Mobile Health
Lecture 5
Audio Signal and Health

Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

Another Cheap Sensor

UNIVERSITY OF CAMBRIDGE

Automated Sound based Diagnostics

Voice-based Diagnostics

MIT Technology Review

Artificial Intelligence / Machine Learning

Voice Analysis Tech Could Diagnose Disease

Researchers enlist smartphones and machine learning to find vocal patterns that might signal post-traumatic stress disorder or even heart disease.

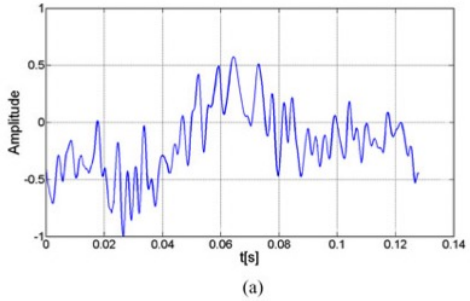
by **Emily Mullin** Jan 19, 2017

UNIVERSITY OF CAMBRIDGE

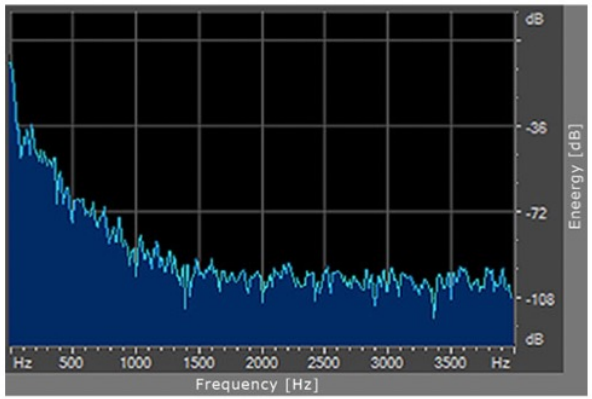
Type of diseases for which audio has been tried on

- Depression and PTSD
- Sleep Apnea
- Wheezing (Asthma)
- Parkinson’s
- Alzheimer’s
- Autism
- Cardiovascular: coronary heart disease, arteriosclerosis
- …

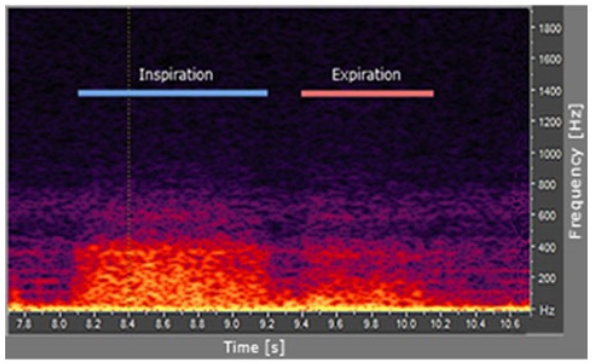
Wheezing



(a)



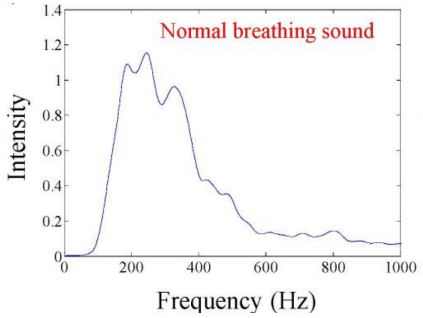
(b)

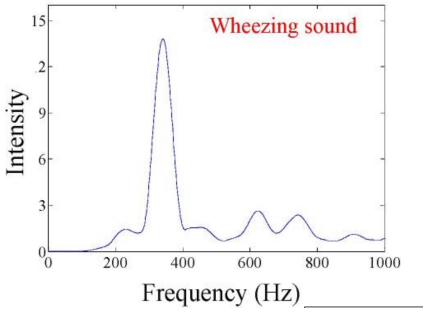


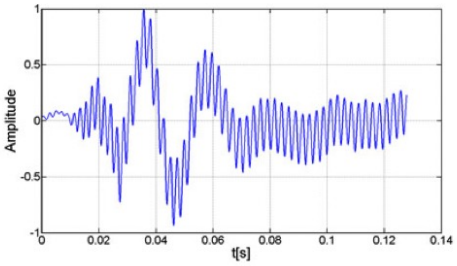
(c)

<- Normal

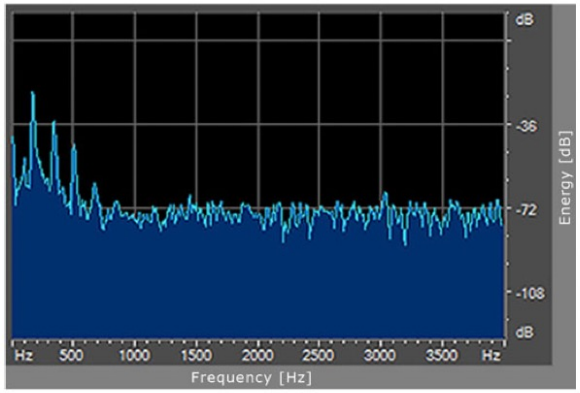
Diseased ->



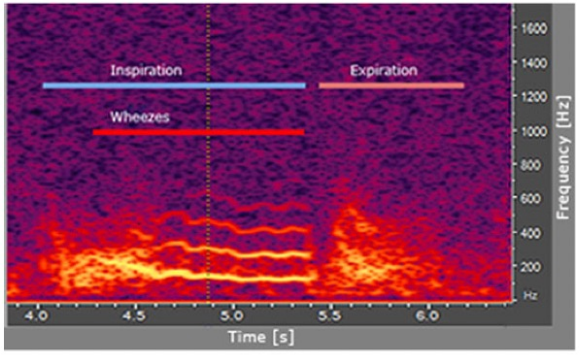




(a)



(b)



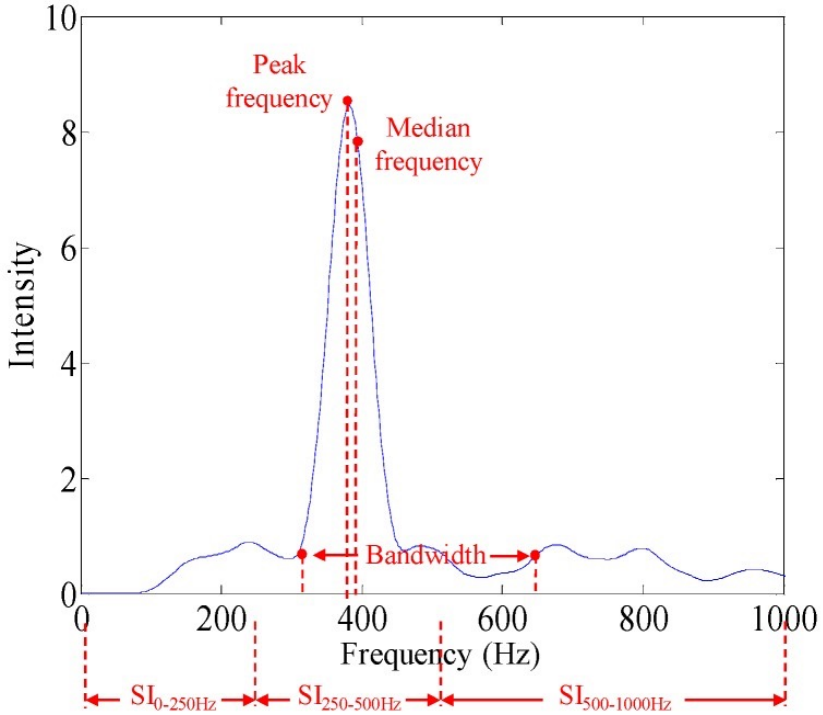
(c)

