Inhaltsverzeichnis

Kapitel 1

Mathematical and Physical Background of Sound Production and Propagation

1.1 Notation and some general useful formulae

For differentiations I write:

$$\partial_t = \frac{\partial}{\partial t}; \quad \partial_x = \frac{\partial}{\partial x}; \quad \partial_i = \frac{\partial}{\partial x_i}.$$

$f'(\texttt{expr})$ denotes the differentiation with respect to `expr`.
$(e^{ik\,x - i\omega\,t})' = e^{ik\,x - i\omega\,t}$,
whereas
$\partial_t(e^{ik\,x - i\omega\,t}) = -i\omega\,(e^{ik\,x - i\omega\,t})$

We use i as imaginary unit: $i^2 = -1$, and * for complex conjugation: If a and b are real numbers:

$$(a + i\,b)^* = \overline{a + i\,b} = a - i\,b$$

Frequently used **special functions**:

The θ- function.

In order to express case distinctions in formulae, it is convenient to introduce the θ-function:

$$\theta(x-a) = \begin{cases} 1 & \text{for } for \quad x > a \\ 0 & \text{for} \quad x \leq a \end{cases} \tag{1.1}$$

Exponential Function

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \ldots$$

with the convenient properties:

$$e^a \cdot e^b = e^{a+b}; \quad (e^x)' = e^x; \quad \partial_x e^{f(x)} = e^{f(x)} \partial_x f(x)$$

$$\int dx\, e^x = e^x; \quad \int dx\, e^{a\,x} = \frac{1}{a} e^{a\,x}$$

$$e^{ix} = \cos(x) + i\, \sin(x) \tag{1.2}$$

respectively:

$$\sin(x) = (e^{ix} - e^{-ix})/(2\,i); \quad \cos(x) = (e^{ix} + e^{-ix})/2; \tag{1.3}$$

Since $(e^{ix})^* = e^{-ix}$ one has:

$$\cos[x] = \mathrm{Re}\, e^{ix}, \quad sin[x] = \mathrm{Im}\, e^{ix}$$

With these relations one can forget all formulae about sinus and cosinus !

Logarithm

log or ln is the inverse function of e^x:

$$e^{\log[x]} = x; \quad \log[e^x] = x$$

$$\log[a \cdot b] = \log[a] + log[b]$$

The decadic logarithm, $\log[10, x]$ is the inverse function of 10^x.

$$\log[10, x] = \log[x]/\log[10] \approx \log[x]/2.3$$

1.2 The wave equation

The basis for the production and propagation of sound is the Newton equation

$$m\partial_t^2 x = K,$$

where the force K is determined by the matter in question. For a metal string e.g. it is Hooke's "Law": $K \sim \Delta L$. For sound it is the law of adiabatic compression: the variation of pressure with density is (approximately) constant

$$\frac{\partial p}{\partial \rho}_{/\rho} = \text{const}$$

.

If a string deviates from its straight line, due to Hooke's Law a force pulls it back, and the resulting equation is the wave equation in 1+1 dimensions (1 time, 1 space).

If a small volume of air is displaced, pressure increases in the direction of displacement and decreases in the opposite direction. This drives the air back and leads again to a wave equation, but in 1+3 dimensions.

Wave equation in 1+1 dimensions

We discuss especially the wave equation in 1+1 dimensions,that is in one time and one space dimension. It is:

$$(\partial_t^2 - c^2\partial_x^2)\, y(x,t) = K \tag{1.4}$$

The variable $y(x,t)$ can e.g. be the transverse excitation of an ideal string, the pressure or density of an ideal tube filled with gas. c is the velocity of sound (pressure wave) in the material.

Schallgeschwindigkeit in mehreren Medien
(Normalbedingungen):

Medium	c [m/sec]
Luft	331
Wasser-Dampf	401
Wasser, flüssig	1497
Buchenholz	3412
Tannenholz	5256
Messing	3400

There exists a general solution for the homogeneous case ($K = 0$) (Laplace ca. 1750):

$$y(x,t) = f(x - ct) + g(x + ct) \tag{1.5}$$

where f and g are arbitrary functions.

Proof by insertion:

$$\partial_t^2 y = f''(x - ct)(-c)^2 + g''(x + ct)c^2; \quad \partial_x^2 y = f''(x - ct) + g''(x + ct)$$

This solution consists of two wave packets with arbitrary form $f(z)$ and $g(z)$. f is moving to the right, g to the left.

