# 8.3 Basic Parameters for Audio Analysis

**Physical audio signal: simple**

- one-dimensional
- amplitude = loudness
- frequency = pitch

**Psycho-acoustic features: complex**

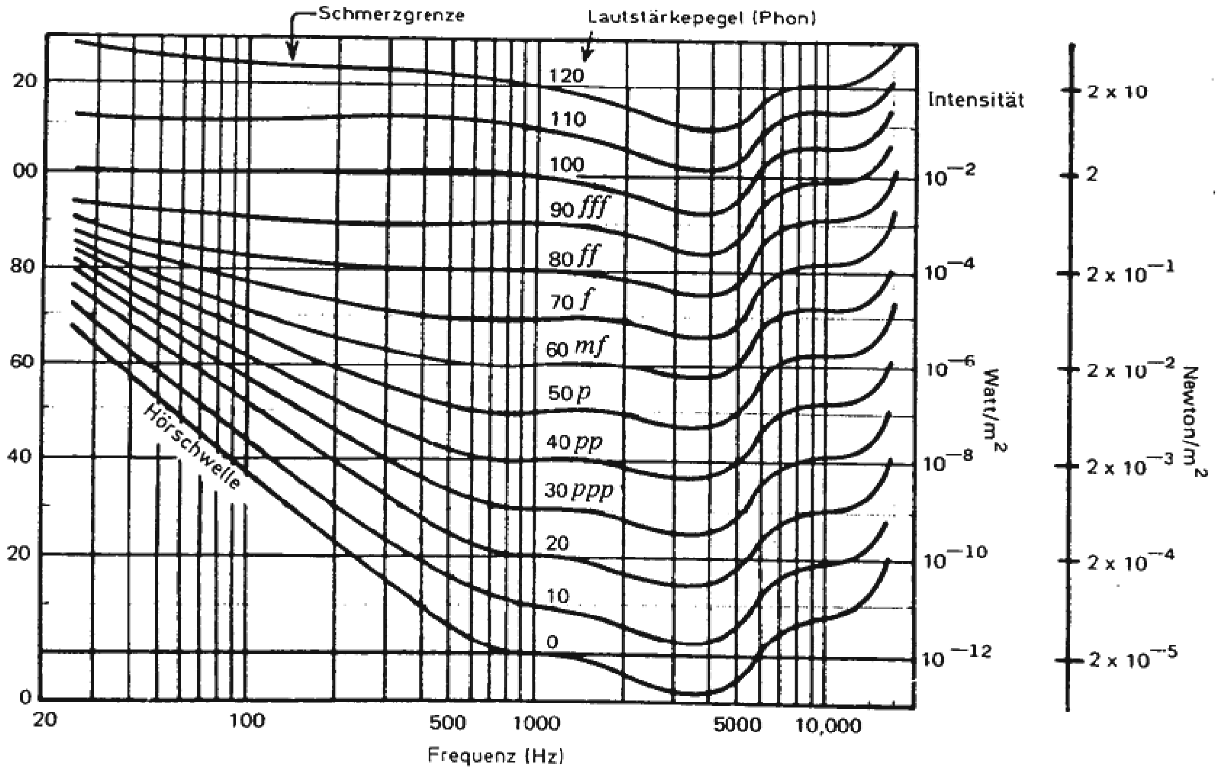
- A real-life tone arises from a complex superposition of various frequencies.
- For human audible perception, the emerging and fading away of a tone are very important (e.g., they distinguish the tone of a piano from the tone of a guitar).

