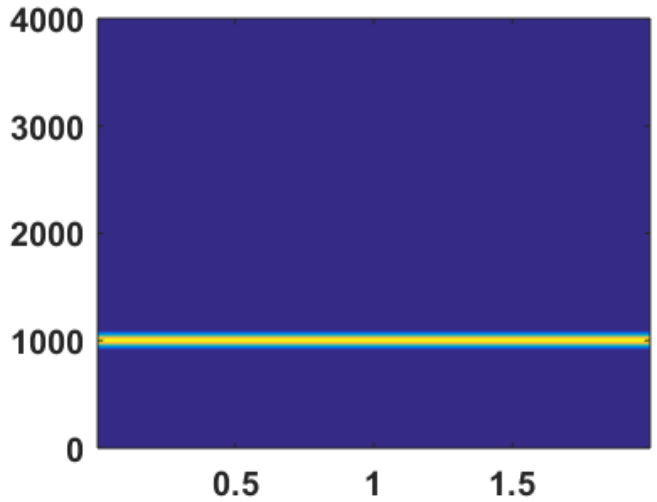
Chirp Signal:



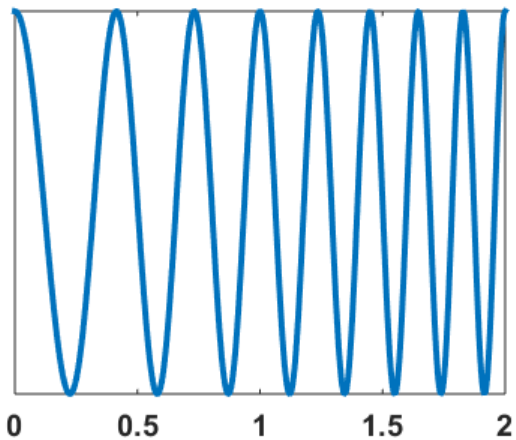
Short-time Fourier Transform (STFT)

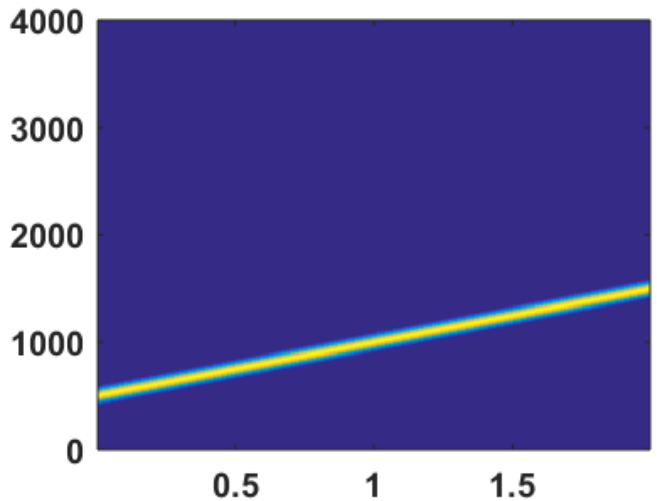
Sine signal:





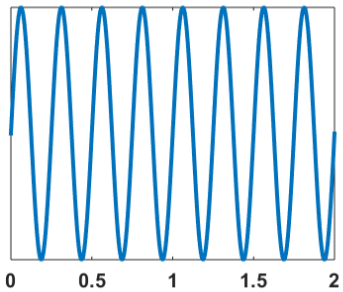
Chirp Signal:



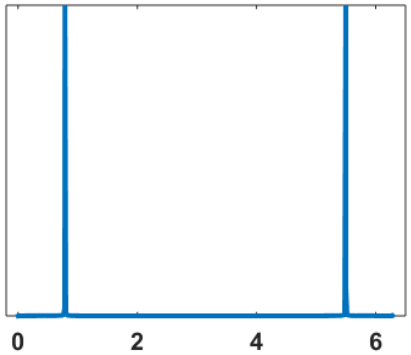


Why STFT?

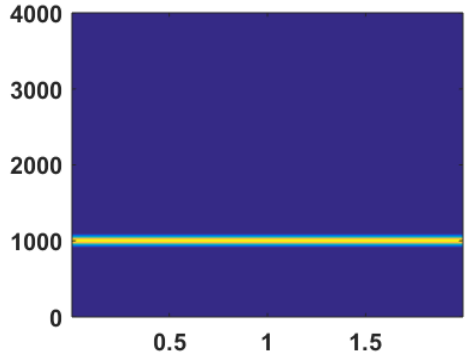
Sine signal:



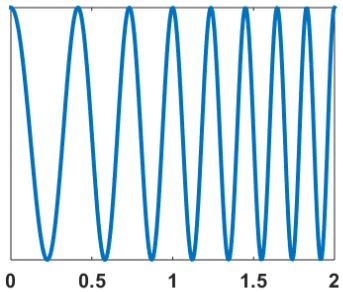
DFT:

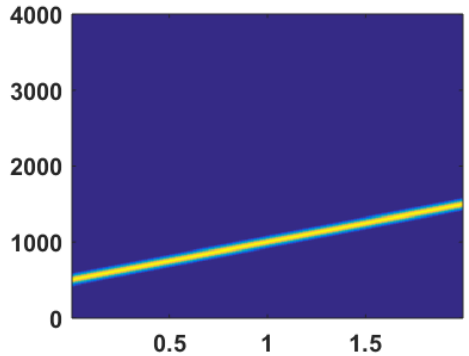


STFT:



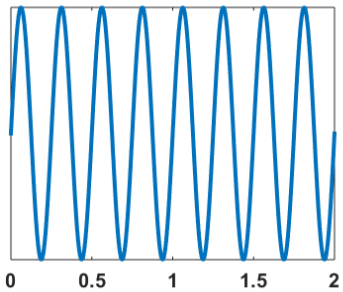
Chirp Signal:



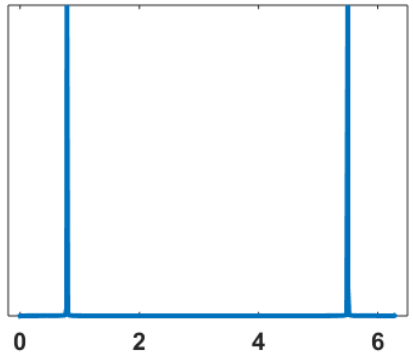


Why STFT?

