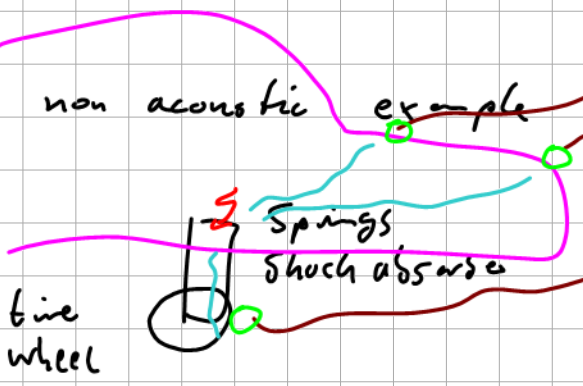
Source Filter Model

$x(n) \rightarrow [\text{LTI}] \rightarrow y(n)$

vocal chords — Lips, tongue, nose — Voice, that you are hearing

Source $x(n)$ * Filter $h(n)$ = $y(n)$

non acoustic example



Microfone

body of the car

transport of vibration through materials

transport over air

$y(n) = h_1(n) * h_2(n) * x(n)$

- Piezo elements

$y(n) = h_1(n) * x(n)$

hopsize in ms = 10

hopsize in samples = $10 \text{ ms} \cdot \frac{1}{1000} \frac{\text{s}}{\text{ms}} \cdot 48000 \frac{\text{samples}}{\text{s}}$

$\sum_{n=0}^{N-1} x(n)$ = np.sum(x(0:N))

$x = [7, 5, 3, 2, 12, \dots, 3]$

$n = \underbrace{0, 1, 2, \dots, N-1}_{N \text{ elements}}$

$\sum_n x(n)$ = np.sum(x)

↳ over all n

$x = [\ \ldots\ ]$

M elements $\quad M = 1000$

$\sum_{n=300}^{500} x(n) \quad \hat{=} \quad np.sum\left(x[300:500]\right)$

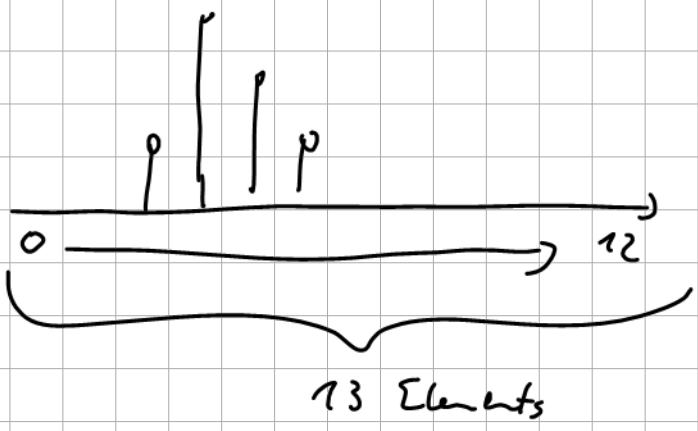
$\varphi = \ldots$

return $\dfrac{\varphi[1]}{\varphi[0]}$

$$COG = \frac{\sum_n n \cdot X(n)}{\sum_n X(n)}$$

↓

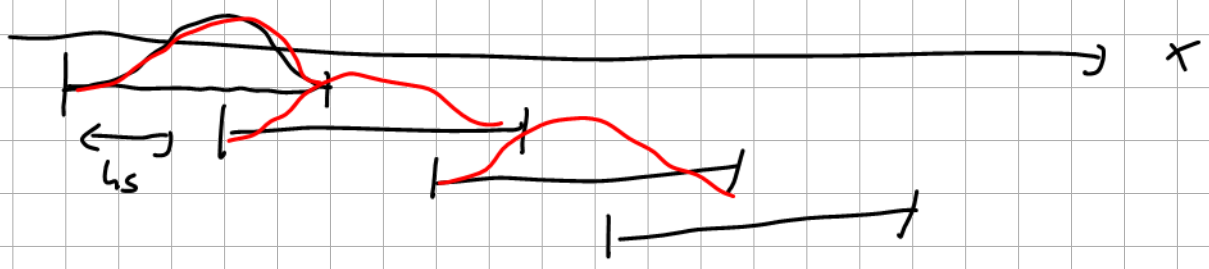
a position of n

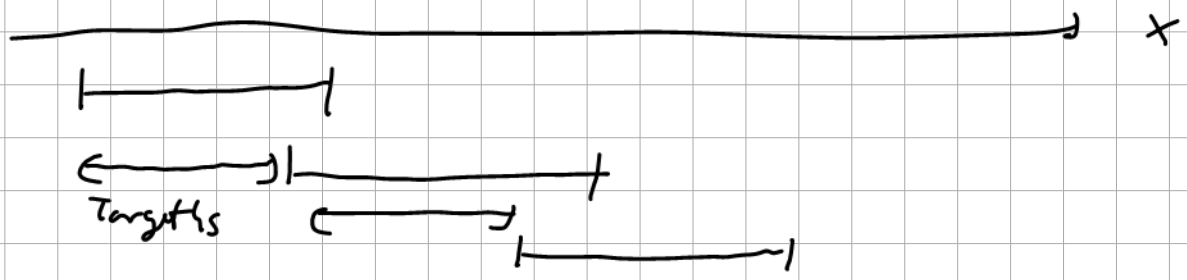


Bandpass

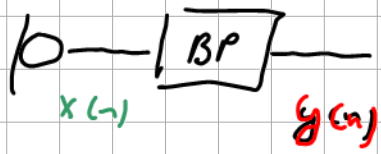
$X(f)$

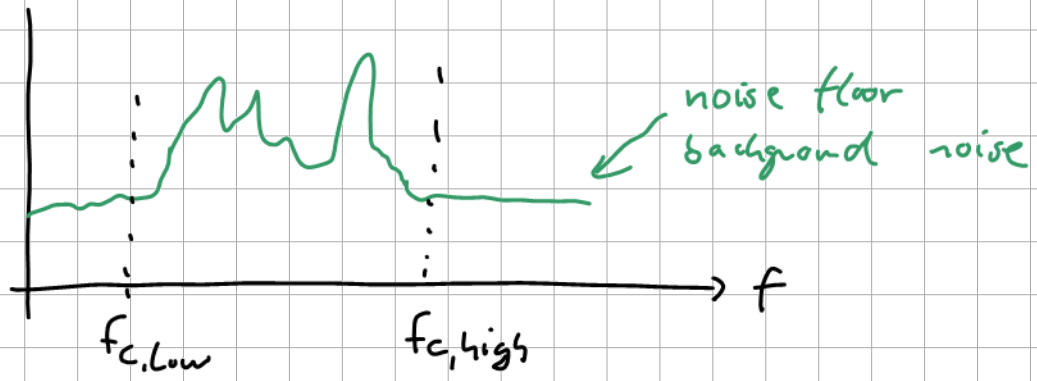
- Timestretch





Why are we using a Bandpass (always at the beginning)





Vocoder – Voice Codec

- understanding human voice
- modifying human voice

human voice
- Source – Pitch
  - Speed
- Filter – Envelope of the local spectra

critical for understanding spoken words

↳ Local envelope of the spectra is sufficient for understanding

If you know what modification does not disturb human classification

→ you can apply these modifications

→ by this increase your amount of training data

⇒ Dataset augmentation