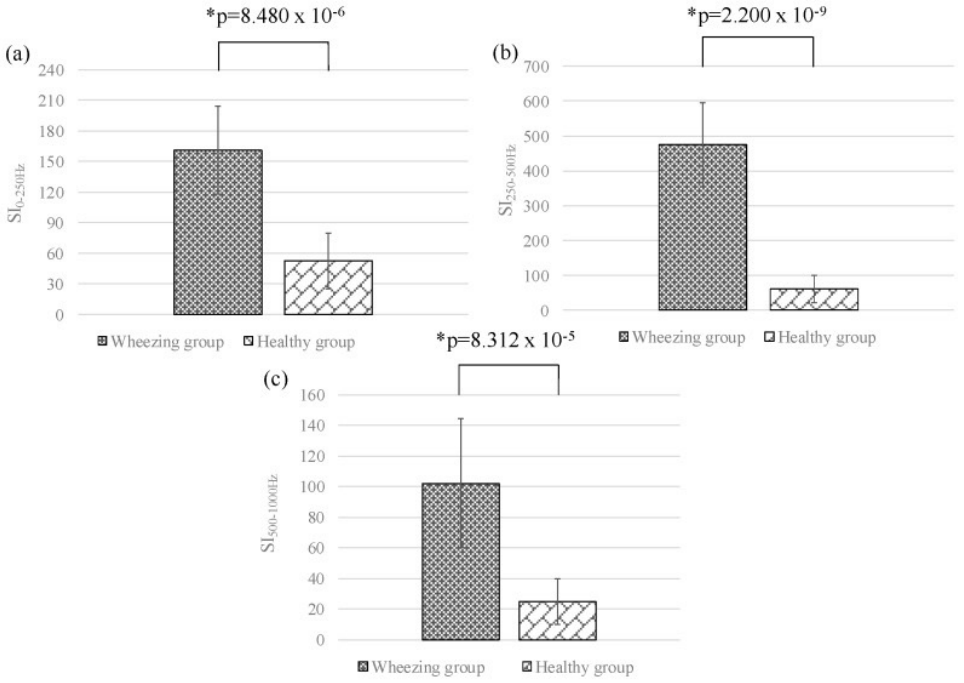
Preprocessing of this signal

- Band pass filter applied: (150 Hz–1000 Hz)
  - Limit heart sound, muscle interference sound and blood sound.
- Raw breathing sound split into 250-ms breathing sound segments with 200-ms overlapping.
- Power spectrums of these breathing sound segments will be calculated by using FFT with a Hann window.

Li SH, Lin BS, Tsai CH, Yang CT, Lin BS. Design of Wearable Breathing Sound Monitoring System for Real-Time Wheeze Detection. Sensors (Basel). 2017 Jan 17;17(1):171.

Wheezing Features





UNIVERSITY OF CAMBRIDGE

Li SH, Lin BS, Tsai CH, Yang CT, Lin BS. Design of Wearable Breathing Sound Monitoring System for Real-Time Wheeze Detection. Sensors (Basel). 2017 Jan 17;17(1):171.

Important Features

ASE= audio spectral envelope
TI = tonal index
CF1= correlation feature
ER = Energy ratio
K= Kurtosis, Difference to mean ratio
EVD = Eigen Value Decomposition
VC = Vector Comparison
LP = Linear prediction
SPE = Spectral peak entropy
SF = Spectral flatness

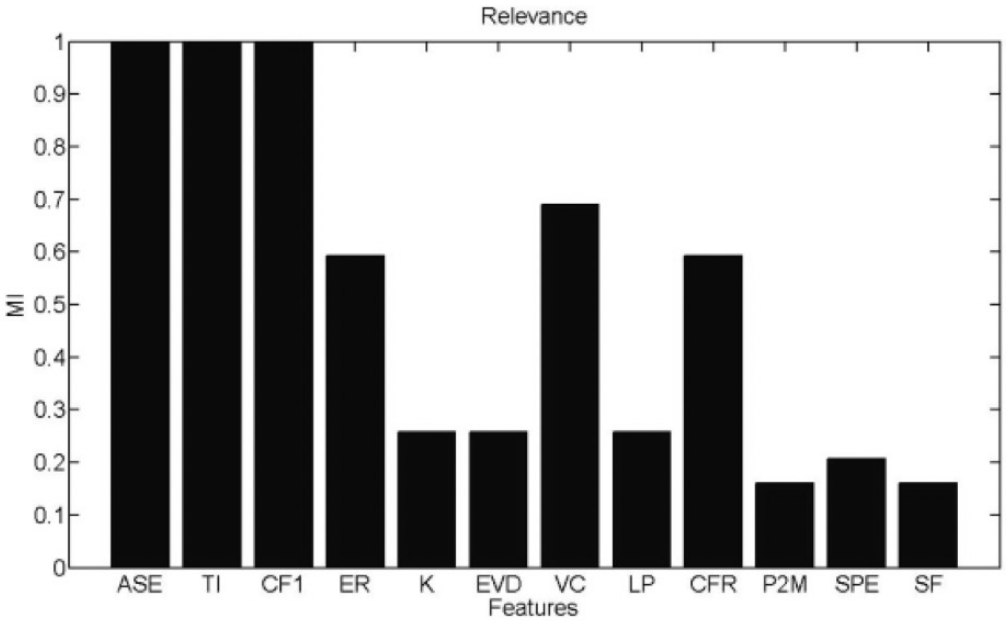


Fig. 2. Relevance of the tested features calculated for real lung recordings using the mRMR algorithm

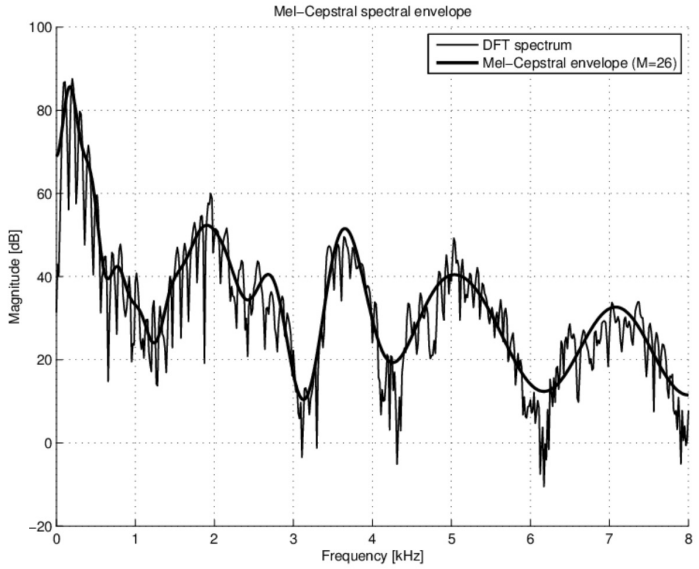
UNIVERSITY OF CAMBRIDGE

MRMR based feature selection for automatic asthma wheezing recognition. M. Wisniewski, T. Zielinski. Signals and Electronic Systems (ICSES), 2012

Frequency Spectrum Envelope

- Identifies the max frequency of each frame and creates a curve which represents all those maximum frequencies.