# Perception of Loudness

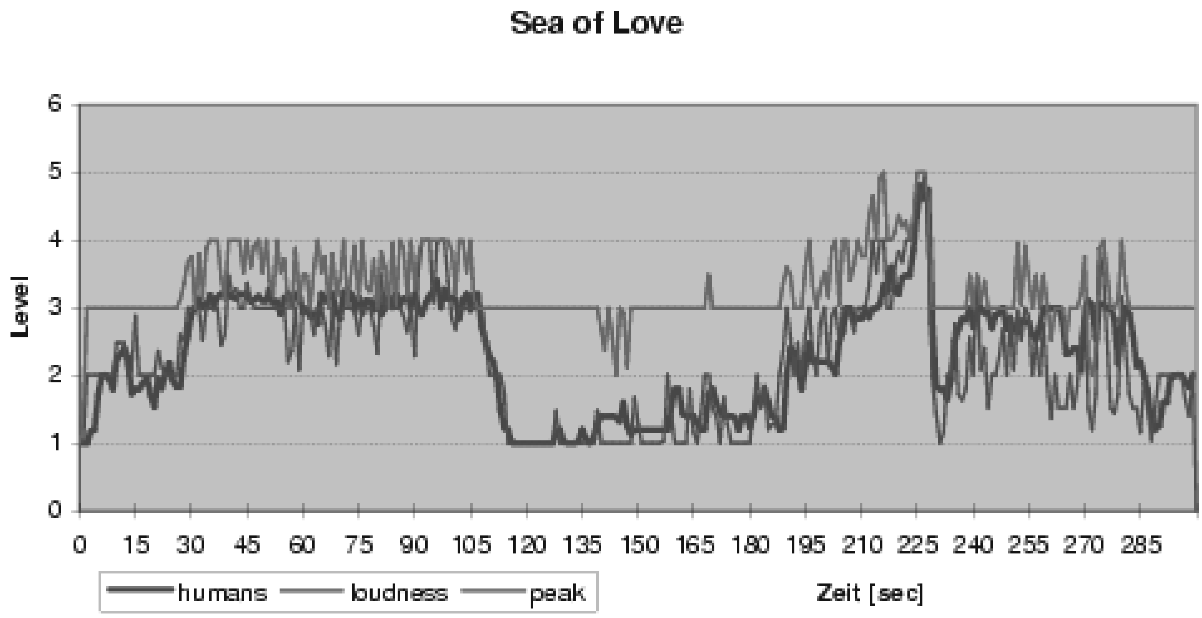
The physical measure is called **acoustic pressure**, the unit is decibel [dB].

The human audible perception is called **loudness**, the unit is phon.

We can empirically derive a set of curves that depicts the perceived loudness as a function of acoustic pressure and frequency. They are called **isophones**.

