

## 8.4 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.

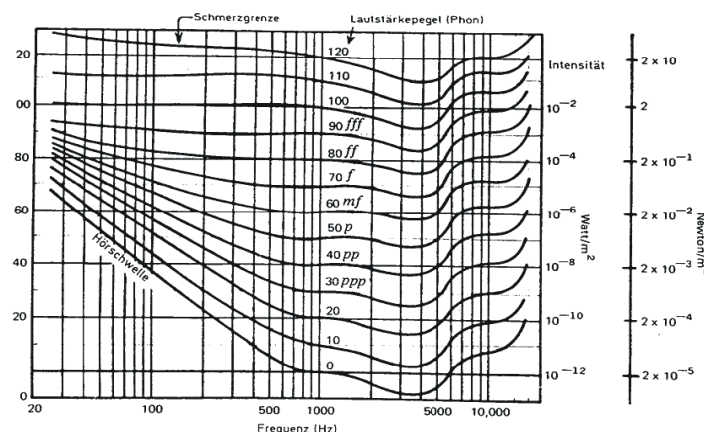
For example, both features are different for a violin and a piano.

## Perception of Loudness

The physical measure is called **acoustic pressure**, the unit is decibel [dB-SPL, Sound Pressure Level].

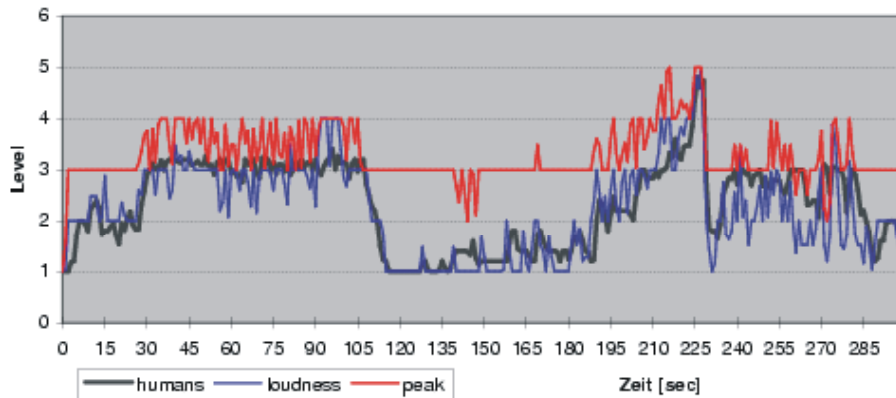
The human audible perception is called **loudness**, the unit is phon.

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

Sea of Love

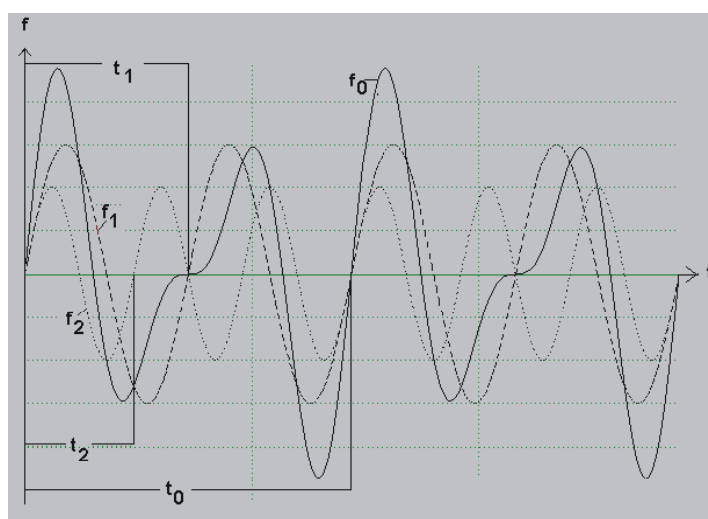


red curve: acoustic pressure

black curve: loudness as perceived by test subjects

blue curve: computationally predicted perceived loudness

# Fundamental Frequencies in Harmonic Sounds



The period of the composite tone  $f_0$  corresponds to the least common multiple of the periods of the two composing frequencies  $f_1$  and  $f_2$ .

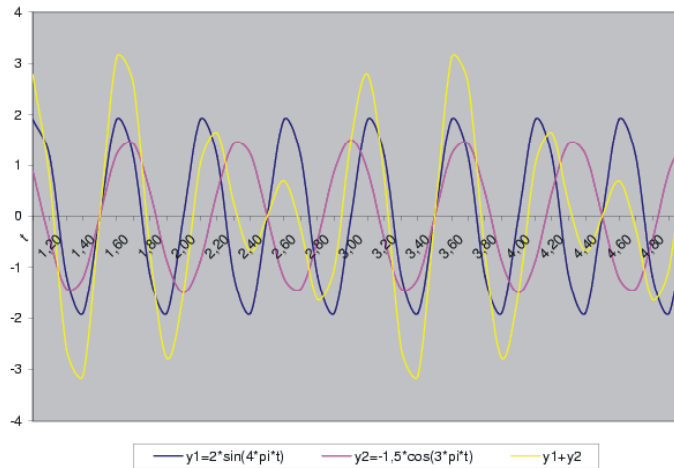
# Frequency Transformations

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

$$s(t) = \frac{B_0}{2} + \sum_{f=1}^{N-1} [A_f \sin(\frac{2\pi ft}{N}) + B_f \cos(\frac{2\pi ft}{N})]$$