- Sometimes the frequency spectrum is “smoothened” first (e.g. with a log in cepstral analysis)

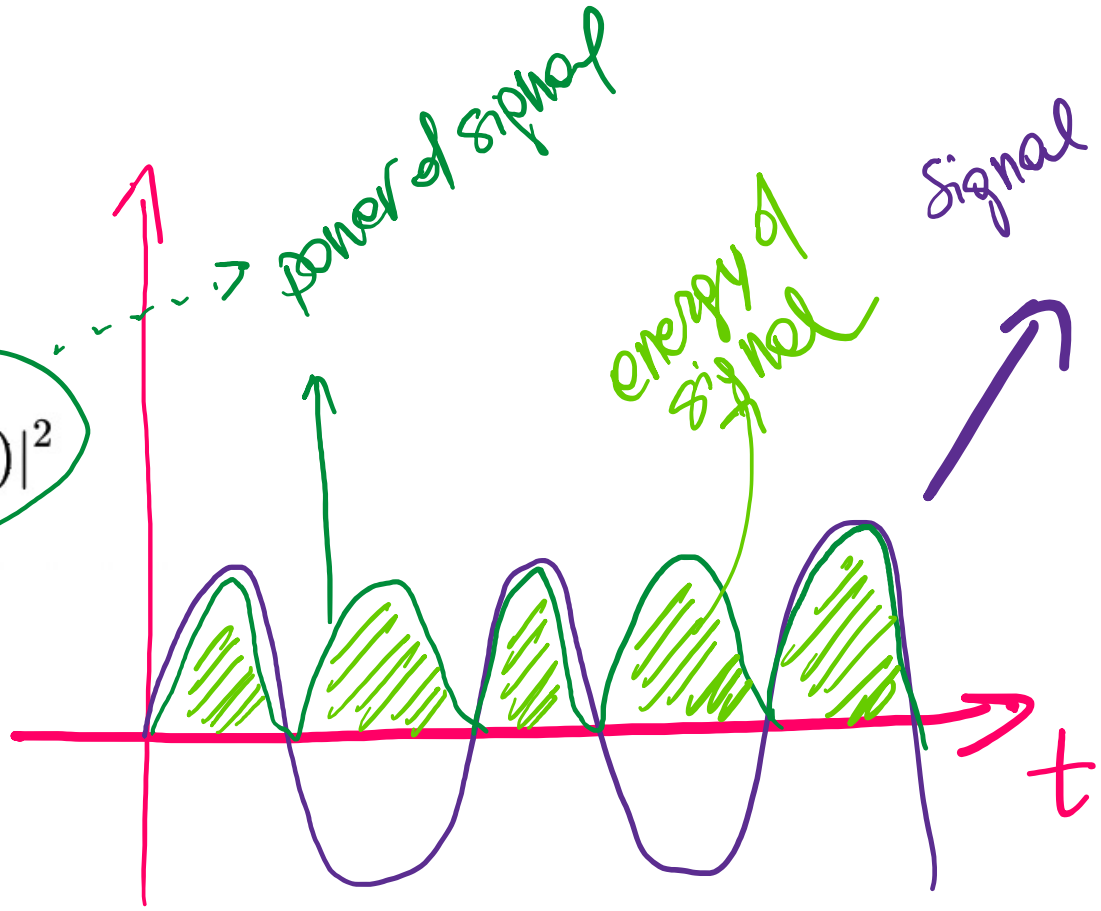


UNIVERSITY OF CAMBRIDGE

Figure from F. Tesser, E. Zovato, P. Cosi (2022). Statistical spectral envelop transformation applied to emotional speech. 13th Int. Conference on Digital Audio Effects. 2010.

Energy of the Signal

$$\sum_{n=-\infty}^{\infty} |x(n)|^2$$



Linear Prediction Coefficients

- Technique that calculates coefficients of a linear prediction model which predicts the next sample of audio from a sequence of k previous samples. The coefficient of this linear model are those coefficients. This technique is used to predict **pitch period** accurately (ie the period at which a signal pattern repeats).

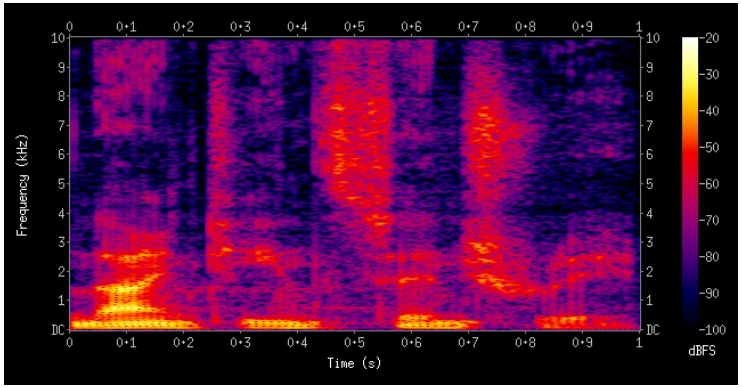
MFCC (Mel-frequency cepstral coefficients) Features

- Very used in audio processing

- The intuitively match our auditory way of perceiving sounds

From Spectrograms to MFCCs

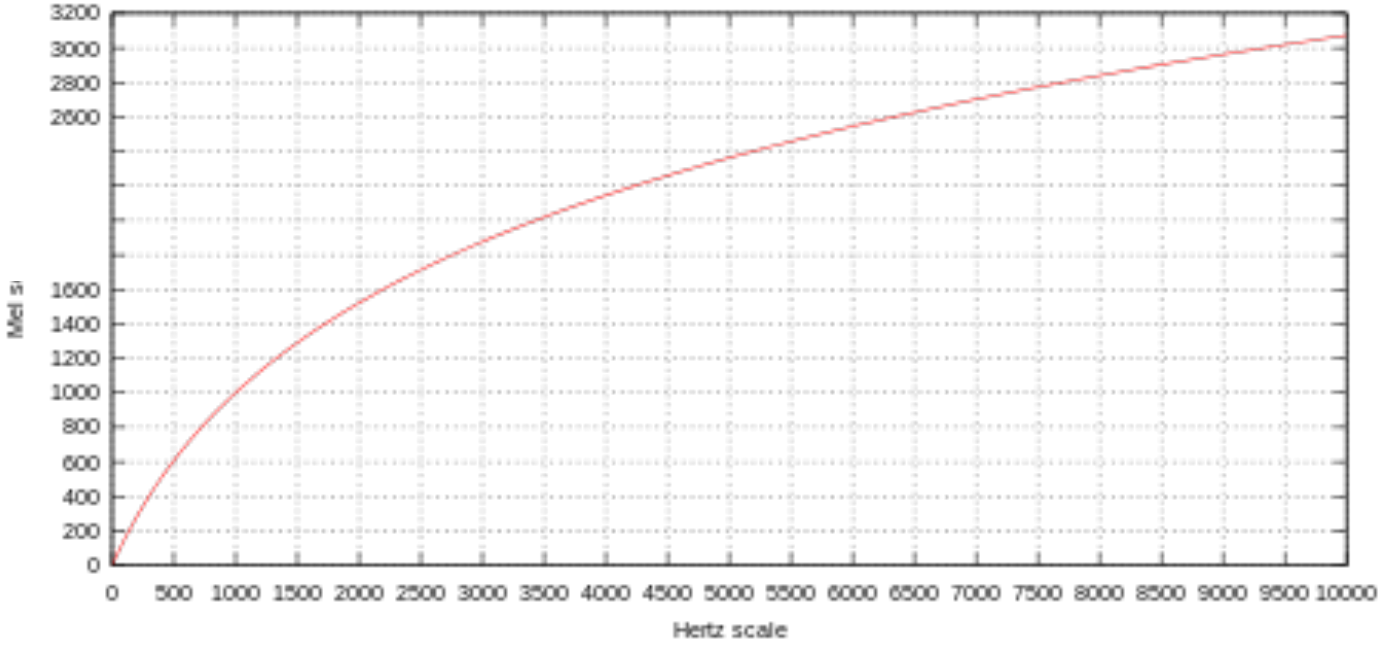
- Y axis in spectrogram is linear
- However humans perceive frequency of sound “logarithmically”
  - Difference between sounds at lower frequency seems more than distance of sounds at higher frequency