Example: Be f a Gaussian $f(z) = \exp[-z^2]$, then at $t = 0$ $f(x - ct) = \exp[-(x - ct)^2]$ is a Gaussian, centered at $x = 0$, at the later time t_1, it is a Gaussian centered at $x - ct_1$, that is it has moved with the velocity c to the right. A similar reasoning shows that the packet $g(x + ct)$ moves with velocity c to the left, see figures 1.2, 1.2.

The solution can be made unique by prescribing initial conditions and boundary conditions. We shall not dwell on that, but look for another form of solution found by Bernoulli.

We consider the boundary conditions

$$y(0,t) = y(L,t) = 0 \tag{1.6}$$

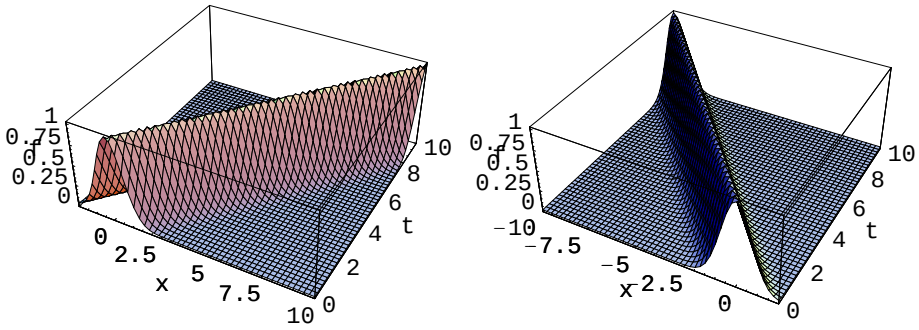
Then the most general solution is:

$$y(x,t) = \sum_{n=0}^{\infty} A_n \, \sin(n\,\omega_0\,t + \delta_n) \, \sin(\frac{n\,\pi}{L}\,x) \tag{1.7}$$

with $\omega_0 = \frac{\pi\,c}{L}$

Solutions of this type are relevant for all melody instruments (air and string). The arbitrary constants can be fixed by initial conditions, $y(x,0)$, $\partial_t y(x,0)$

Exercise: Find the general solution with $y(0,t) = 0$, $\partial_x\,y(L,t) = 0$ (Flute).



$f(x-ct)$ $g(x+ct)$

Abbildung 1.1: The right and left moving wave packet $f(x-ct)$ and $g(x-ct)$.

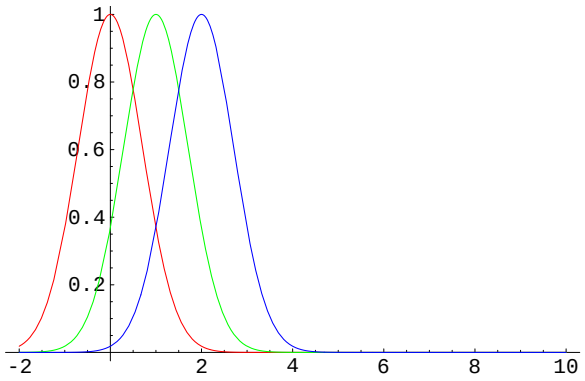


Abbildung 1.2: Time slices $t = 0, 1, 2$ for the right moving packet $f(x-ct)$.

Wave equation in 1+3 dimensions

The wave equation in three space dimensions , relevant for instance for propagation of sound, is

$$(\partial_t^2 - c^2 \sum_{i=1}^{3} \partial_i^2)\, y(\vec{x}, t) = 0 \tag{1.8}$$

There is no general analytic solution in that case. If sound is produced near $\vec{x} = 0$, then for $r = |\vec{x}\,| \to \infty$ a solution is

$$y(\vec{x}, t) \sim \frac{1}{r} e^{ikr} e^{-i\omega t}$$

with $k = c\omega/c = 2\pi/\lambda;\ \omega = 2\pi\nu$

Kapitel 2

Mathematical tools

2.1 Fourier transform

2.1.1 The Fourier series

The beginning of Fourier analysis was the Fourier series (Fourier 1835). Any (sufficiently well behaved) function $f(t)$ in the interval $[-L/2, L/2]$ can be expressed as a series:

$$f(t) = \sum_{n=0}^{\infty} \Big(A_n \, \sin(2\pi \, n \, t/L) + B_n \, \cos(2\pi n \, t/L) \Big) \tag{2.1}$$