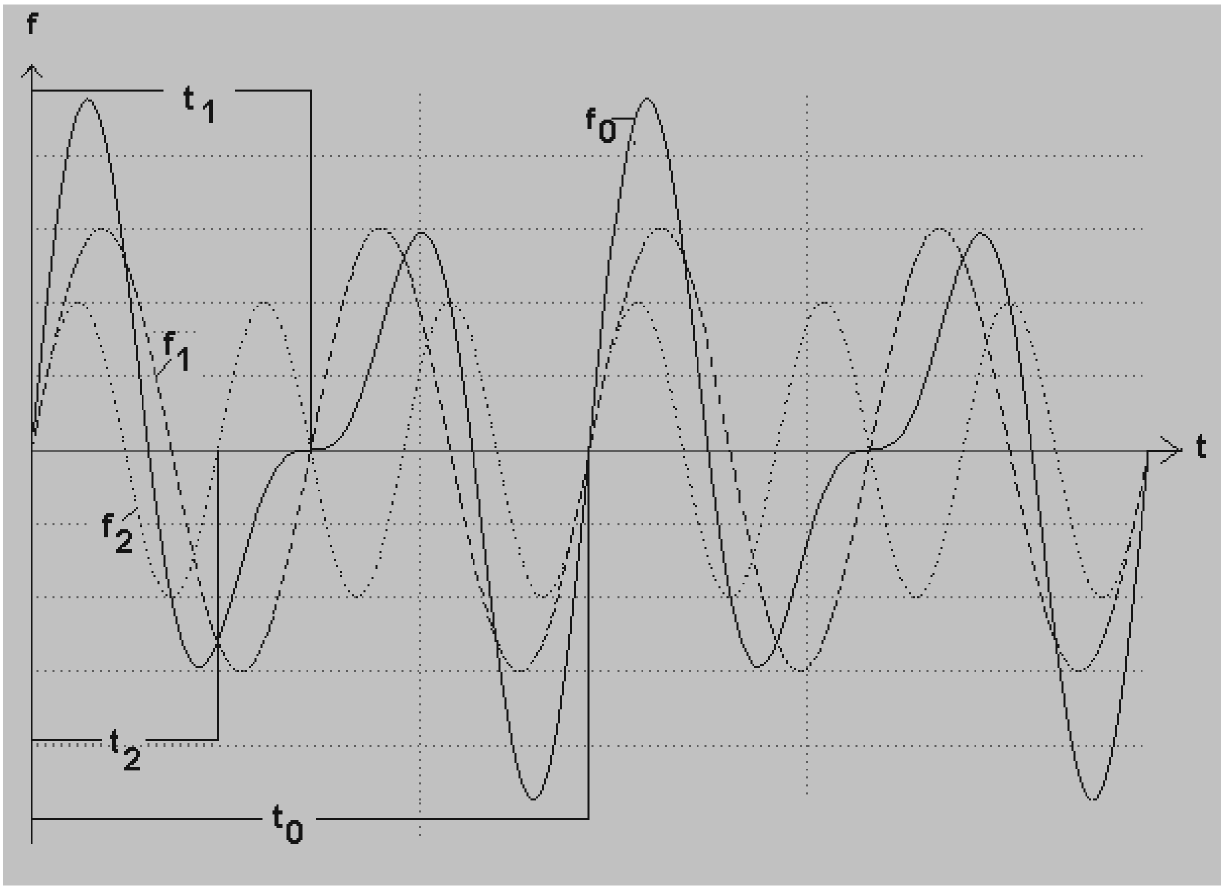


# Experimental Results




| | |
|---|---|
| red curve: | acoustic pressure |
| black curve: | loudness as perceived by test subjects |
| blue curve: | computationally predicted perceived loudness |

# Fundamental Frequencies in Harmonic Sounds




The fundamental frequency of the composite tone $f_0$ corresponds to the minimum common multiple of the two composing frequencies $f_1$ and $f_2$.

# Frequency Transformations

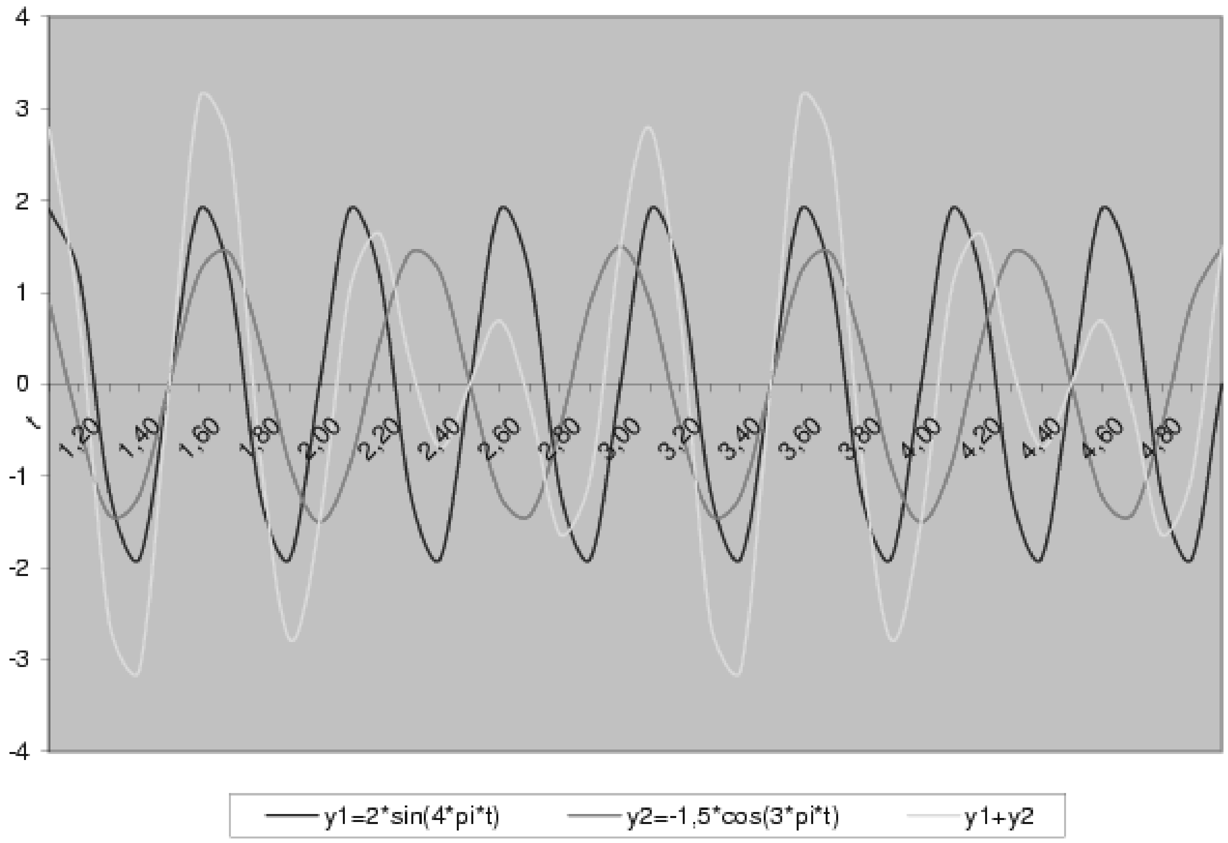
**J.B.J. Fourier** (1768-1830): Each periodic oscillation can be written as the sum of harmonic frequencies:

$$s(t) = \frac{B_0}{2} + \sum_{n=1}^{\infty} [A_n \sin(2\pi nft) + B_n \cos(2\pi nft)]$$

*f*: basic frequency

$A_n, B_n$: amplitudes

$\sin(2\pi nft)$ = multiples of the basic frequency



# Frequency Transformation of an Audio Signal

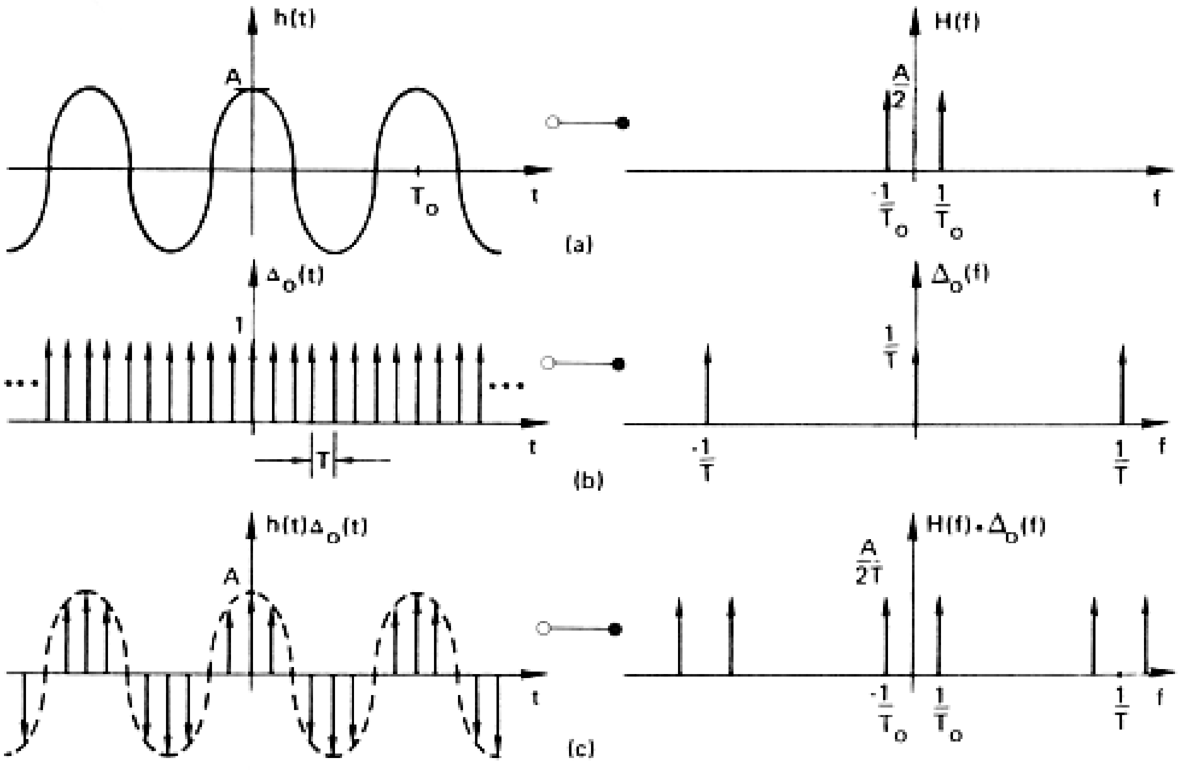
Here: discrete Fourier transform (DFT) with N sampling points

$$S(f) = \sum_{n=0}^{N-1} s(n) e^{-if\frac{2\pi}{N}n}, \; f = 0,1,\ldots, N-1$$

| | | |
|---|---|---|
| | s(t) | continuous original signal |
| step 1 | | sampling at rate $f_s = \frac{1}{T}$ |
| | s[t] | discrete original signal |
| step 2 | | temporal restriction to a window w(t) |
| | s[t] | discrete original signal containing N sampling values [0, *NT*] |
| step 3 | | N-point DFT |
| | S(f) | continuous Fourier transform |
| step 4 | | sampling at rate *N* per *T* |
| | S[f] | discrete Fourier transform |