Mel Scale

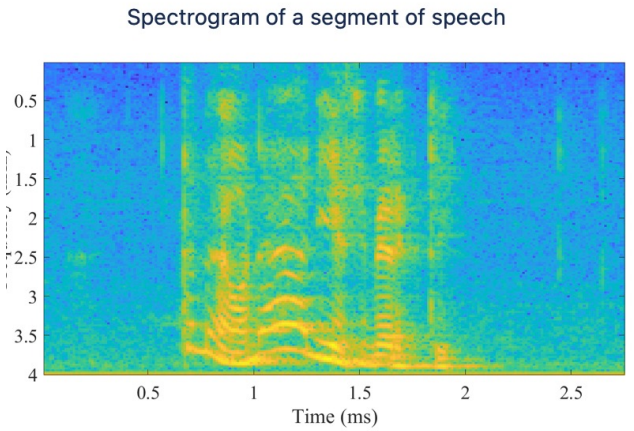
- Mel Scale is perceptually relevant scale for frequency
  - Matching our hearing

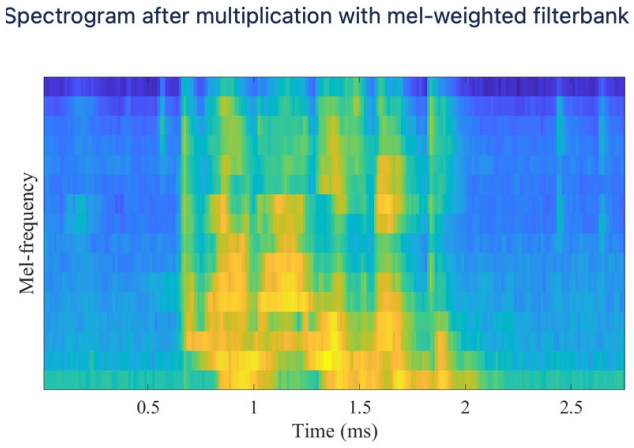
$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$



Spectrogram to Mel Spectrogram

- Mel spectrograms use Mel Filter banks
  - Effectively bins for the frequency mapping
- Shape retained. Fine structure smoothened.





UNIVERSITY OF CAMBRIDGE

Figure from
https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC

Mel Frequency Cepstral Coefficients

- Steps
  - DFT
  - Log
  - Apply Mel Scaling
  - Discrete Cosine Transform (a transformation which highlights the important parts of the log DFT graph)
  - Result is a number of “coefficients”: first 12-13 are the most relevant generally for audio

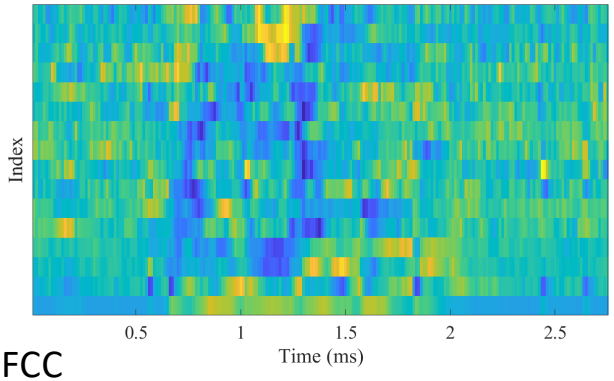
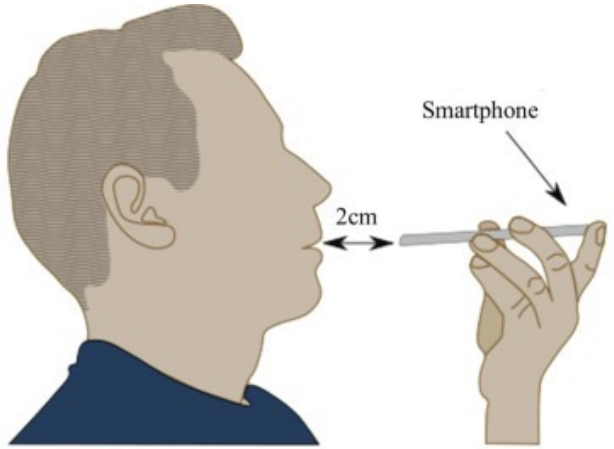


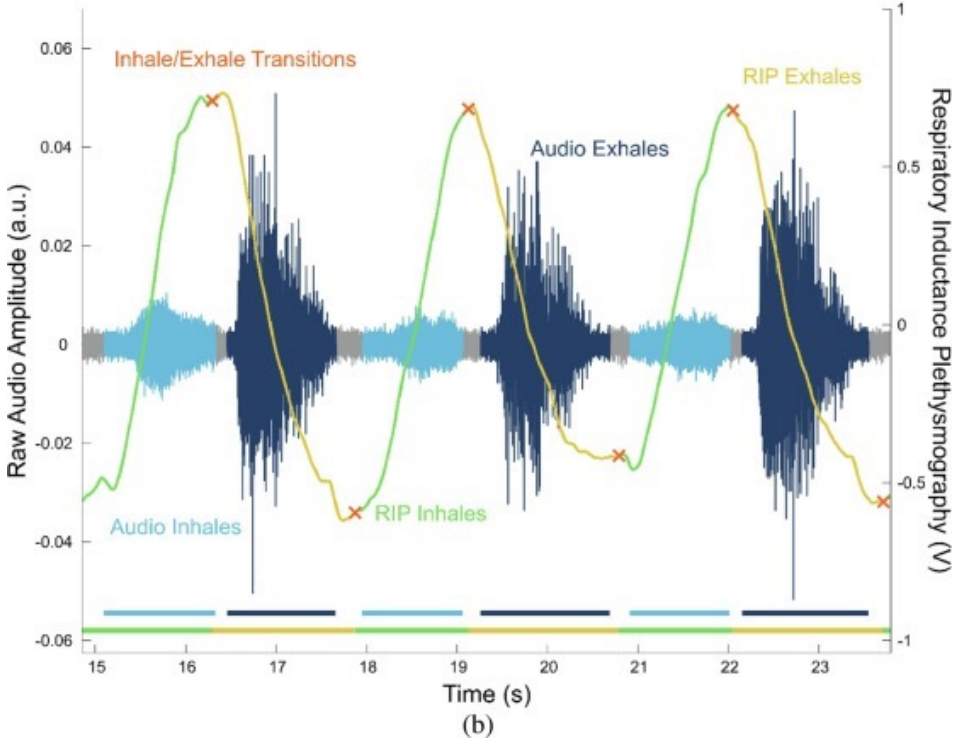
Figure from
https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC

Respiratory Pattern Detection from Audio

- Green/yellow line is ground truth data from chest strap.
- In-out breathing audio patterns can be visibly distinguished on time domain audio plot.



(a)