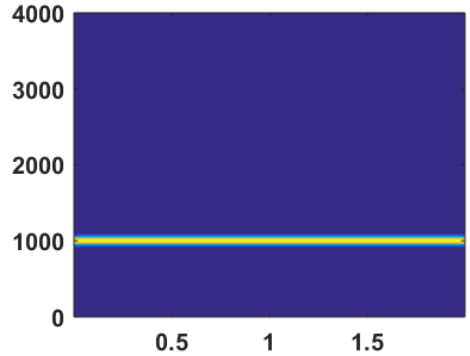
Sine signal:



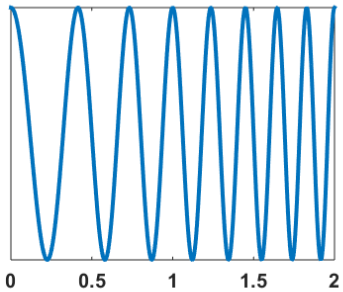
DFT:

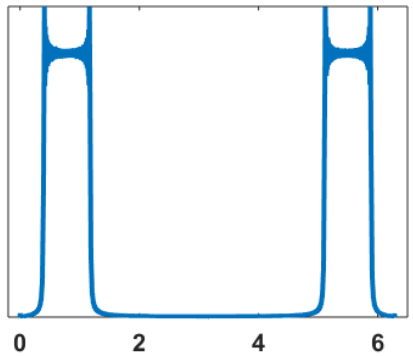


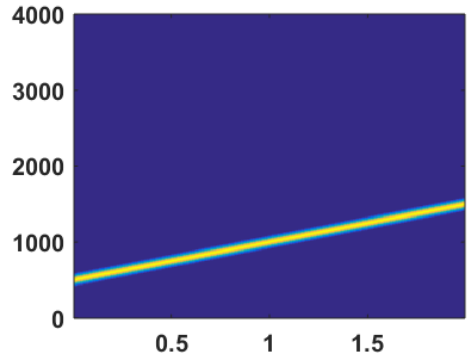
STFT:



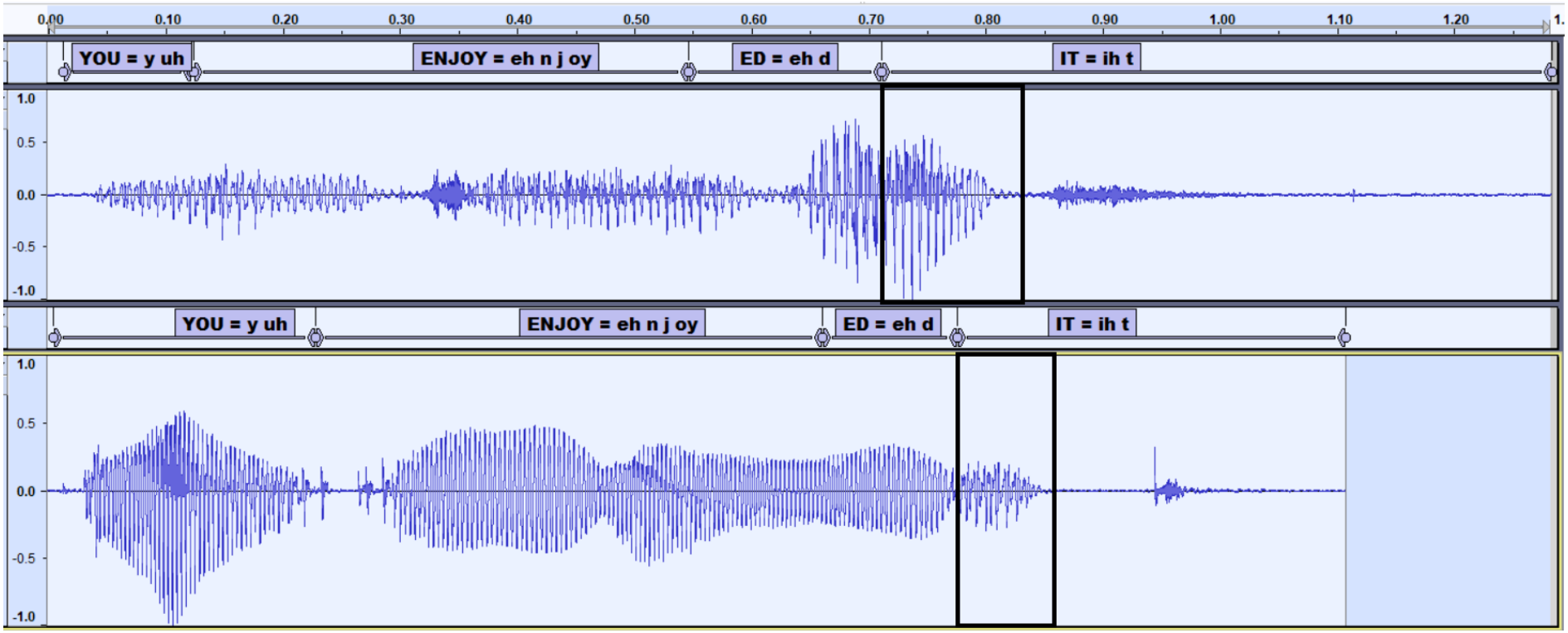
Chirp Signal:



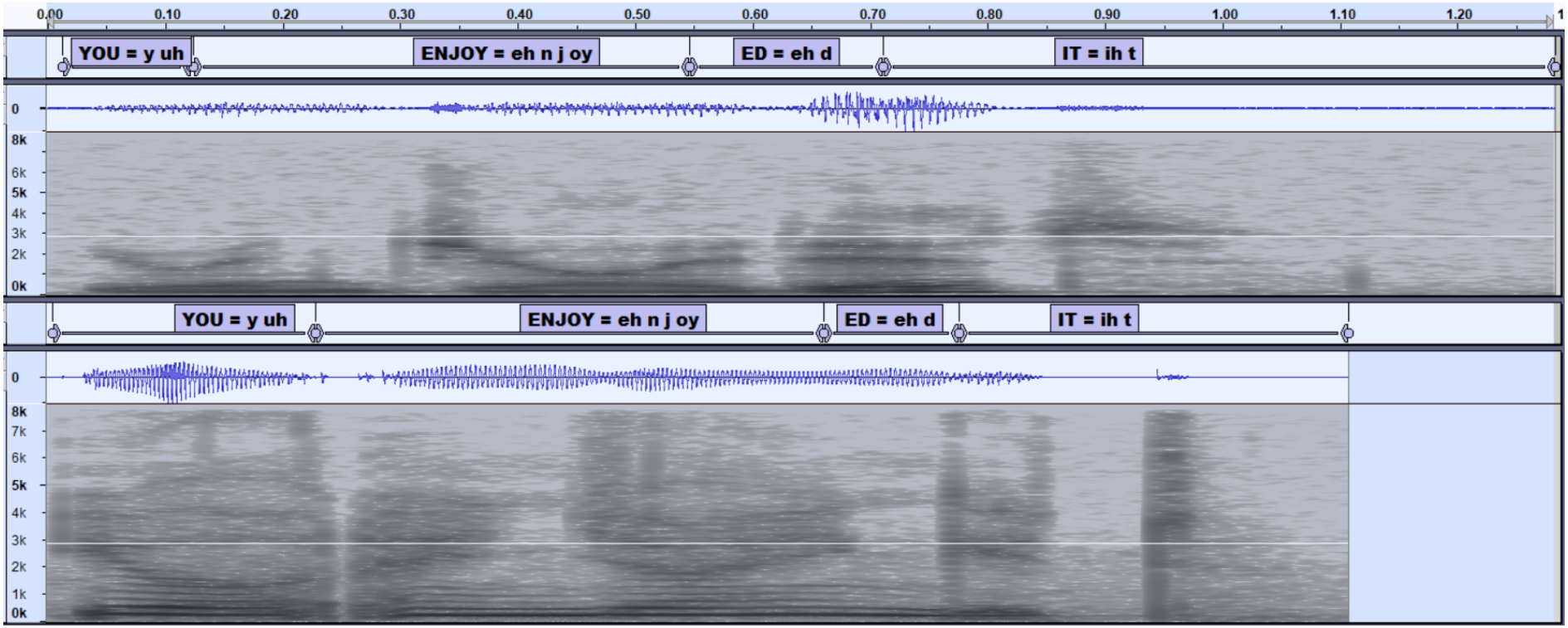




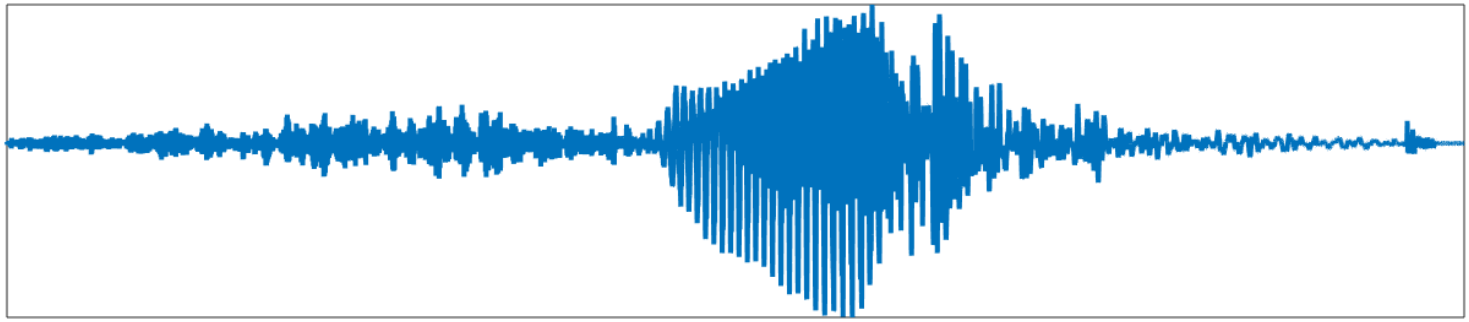
For the exemplary speech

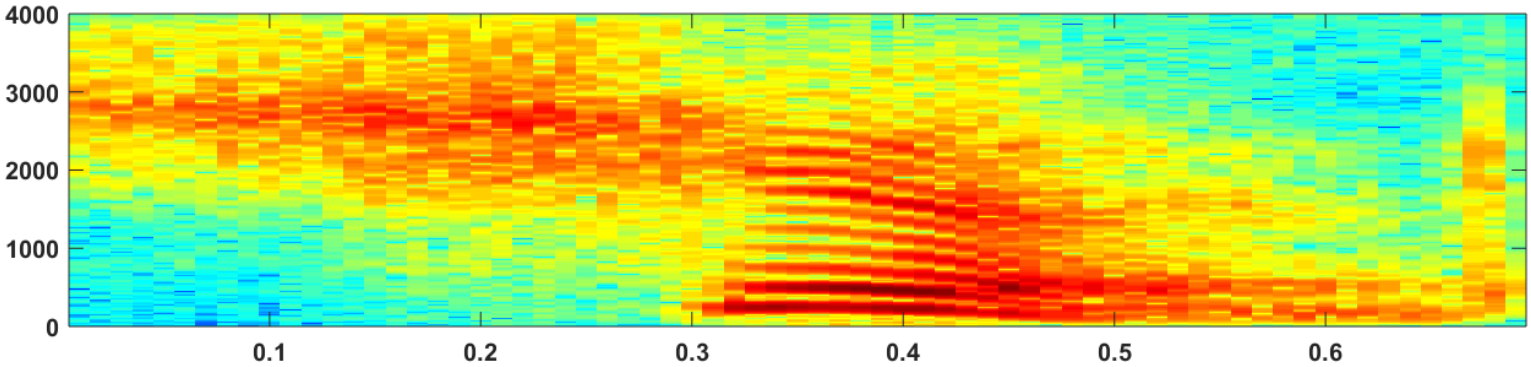


For the exemplary speech

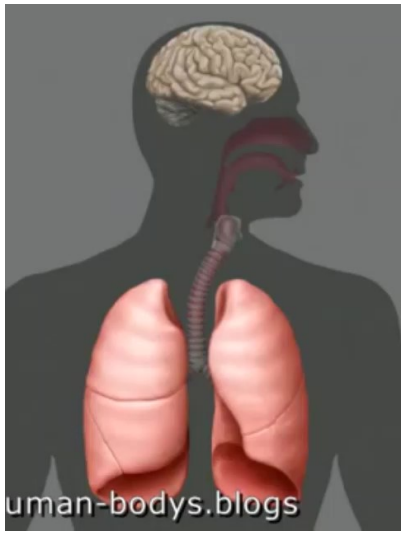


For an exemplary speech

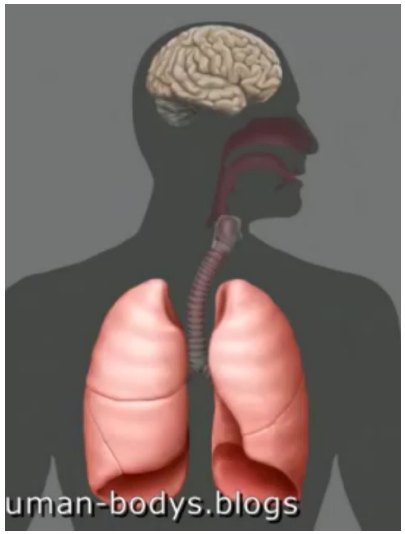


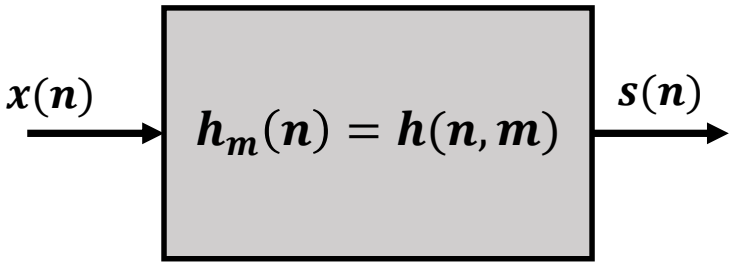


Speech production approximation



