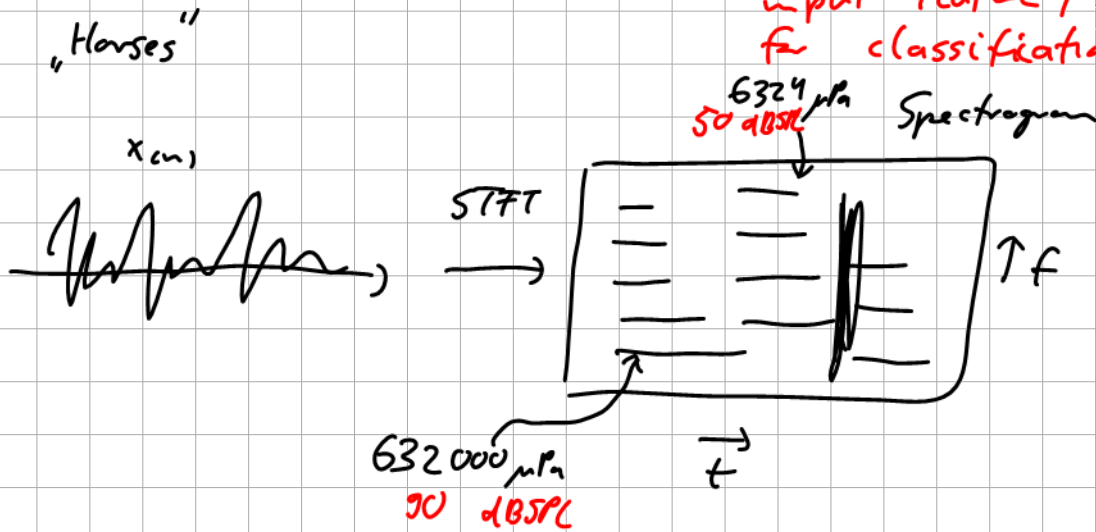


input Feature / Signal for classification

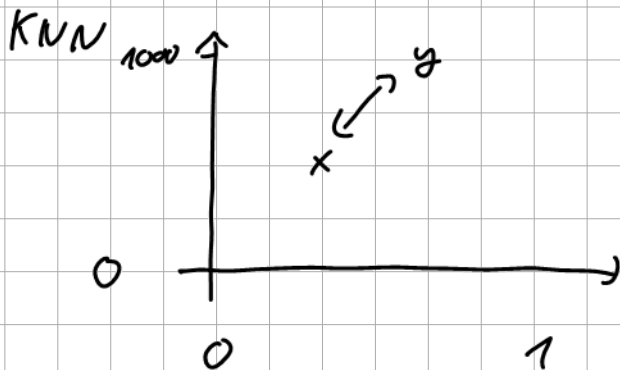


Human Voice $L = 50 \text{ dB SPL} \dots 90 \text{ dB SPL} = 20 \log_{10} \frac{p}{20 \mu Pa}$