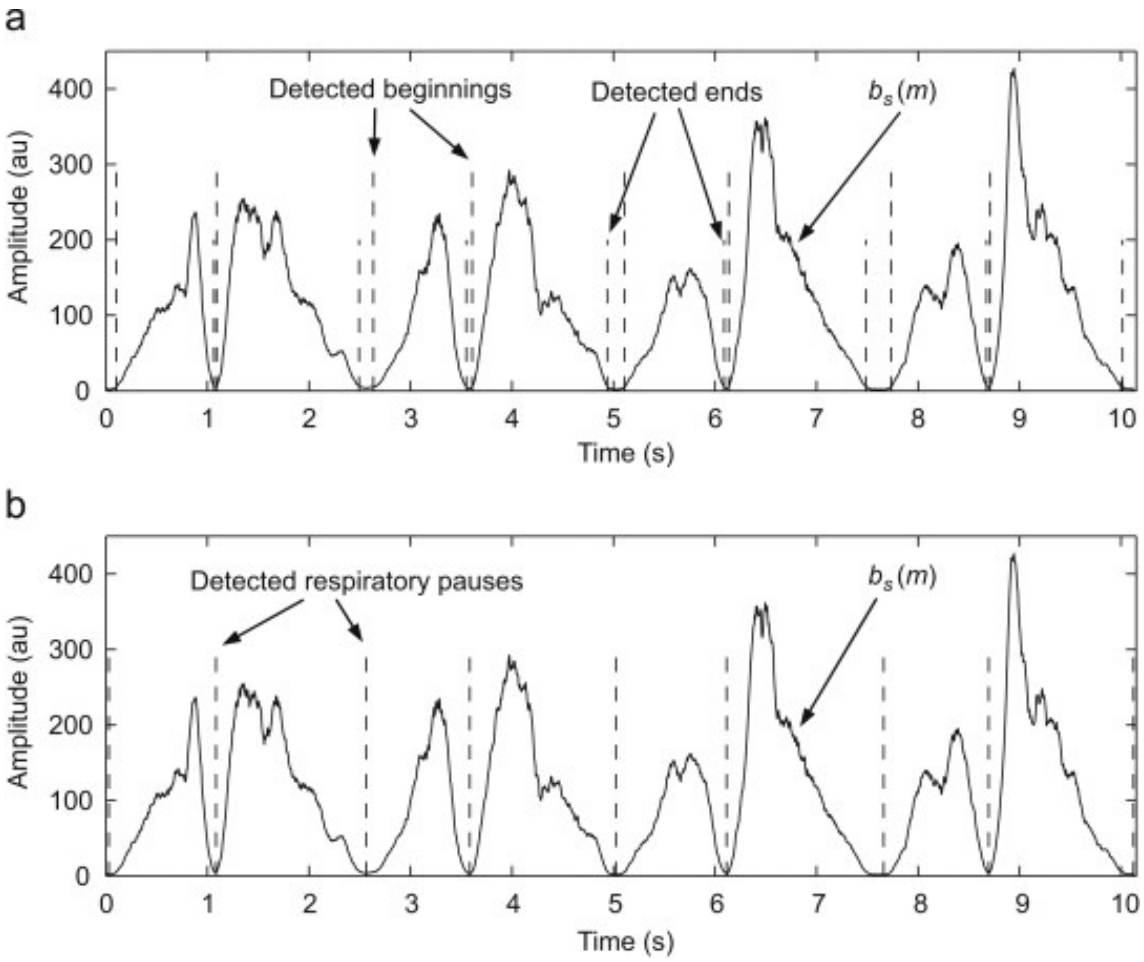


(b)

E Doheny, Ben P.F. O'Callaghan, Vitória S. Fahed, Jérémy Liegey, Cathy Goulding, Silke Ryan, Madeleine M. Lowery, Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones, Biomedical Signal Processing and Control, Volume 80, Part 1, 2023,

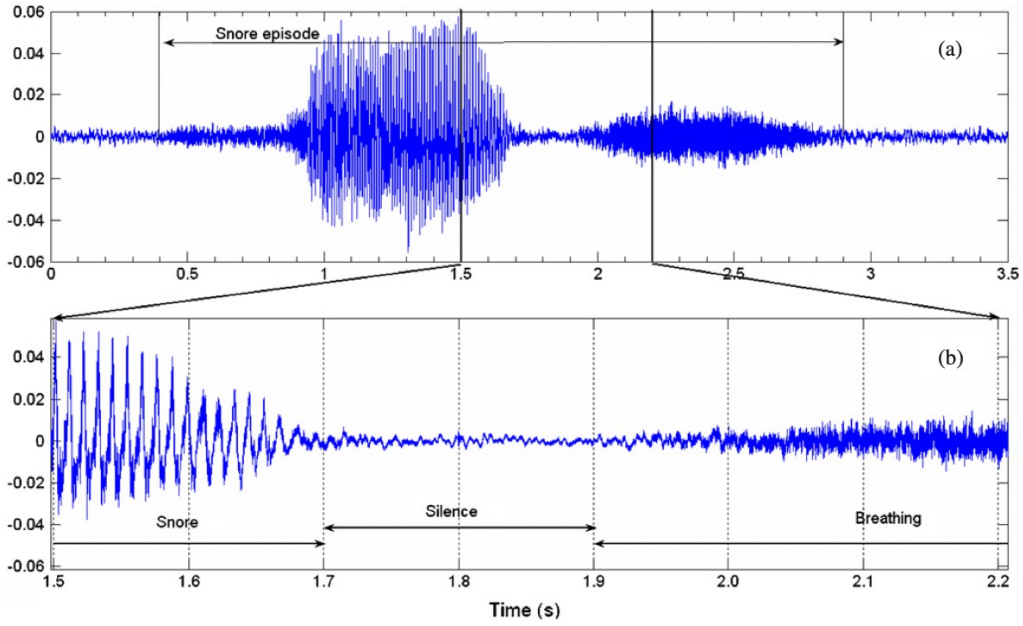
Breathing Pattern Detection

- Signal passed with filters to isolate the right frequencies
  - 8th order Butterworth low-pass filter with cut-off frequency of 1 kHz
- Elimination of signal with coughs, yawning etc: windowing and max and median peak amplitude used.
- Time domain features used to detect pauses in respiration



Snoring-non Snoring

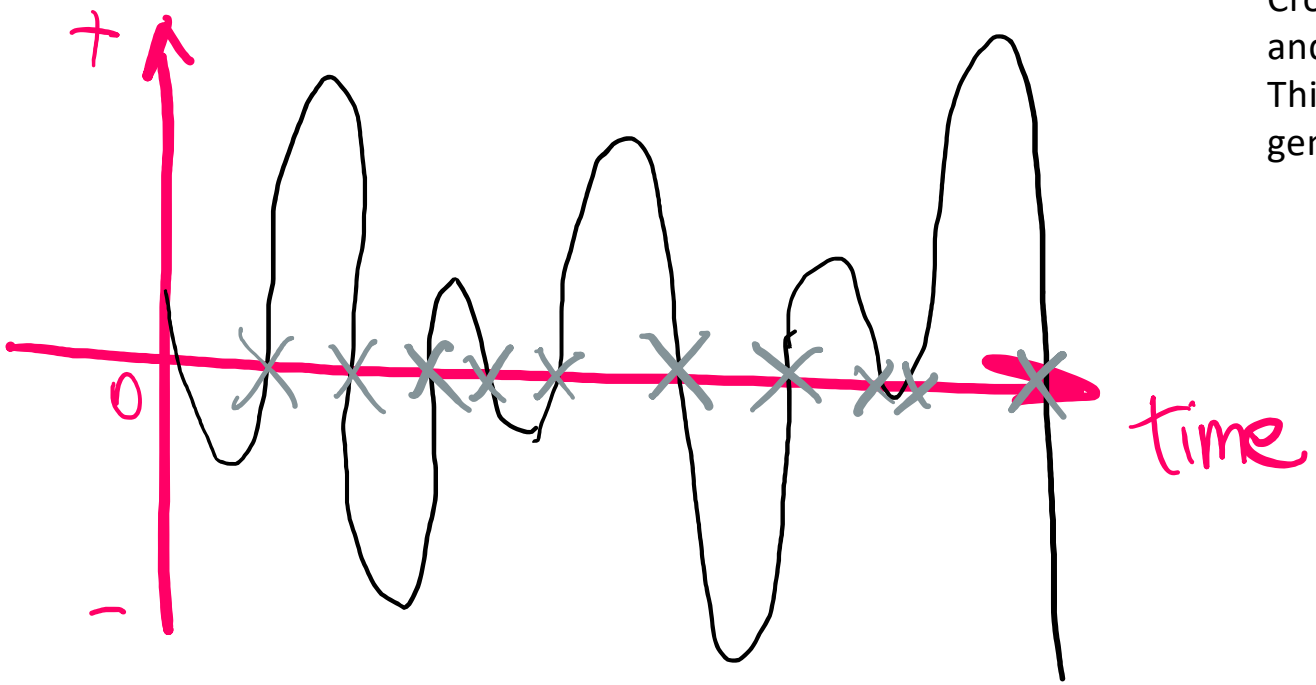
- Identification of voiced snoring, non voice snoring and silence.
- Features
  - number of zero crossings in a given length of time
  - the energy of the signal
  - normalized autocorrelation coefficient at 1 ms delay
  - first predictor coefficient of linear predictive coding (LPC) analysis



UNIVERSITY OF CAMBRIDGE

Karunajeewa AS, Abeyratne UR, Hukins C. Silence-breathing-snore classification from snore-related sounds. Physiol Meas. 2008 Feb;29(2):227-43.