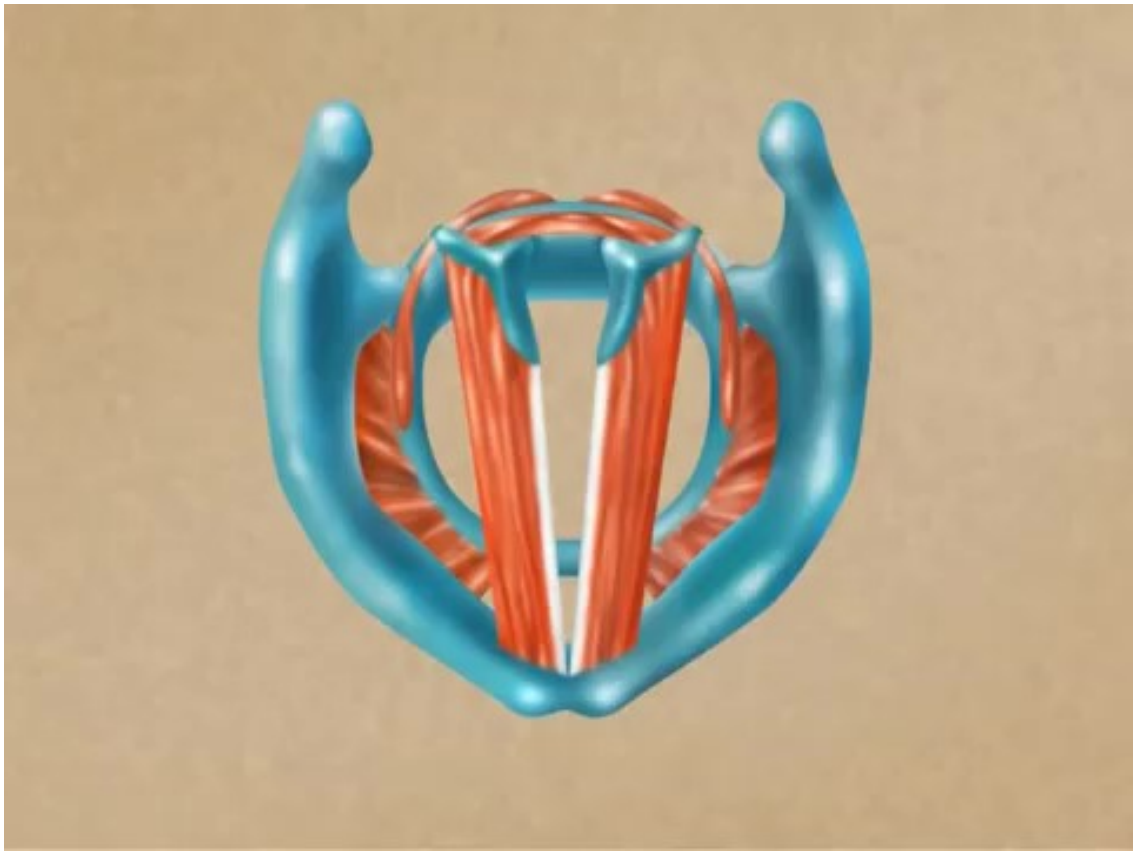
Speech production approximation



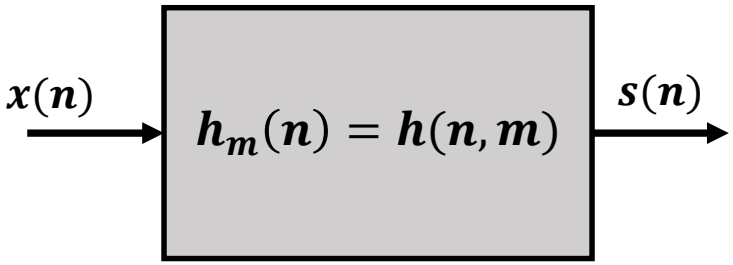


$$s(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(n-m)$$

Vocal chord movements

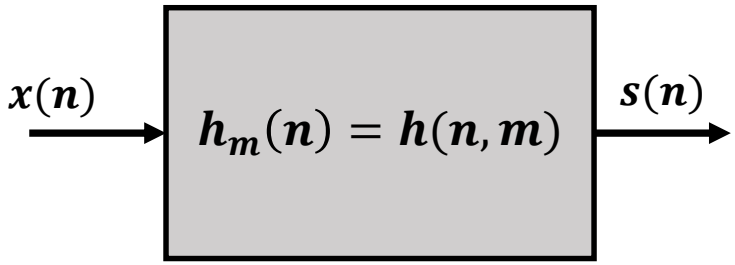


VuV based modelling

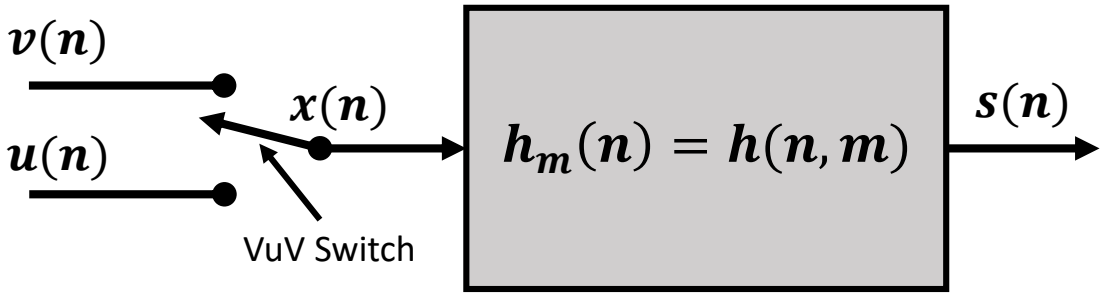


$$s(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(n-m)$$