Pressure : $p = 20 \mu Pa \cdot 10^{\frac{L}{20}}$

$= 20 \mu Pa \cdot 10^{\frac{50}{20}} = 6324 \mu Pa$

$= 20 \mu Pa \cdot 10^{\frac{90}{20}} = 632000 \mu Pa \cdot 100$

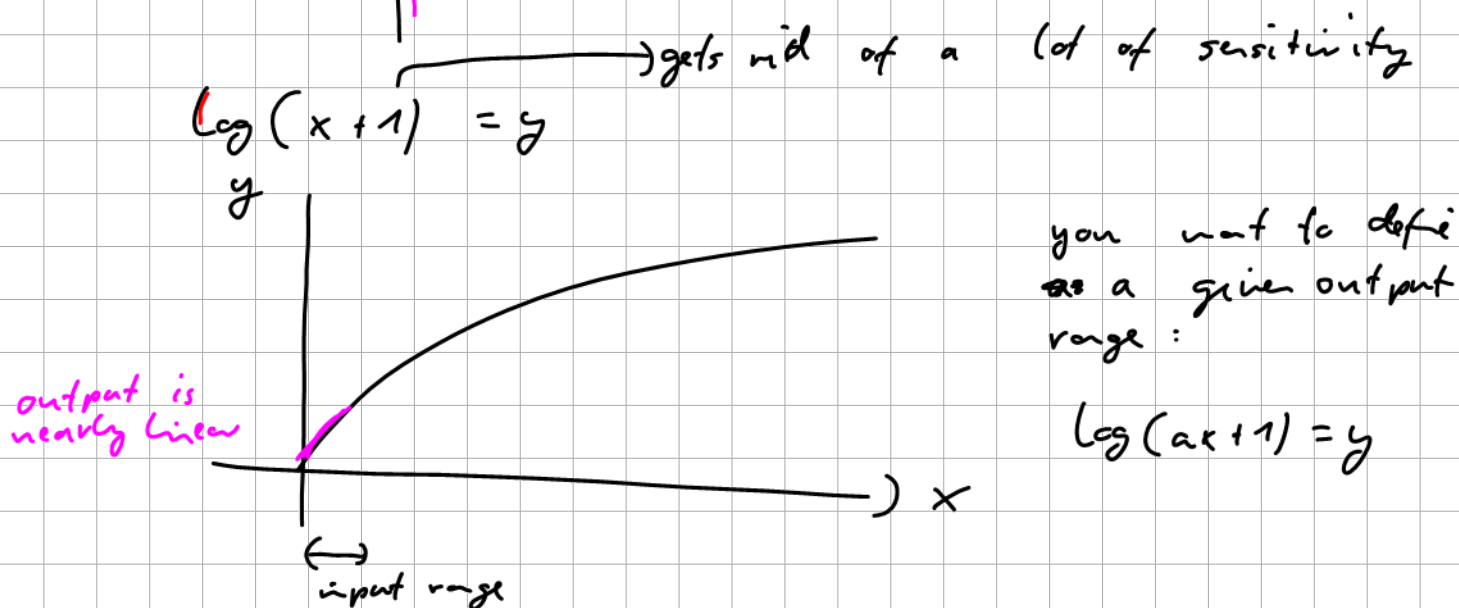
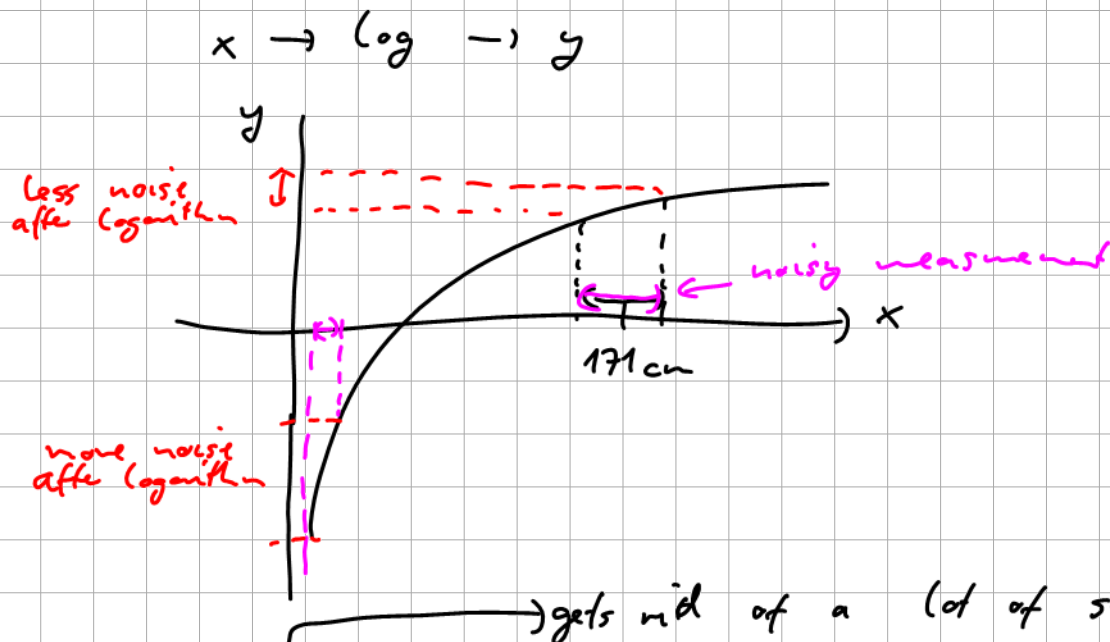


$d(\vec{x}, \vec{y}) = \sqrt{\sum_n (x(n) - y(n))^2}$

Logarithm compresses the Feature for classification algorithms.