It is more convenient to work with the series

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \, e^{-2\pi \, i \, t \, n/L} \tag{2.2}$$

The coefficients C_m are:

$$C_m = \frac{1}{L} \int_{-L/2}^{L/2} dt \, f(t) e^{+2\pi \, i \, t \, m/L} \tag{2.3}$$

This follows from

$$\frac{1}{L} \int_{-L/2}^{L/2} dt \, f(t) e^{2\pi \, i \, t \, (m-n)/L} = \delta_{nm} \tag{2.4}$$

$$\delta_{nm} = \begin{cases} 1 & \text{for} \quad m - n \\ 0 & \text{for} \quad m \neq n \end{cases} \qquad \text{(Kronecker delta)} \tag{2.5}$$

Eq. 2.2 is the basis of all numerical Fourier transformation programs.

2.1.2 The Fourier Integral

If one lets L go to ∞, the steps $2\pi/L$ become smaller and one ends with the integral:

$$f(t) \equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\ \tilde{f}(\omega) e^{-i\omega t} \tag{2.6}$$

$\tilde{f}(\omega)$ is called the Fourier transform or *spectrum* of $f(t)$.

To simplify language we shall call the t-dependent function *signal*. The variable ω is called *circular frequency*, it is related to the usual frequency ν by

$$\omega = 2\pi\nu \tag{2.7}$$

. The integral 2.6 can therefore also be written as:

$$f(t) \equiv \int_{-\infty}^{\infty} d\nu\, \tilde{f}(2\pi\nu) e^{-i2\pi\nu t}. \tag{2.8}$$

The spectrum is obtained from the signal $f(t)$ by the formula (Fourier transform):

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} dt\ f(t) e^{i\omega t} \tag{2.9}$$

Note: In physics the convention is usually to have the $\int f dt\, e^{+i\omega t} \ldots$ for the Fourier transform $f(t) \to \tilde{f}(\omega)$, 2.9, and $\int \frac{d\omega}{2\pi}\, e^{-i\omega t} \ldots$, 2.6, for the back-Fourier transform $\tilde{f}(\omega) \to f(t)$. Although this is pure convention, I recommend to adhere strictly to it. If you are not consequent, you will end up in a mess!!.

The squared modulus

$$|\tilde{f}(\omega)|^2 \tag{2.10}$$

is called *power spectrum*.

The Fourier transform is linear:

$$\alpha\, f(t) + \beta\, g(t) \to \alpha \tilde{f}(\omega) + \beta \tilde{g}(\omega) \tag{2.11}$$

If the function $f(t)$ is real, i.e. if f(t) = f*(t), then

$$\tilde{f}(\omega) = \tilde{f}^*(-\omega). \tag{2.12}$$

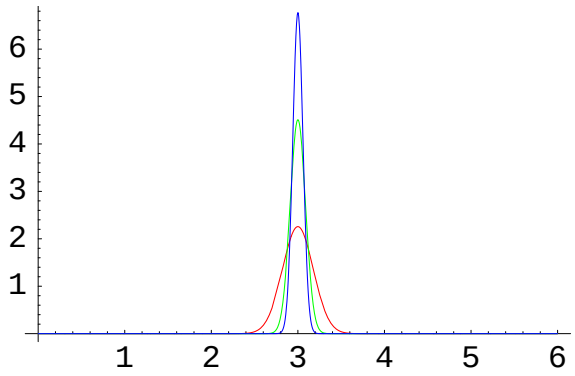


Abbildung 2.1: The delta distribution as the limit of functions. $\lim_{\lambda\to\infty} \frac{\lambda}{\sqrt{\pi}}\, e^{-\lambda^2(t-\tau)^2}$ does not exist, but $\lim_{\lambda\to\infty} \int dt\, \frac{\lambda}{\sqrt{\pi}}\, e^{-\lambda^2(t-\tau)^2} f(t) = f(\tau)$ for any function continuous at $t = \tau$.