$f$ : frequency

$A_f, B_f$ : amplitudes

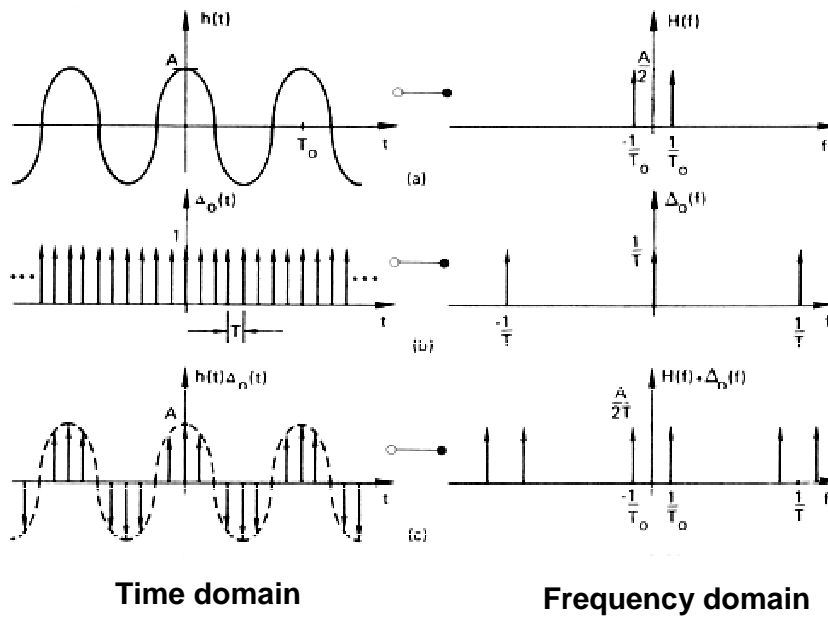


# Frequency Transformation of an Audio Signal

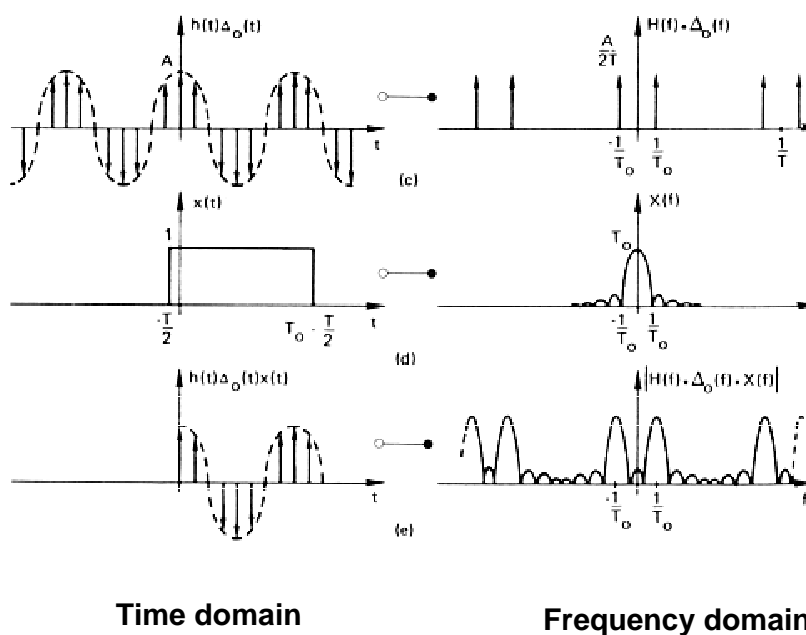
	$s(t)$	continuous original signal
step 1		sampling at rate $f_s = \frac{1}{T}$
	$s[t]$	discrete original signal
step 2		temporal restriction to a window $w(t)$
	$s[t]$	discrete original signal containing $N$ sampling values $[0, NT]$
step 3		$N$ -point DFT
	$S(f)$	continuous Fourier transform
step 4		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). The complexity of FFT is  $O(n \log n)$  compared to  $O(n^2)$ .

# Step 1: Sampling in the Time Domain

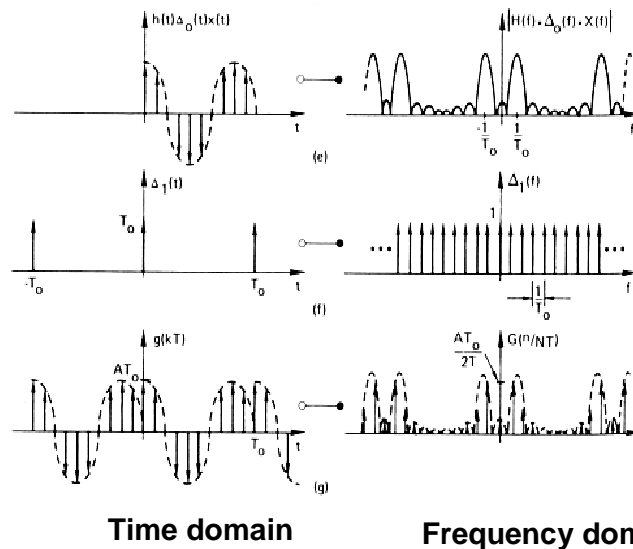


# Step 2: Time Restriction to $[0, NT]$



## Step 3: Sampling in the Frequency Domain

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



**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

$$\frac{5 \times 60 \times 100}{8} = \frac{30.000}{8} = 3.750$$

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**. Computing the energy in the frequency space makes sense if one is interested in knowing frequency ranges in which the energy occurs.

# Signal Analysis – Properties (2)

## 2. Zero-crossings

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

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

- 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.