:-)

Logarithm is sensitive to noise :-C



assuming

$$0 \leq x \leq 1$$

$$0 \leq y = \log(ax+1) \leq 100$$

$$a = ?$$

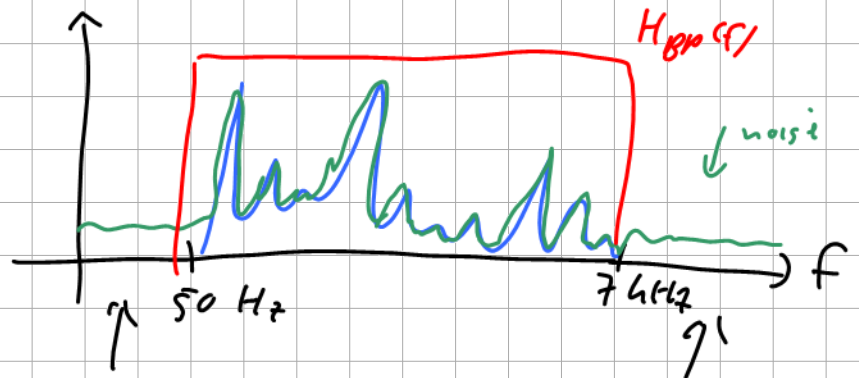
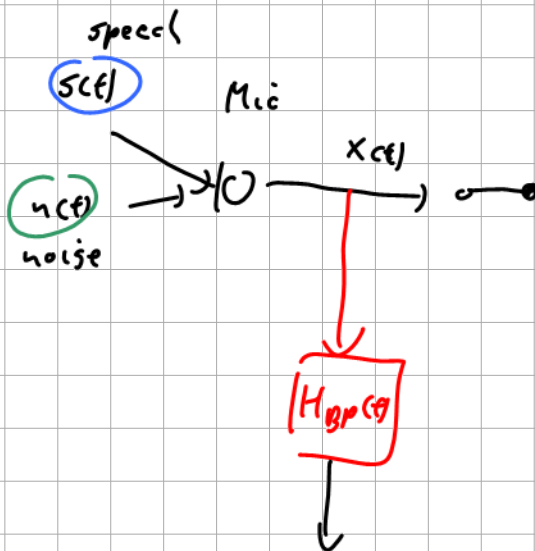
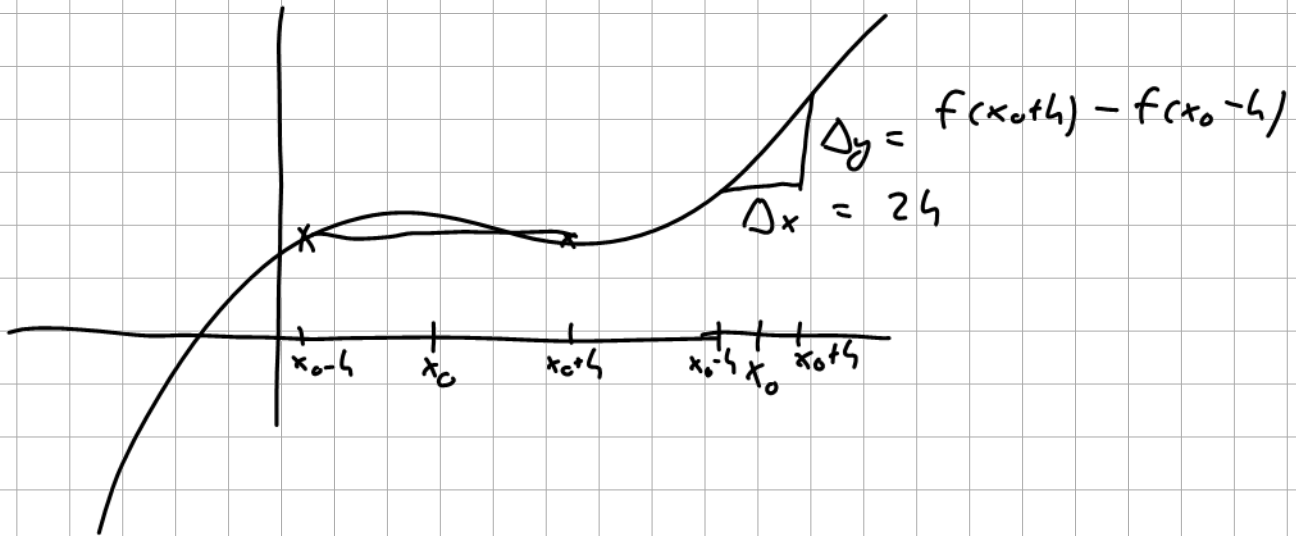
$$x=0 : y = \log(a \cdot 0 + 1) = \log(1) = 0$$