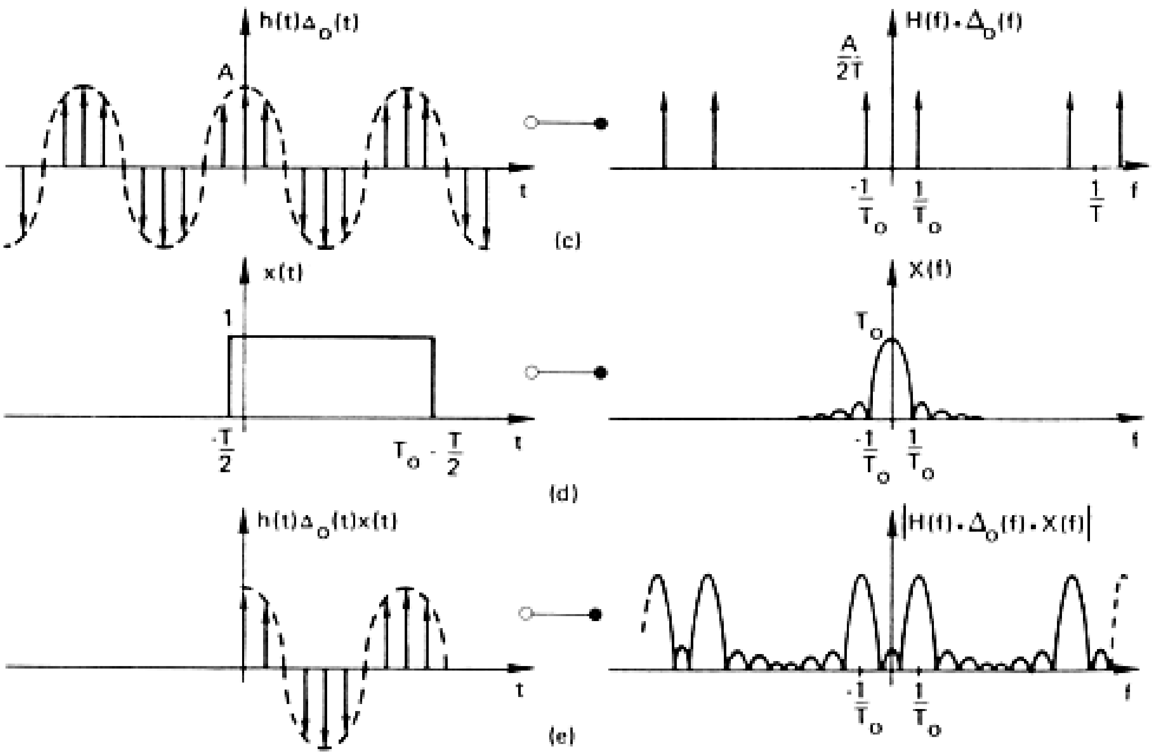
Steps 3 and 4 can be sped up considerably by means of the fast Fourier transform (FFT).