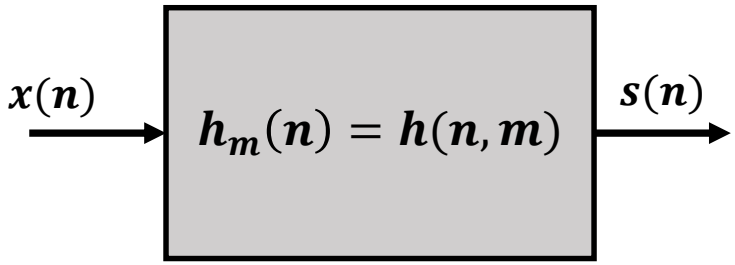
VuV based modelling



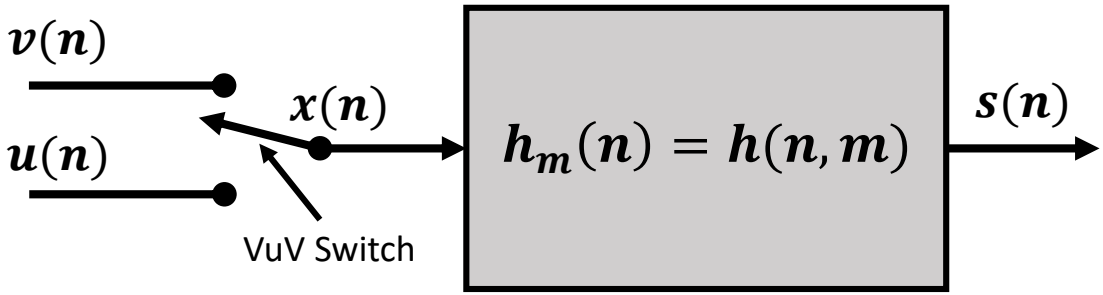
$$s(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(n-m)$$



VuV based modelling



$$s(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(n-m)$$



$$s_v(n) = \sum_{m=-\infty}^{\infty} h(n,m)v(n-m); \quad s_u(n) = \sum_{m=-\infty}^{\infty} h(n,m)u(n-m)$$