For a real function the spectrum is therefore completely determined by the positive values of ω:

$$\begin{aligned} f(t) &= \int_0^\infty \frac{d\omega}{2\pi} (\tilde{f}(\omega) e^{-i\omega t} + \tilde{f}^*(\omega) e^{i\omega t}) \\ &= \int_0^\infty \frac{d\omega}{2\pi}\, 2\,\mathrm{Re}\left(\tilde{f}(\omega) e^{-i\omega t}\right) \end{aligned} \tag{2.13}$$

The power spectrum 2.10 is also determined completely by positive values of ω, since

$$|\tilde{f}(\omega)|^2 = |\tilde{f}(-\omega)|^2| \tag{2.14}$$

The δ distribution The extension of the Kronecker delta into the continuum is the δ distribution, $\delta(t-\tau)$, defined by:

$$\int_{-\infty}^{\infty} d\tau\, \delta(t-\tau) f(\tau) = f(t) \tag{2.15}$$

One can view the δ distribution as the limit of a function which is different from zero only for a small region $t \approx \tau$ and which has the integral 1.

Example: $\frac{\lambda}{\sqrt{\pi}}\, e^{-\lambda^2(t-\tau)^2}$ with $\lambda \to \infty$, see figure 2.1.2. The Fourier transform of the δ function is:

$$\tilde{\delta}(\omega) = \int_{-\infty}^{\infty} dt\ \delta(t-\alpha) e^{i\omega t} = e^{i\omega\alpha} \tag{2.16}$$

The back-Fourier transform 2.6 then yields:

$$\int \frac{d\omega}{2\pi} e^{i\omega\alpha} e^{-i\omega t} = \delta(t-\alpha) \tag{2.17}$$

A true miracle, which can be justified in distribution theory.

The distributions are generalized functions. The Fourier transformations can be applied also to them.

Problem: Show with the help of 2.17 that 2.6 holds.

2.1.3 Filtering

2.1.4 Definition

A filter is a function $\tilde{\phi}(\omega)$ modifying the spectrum $\tilde{p}(\omega)$ of a signal $p(t)$

Signal:

$$p(t) \tag{2.18}$$

spectrum of signal

$$\tilde{p}(\omega) = \int_{-\infty}^{\infty} dt\, p(t) e^{i\omega t} \tag{2.19}$$

Filter:

$$\tilde{\phi}(\omega) \tag{2.20}$$

Filtered spectrum:

$$\tilde{p}_f(\omega) = \tilde{\phi}(\omega) \cdot \tilde{p}(\omega) \tag{2.21}$$

Filtered signal:

$$p_f(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{p}_f(\omega) e^{-i\omega t}. \tag{2.22}$$

The simplest filter is a sharp high or low pass filter, which eliminates all frequencies above (low pass) or below (high pass) a certain frequency $\nu_0 = \frac{\omega_0}{2\pi}$.

high pass: $\tilde{\phi}_{\mathrm{hp}}(2\pi\nu) = \theta(2pi\nu - 2\pi\nu_0)$
low pass: $\tilde{\phi}_{\mathrm{lp}}(2\pi\nu) = \theta(2pi\nu 0 - 2pi\nu)$

Before we go into the details of filtering, I shall show its effects on a realistic case, namely signals evoked by acoustical stimuli in the brain, neural signals and their detection will be discussed in detail in section **??**.

2.1.5 The influence of filtering on a neurological signal

In fig. 2.2 neural signals evoked by an acoustical signal in three different individuals are shown. The strength of the signal in nano-Ampére-meter (nAm) is determined by a dipole fit to the evoked magnetic fields as measured by a Magnetoenkephalogram. There are positive signals at ca 40 ms after tone onset (P1) and 200 ms (P2) and a prominent negative one 100 ms after onset (N1 or N100). In the upper left one (MA) there are only clear transient signals, but no distinct sustained field, that is a cortical activity during the whole tone duration (500 ms), for the second individual (HD) a small sustained field is visible, the sustained field is very prominent for the third one (ZH). There the tone is a piano tone with 750 ms duration. I shall give some examples of high and low pass filtering. Since the signal curve $p(t)$ is given as a list of function values in steps of 1 ms, one has to perform filtering numerically. One performs a numerical Fourier analysis based on eq. **??** and then modifies the resulting spectrum, which will be also discrete, in steps of 1 Hz, when the original signal has a duration of 1 s. The maximal frequency is 500 Hz, this will be explained in detail in sec. 2.1.8. The spectrum, modified according to **??** and **??**, will be back transformed in order to obtain the filtered signal.

The resulting signals for high pass filtering with ν_0 at 1, 2, 4 and 6 Hz are displayed in fig. 2.3. We see low pass filtering with low cutoff has little effect, but that for $\nu > 3$ Hz the signal is rather strongly modified (red curves).