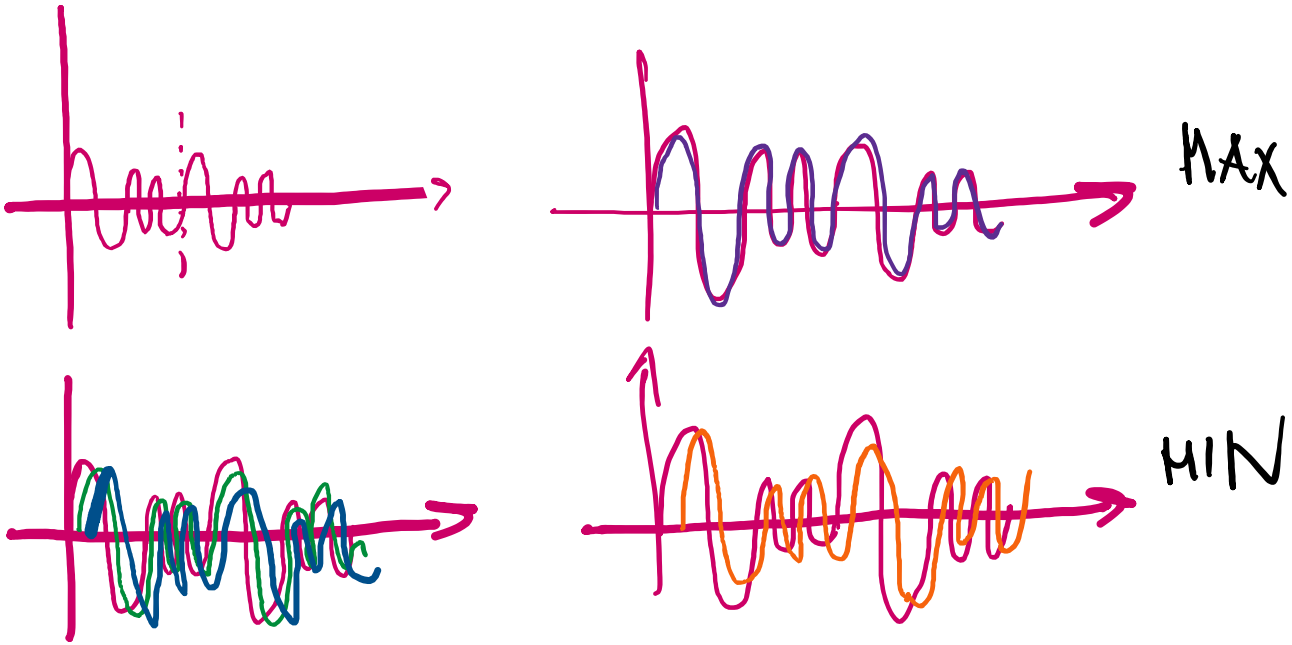
Zero Crossing Rate



Measures the number of times a signal Crosses over from positive to negative and viceversa (per time period).
This is very correlated to frequency generally.

Autocorrelation (also seen in Lecture 3)

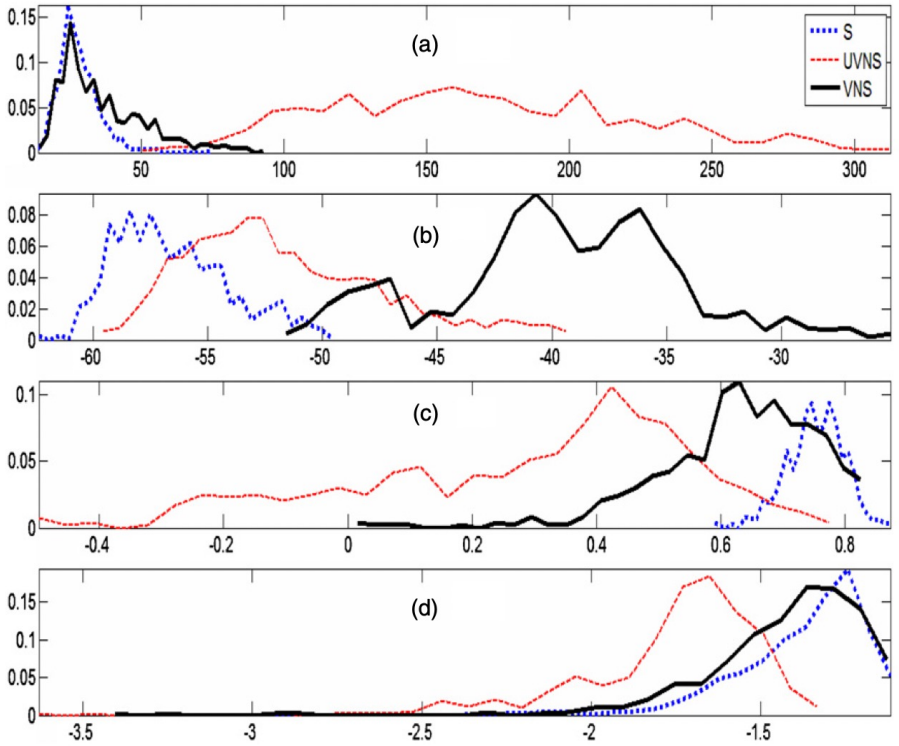
$$R(m) = \sum_{n=0}^{N-1-m} s(n)s(n+m)$$



Performance: Probability distribution of the features for classification

PDF of
-zero crossing
-log of energy
-autocorrelation coefficient
-first linear predictor coefficient