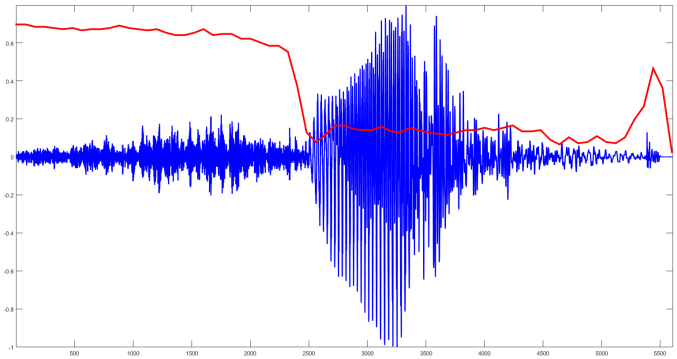
VuV detection

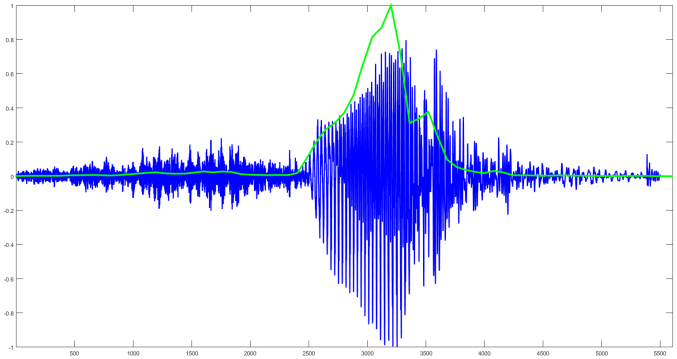
VuV detection

- Short-time zero crossing rate
- Short-time energy

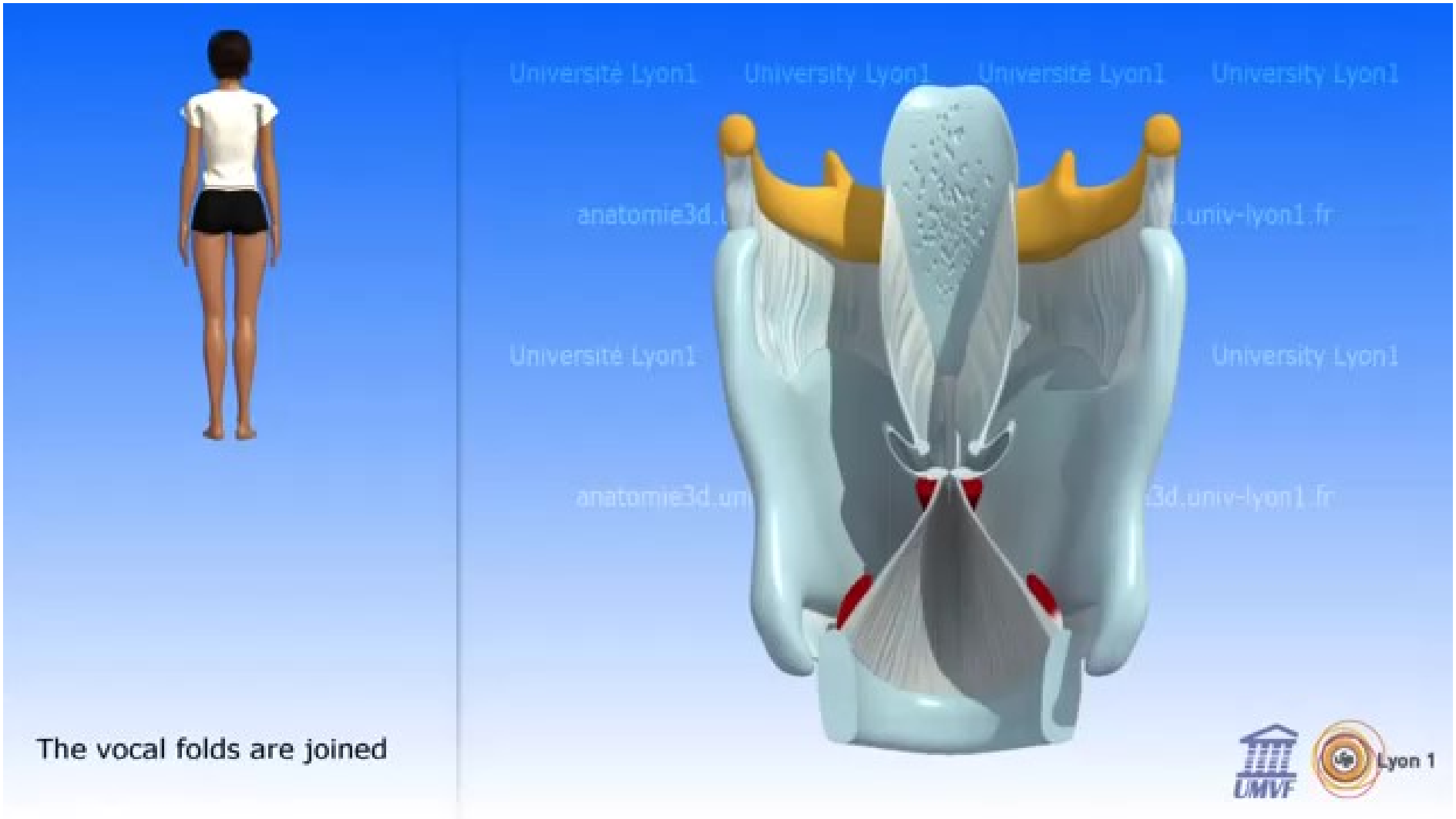
VuV detection

- Short-time zero crossing rate
- Short-time energy

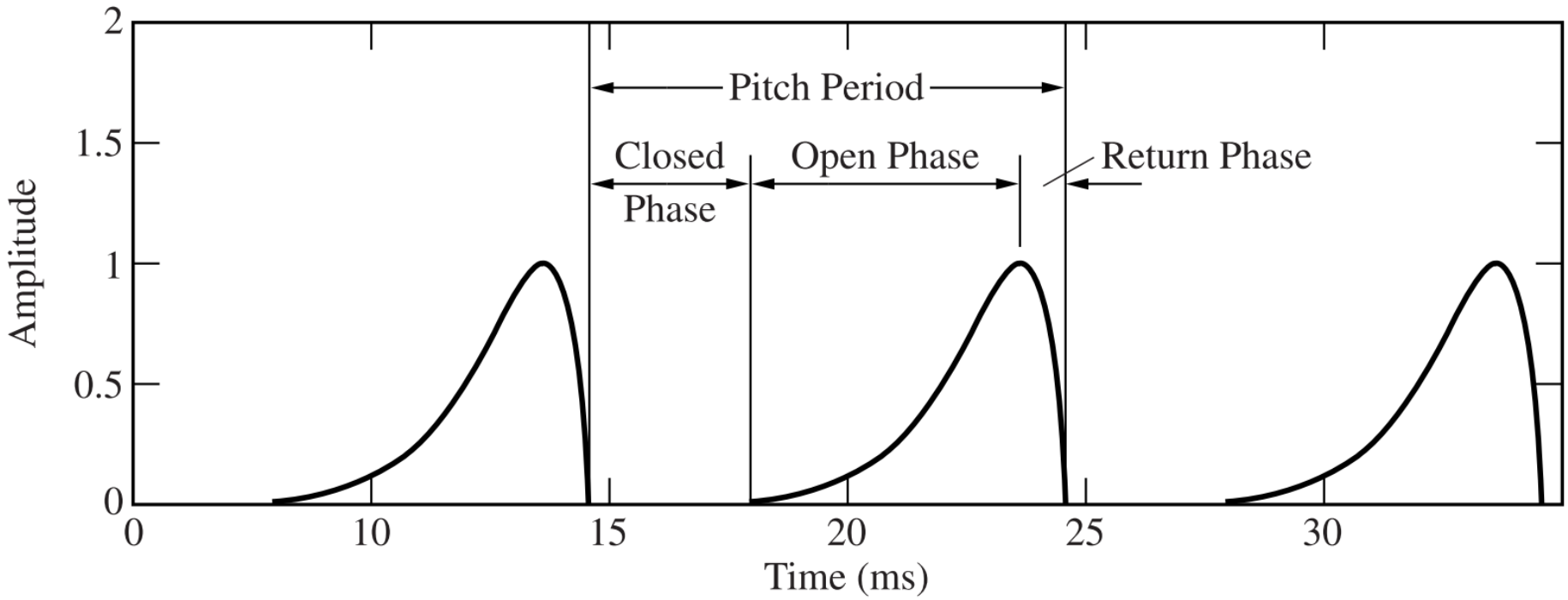