# Step 1: Sampling in the Time Domain




**Time domain** **Frequency domain**

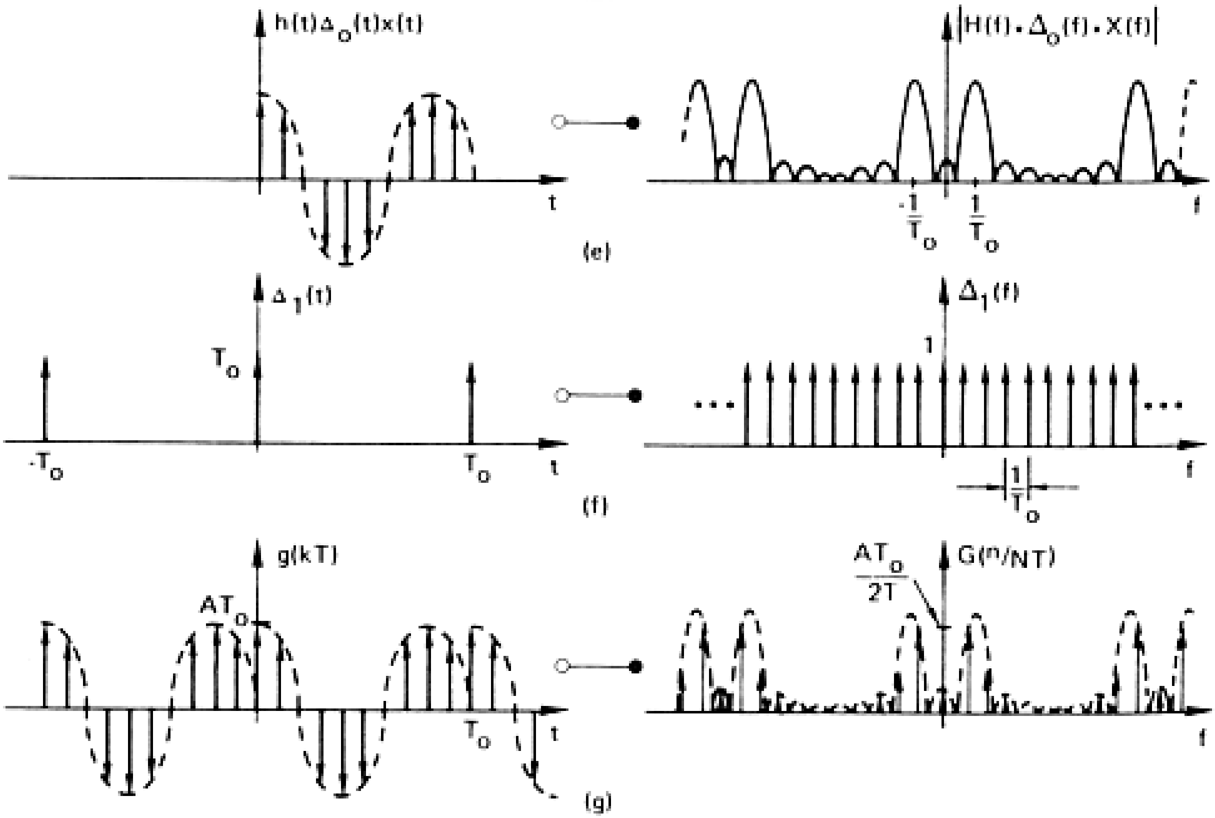
# Step 2: Time Restriction to [0, *NT*]



**Time domain** **Frequency domain**

# Step 3: Sampling in the Frequency Domain

**Goal:** Discretization of the data also in the frequency domain (for representation in the computer)




**Time domain** **Frequency domain**

## Reference:

E.Oran Brigham: Fast Fourier Transform and Its Applications, Prentice Hall, 1997

# Signal Analysis with the DFT

## Assumption

A natural audio signal of sampling length *M* is given, e.g., *M* = 5 min of music.

## Goal

Extraction of features, e.g., musical tones (pitch, loudness, onset, etc.)

## Method

Definition of a window of size *N* which is moved over the audio signal. It represents a window of analysis. The DFT is computed on this window. Only with a **windowed** DFT, we can analyze the behavior of the signal over time.

Example: We can assume that musical tones are stationary for at least 10 ms. We thus choose *N* = 10 ms.

When moving the window, we allow redundancy in order to also analyze the transitions between tones. Here, we chose an overlap of 2 ms. This results in

$$\frac{5x60x100}{8} = \frac{30.000}{8} = 3.750$$

frames.

# Signal Analysis – Properties (1)

It is now possible to compute semantic features for the sample frames.

## 1. Energy

$$E_s(m) = \sum_{n=m-N+1}^{m} s^2(n)$$

$m$ = ending time of the frame

$E_s$ is a measure for the **acoustic energy** of the signal in the frame. It corresponds to the square of the area under the curve in the time domain.

The energy might as well be computed for the frequency-transformed signal. It then denotes a measure for its **spectral energy spread.**

# Signal Analysis – Properties (2)

## 2. Zero-crossings

$$sign(s(n)) = \begin{cases} 1: & s(n) \geq 0 \\ -1: & s(n) \prec 0 \end{cases}$$

$$Z_s(m) = \frac{1}{N} \sum_{n=m-N+1}^{m} \frac{|sign(s(n)) - sign(s(n+1))|}{2}$$

- Counts the number of zero-crossings (i.e., sign changes) of the signal.
- High frequencies lead to a high $Z_s$, while low frequencies lead to a low $Z_s$
- This is closely related to the basic frequencies.

Many other parameters are also used in audio signal analysis.