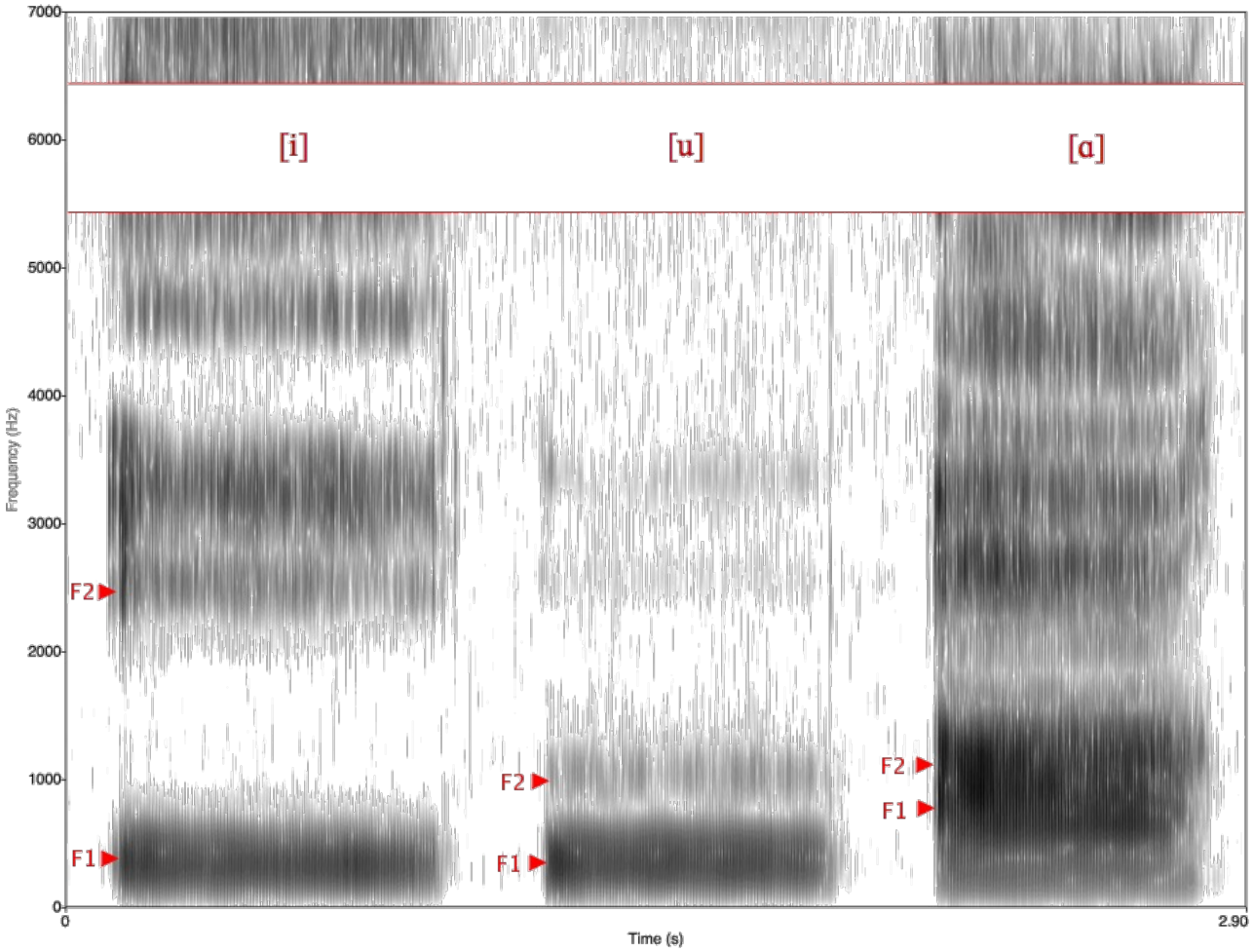
S= silence
UNS = Unvoiced non silence
VNS = voiced non silence



Anatomy of the Voice

- Vocal folds are located within the Larynx
- We produce voiced/non voiced sounds: this depends on the use of the folds
- Vocal folds vibrate at
  - 50:200 Hz for men
  - 150:300 Hz for women
- Additional sounds (eg nasal)

Formants



UNIVERSITY OF CAMBRIDGE

Wikipedia: American English vowels [i, u, ɑ] showing the formants F_1 and F_2

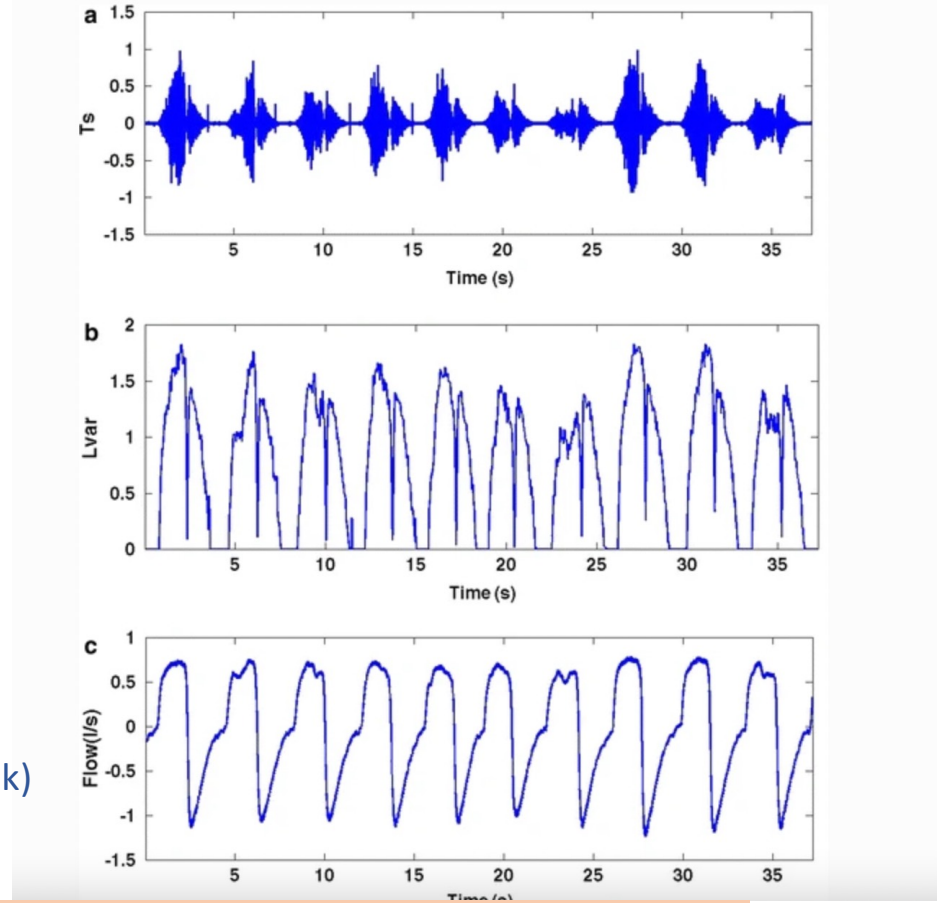
Air Flow and Tracheal Sounds

Log of energy of the tracheal **sound signal** is a good indicator of **air flow**

Signal

Log energy

True flow
(measured with mask)



UNIVERSITY OF CAMBRIDGE

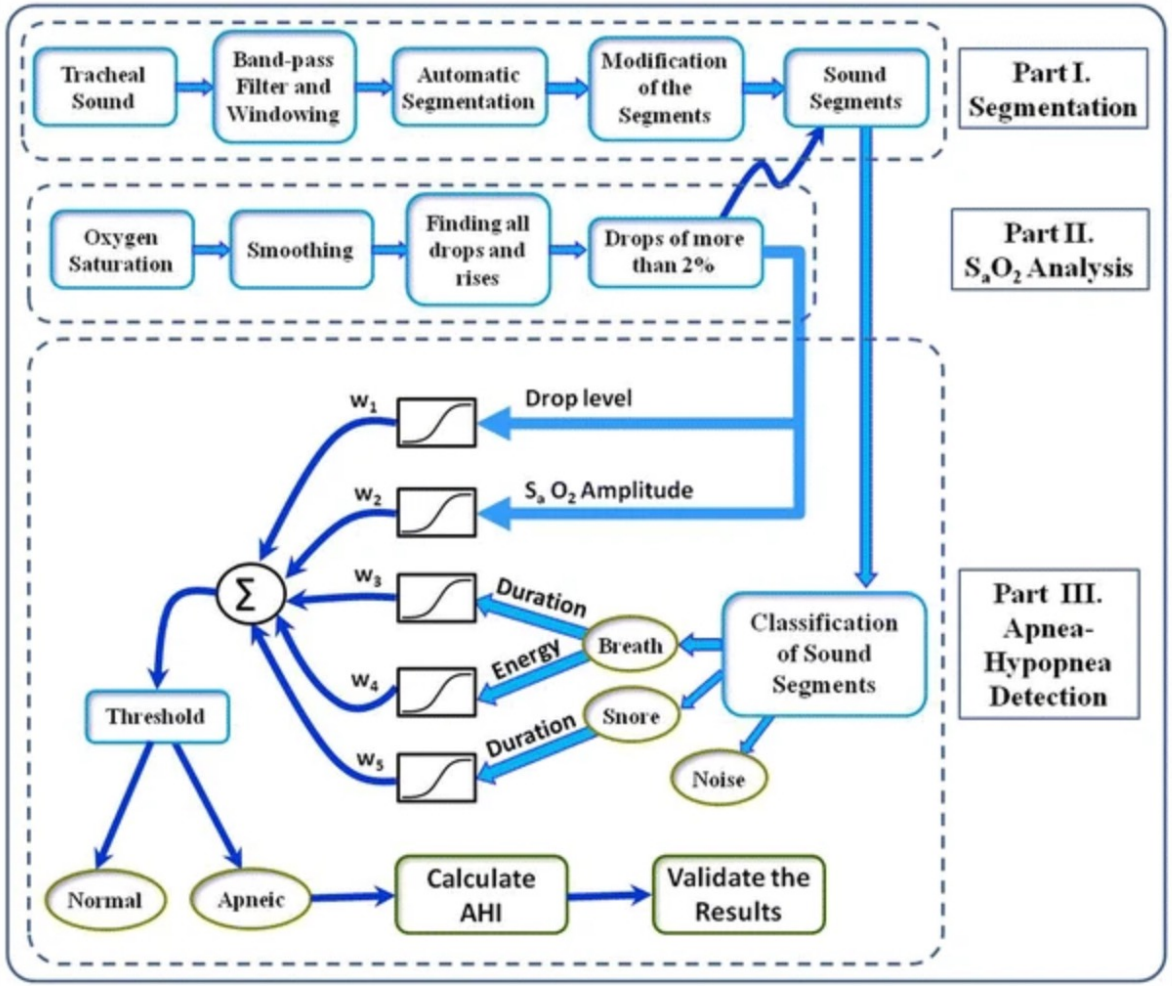
Figures from Yadollahi, A., Giannouli, E. & Moussavi, Z. Sleep apnea monitoring and diagnosis based on pulse oximetery and tracheal sound signals. Med Biol Eng Comput 48. (2010).