$$x=1:$$

$$y = \log(a \cdot 1 + 1) = 100$$

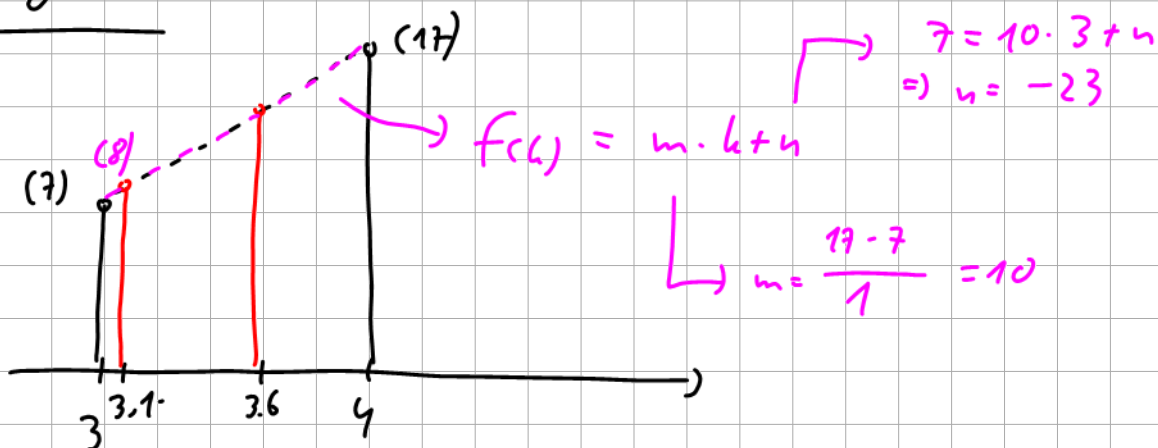
$$a + 1 = e^{100}$$

$$a = e^{100} - 1$$



is shipped
by Bandpass

C2 211

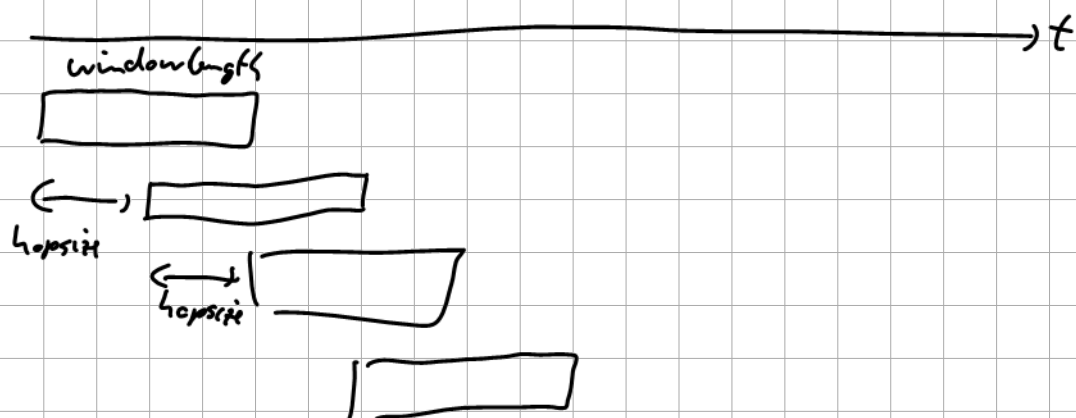


$$f(k) = 10 \cdot k - 23$$

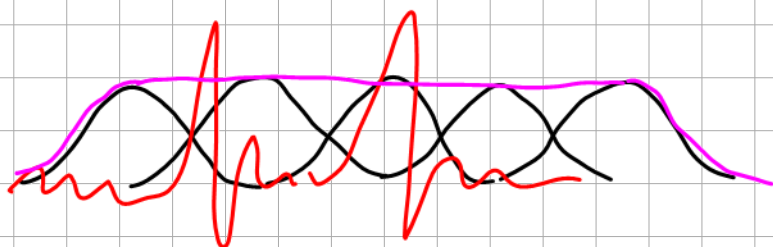
$$f(3.1) = 31 - 23 = 8$$

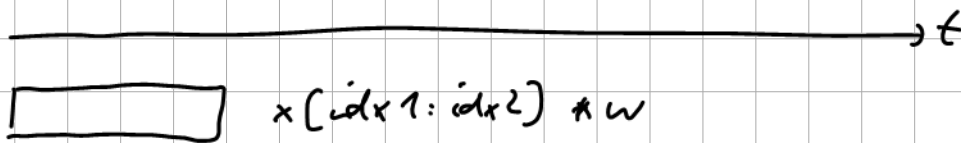
STFT

Constant overlap add property

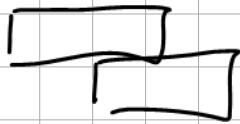
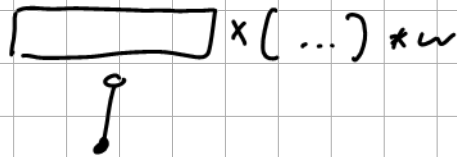


$$\text{window size} = \text{hopsize}$$