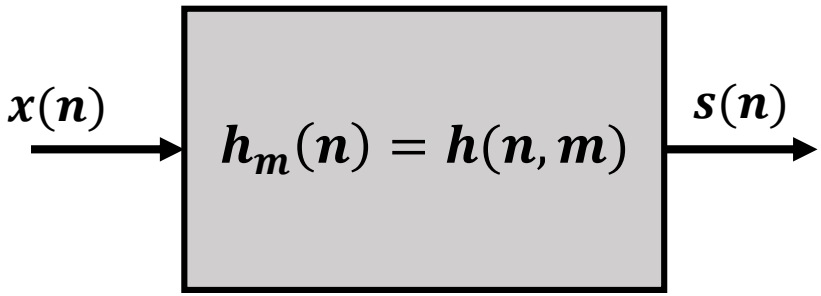
Airflow movement at the vocal chords



Glottal airflow velocity

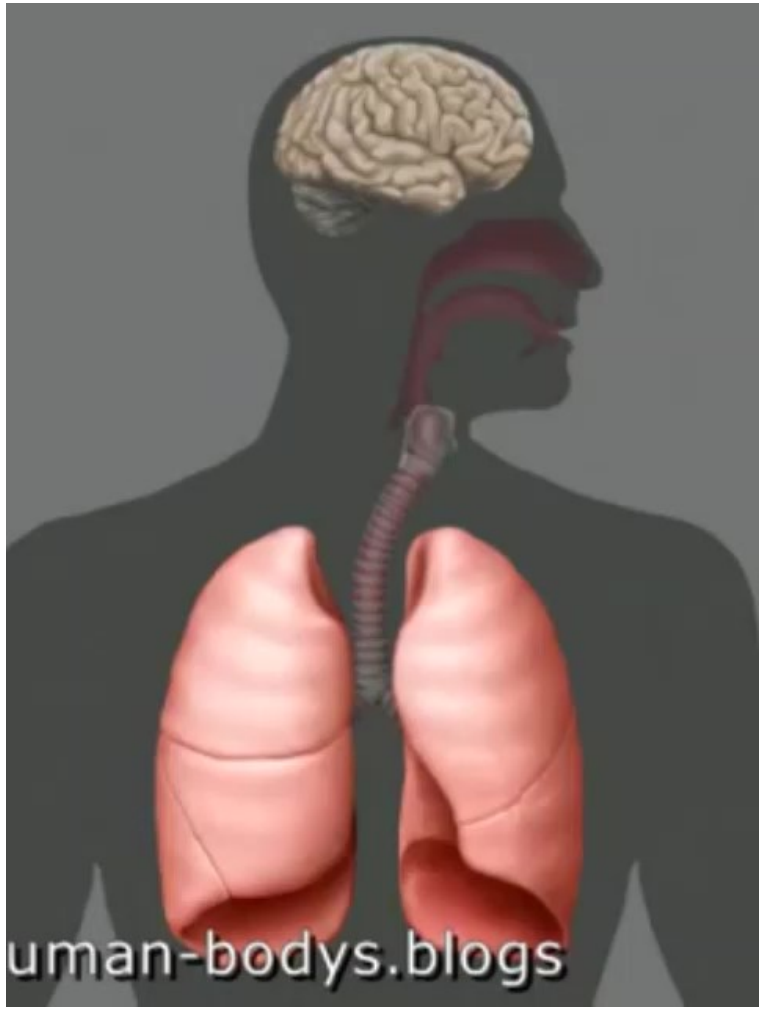


Vocal-tract modelling

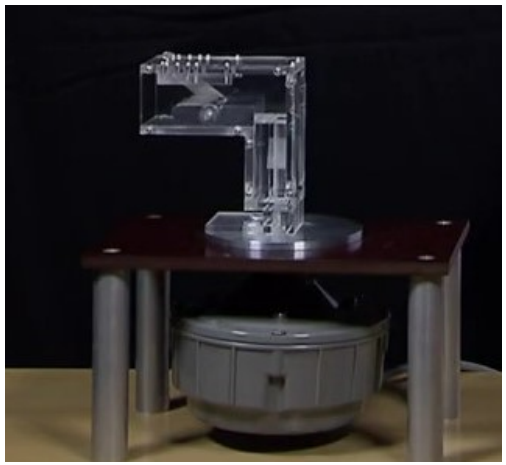


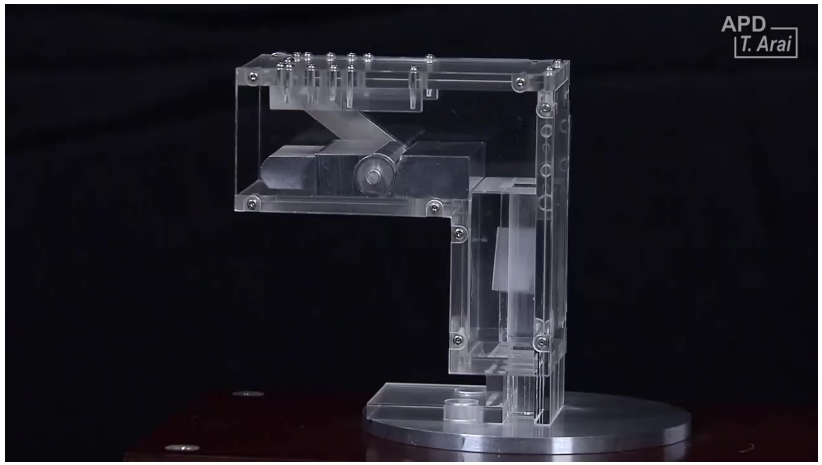
$$s(n) = \sum_{m=-\infty}^{\infty} h(n,m)x(n-m)$$