Sleep Apnea

- Cessation of airflow to the lungs that lasts at least for 10s and is associated with at least 4% drop in blood's oxygen saturation level (SaO2).

UNIVERSITY OF CAMBRIDGE

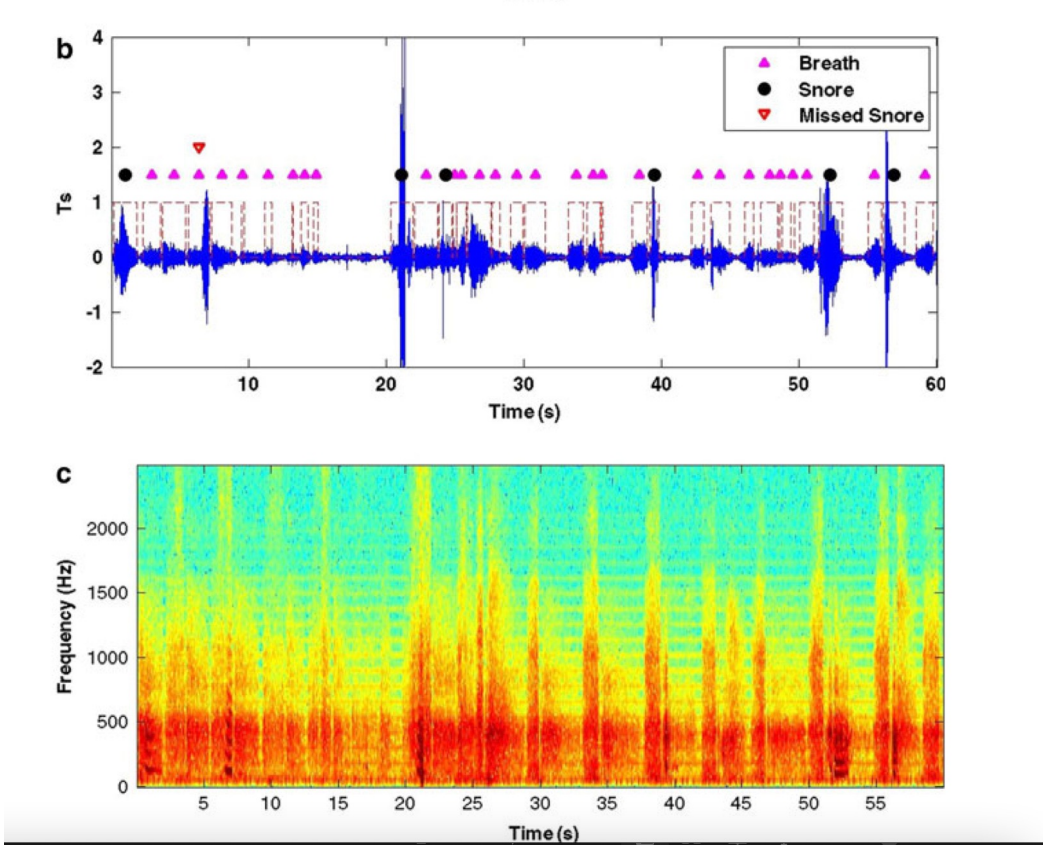
Sleep Apnea Detection with Audio and SaO2 monitoring



UNIVERSITY OF CAMBRIDGE

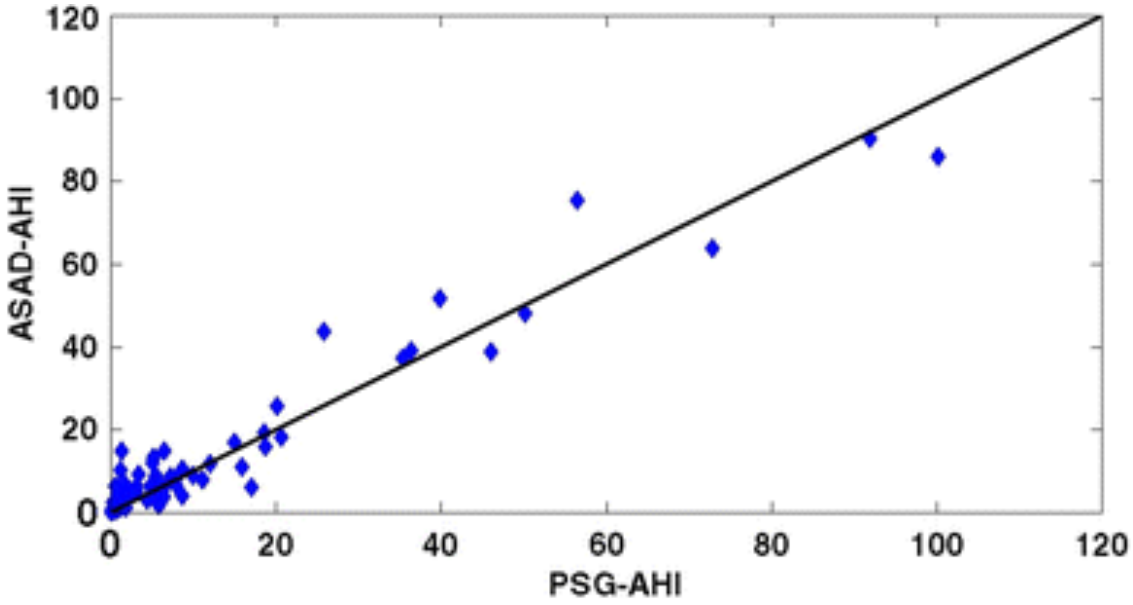
Sleep Apnea Detection

- Classification of the sound in snore, breathing and noise
  - Use of energy and duration (a mixture)
- Spectrogram shows the snoring appearing in deep colours



Results

- AHI: apnea–hypopnea index



UNIVERSITY OF CAMBRIDGE

Questions