DFT



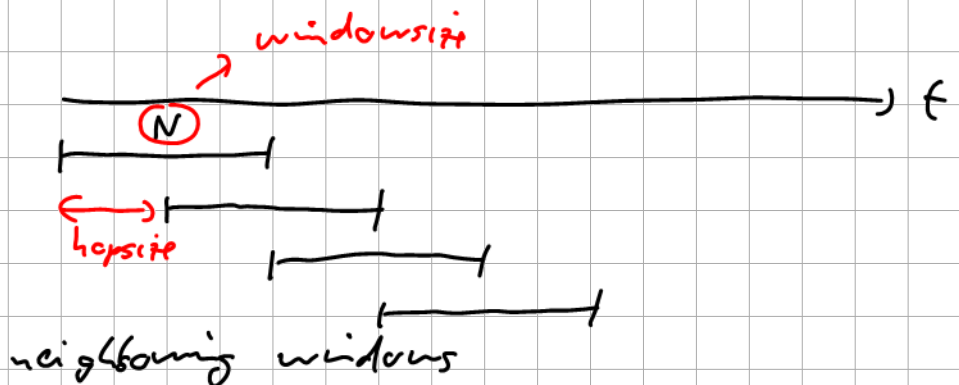
C2 J8 T1

$$\Delta f < 10 \text{ Hz}$$

Audio length 3s

$$r = 16000$$

50 spectra per second
overlap of 50%



$$\text{hop size} = \frac{N}{2}$$

$$\text{hop size} = \frac{16000}{50} = 320 \Rightarrow N = \text{window size} = 640$$

$$\Delta f = 10 \text{ Hz} = \frac{1}{K \cdot T} = \frac{r}{K} \quad (T = \frac{1}{r})$$

$$\rightarrow K = \frac{r}{10 \text{ Hz}} = \frac{16000 \text{ Hz}}{10 \text{ Hz}} = 1600$$