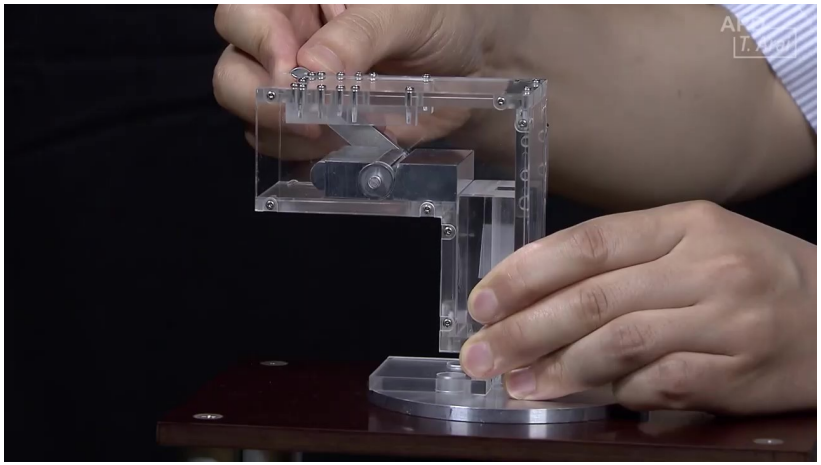
Vocal-tract approximation - cascade of tubes

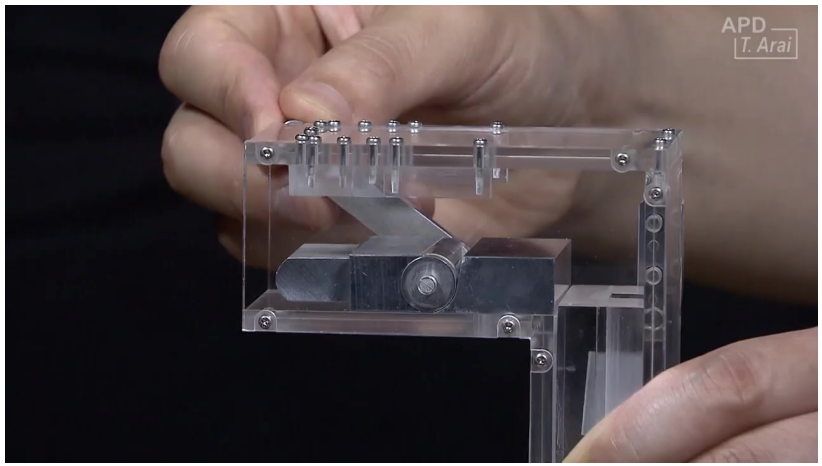


Illustrations of approximation

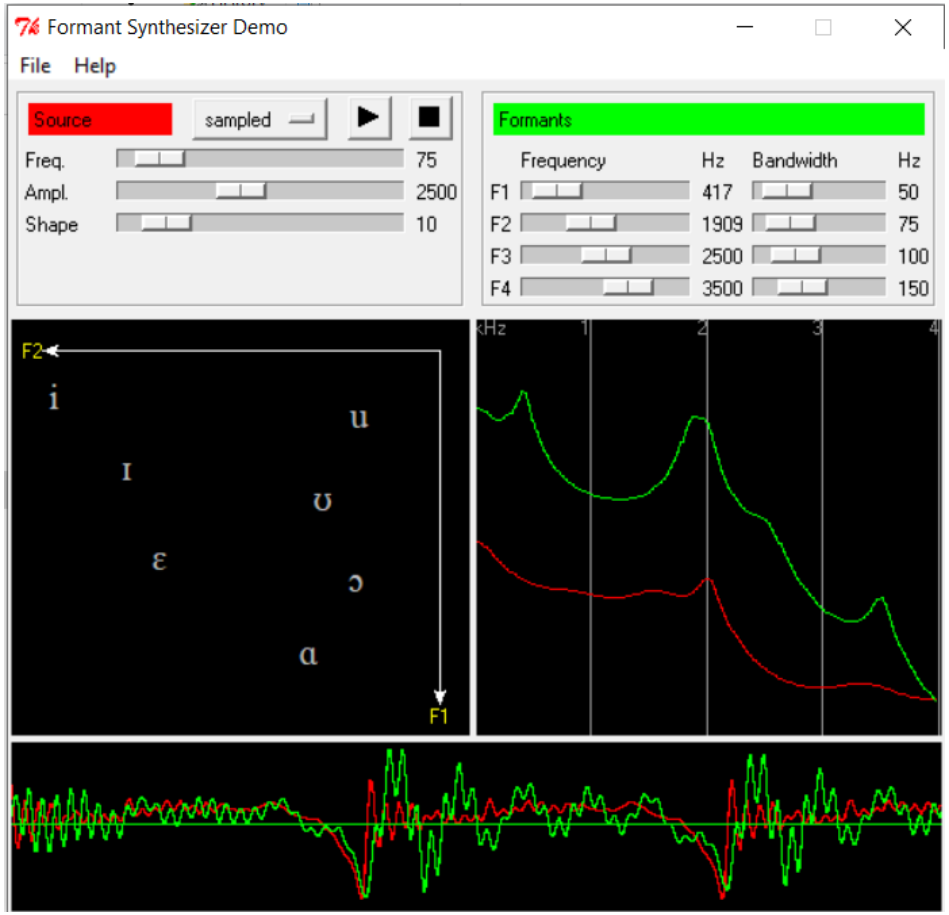








Vocal tract modelling



Conclusion

- What is nature of the speech signal?
- What is short-time analysis and frames?
- What is relative lengths of frames and segments?
- What are the voiced and unvoiced sounds?
- What is pitch?
- What are the formant frequencies?
- How the sine signal characterize?
- How the STFT (spectrogram) is useful over DFT?

Thank you