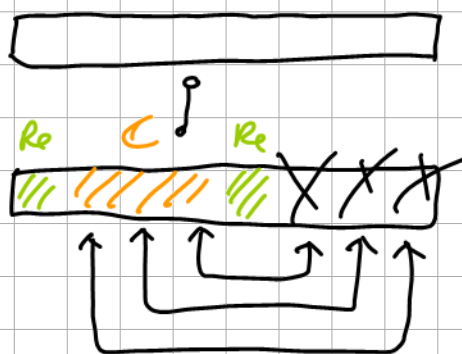
zero padding?

$$K > N \Rightarrow 1600 > 640 \quad \checkmark$$

\rightarrow zero padding is used

3s Audio 50 Spectra per second \Rightarrow 150 spectra

each spectrum



$K=8 \Rightarrow$ 5 elements remain

$K=1600 \Rightarrow$ remaining elements $\frac{K}{2} + 1 \Rightarrow 801$ elements remain

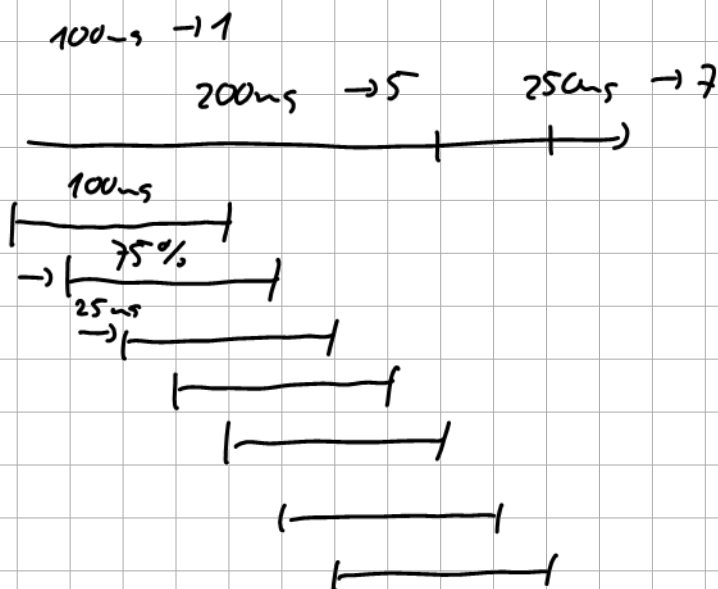
2 real 4 Byte
799 complex 8 Byte

$$150 \cdot [2 \cdot 4 + 799 \cdot 8] = 961200 \text{ Byte}$$

C2 28 74

5s
 $N=100 \text{ ms}$
overlap 75%

$$h_{\text{overlap}} = N \cdot (1 - 0.75) = 25 \text{ ms}$$



$$197 = 1 + \frac{55 - 0.15}{25 \text{ ms}}$$

C2 J8 TS

55 audio

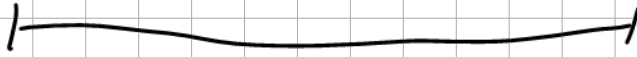
$$0 \leq \zeta \leq 24 \text{ Sa/s}$$

time resolution of human ear: 1ms
hopsize

frequency resolution: $\frac{1}{27} \text{ Sa/s}$

$$0 \leq \zeta \leq 24 \text{ Sa/s}$$

$$\Rightarrow \text{frequencies to analyse} = 27 \cdot 24 + 1 = 649$$

$$0, \frac{1}{27}, \frac{2}{27}, \frac{3}{27}, \dots, 24$$


how many time steps to analyse: 5000

\Rightarrow Number of values in spectrogram: $\begin{matrix} \text{columns} & \text{rows} \\ 5000 & \cdot 649 \\ = 3245000 \end{matrix}$