In fig. **??** a constant term is added to the signal, we see that cutting out the 0-the restores practically the original signal. This is due to the fact, that a constant contributes only to the zero component of the spectrum, that is the average of the curve, and that in this average is near zero.

If a linearly rising term is added, the filtering is no efficient way to eliminate the added background, as can be seen from fig. 2.5

We now come to low pass filtering. In figure 2.6 the signal are is low pass filtered with the upper frequencies $\nu_0 = 30, 20$ and 10 Hz, respectively. Apart from eliminating some jitter the low pass filtering does not disturb the signal for $\nu_0 = 30$ and 20 Hz, but it does so definitely for $\nu_0 = 10$ Hz.

Low pass filtering is an efficient way to eliminate noise. In figure 2.7 a noise is added to the original signal, which is efficiently reduced by low pass filtering with $\nu_0 = 30$ Hz.

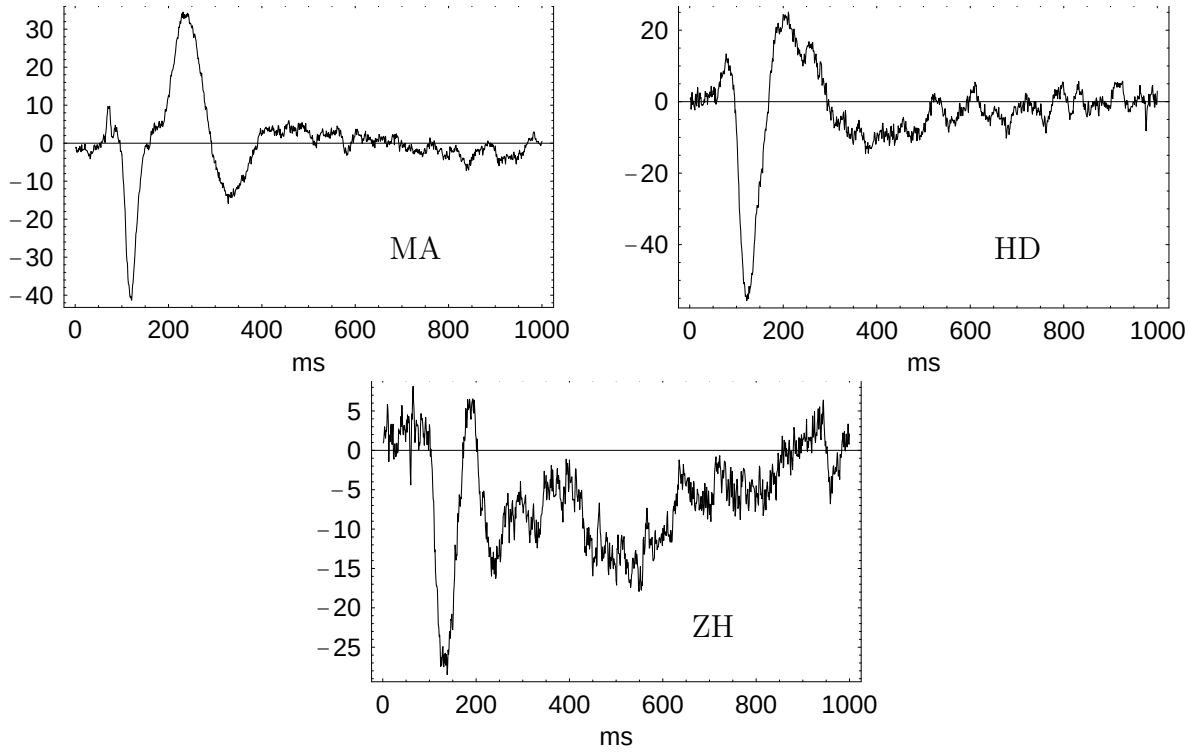


Abbildung 2.2: Cortical signal evoked by a sinusoidal tone, starting at ca 20 ms. There are positive signals at ca 40 ms after tone onset (P1) and 200 ms (P2) and a prominent negative one 100 ms after onset (N1 or N100). In the upper left one (MA) there are only clear transient signals, but no distinct sustained field, that is a cortical activity during the whole tone duration (500 ms), for the second individual (HD) a small sustained field is visible, it is very prominent for the third one (ZH). There the tone is a piano tone with 750 ms duration

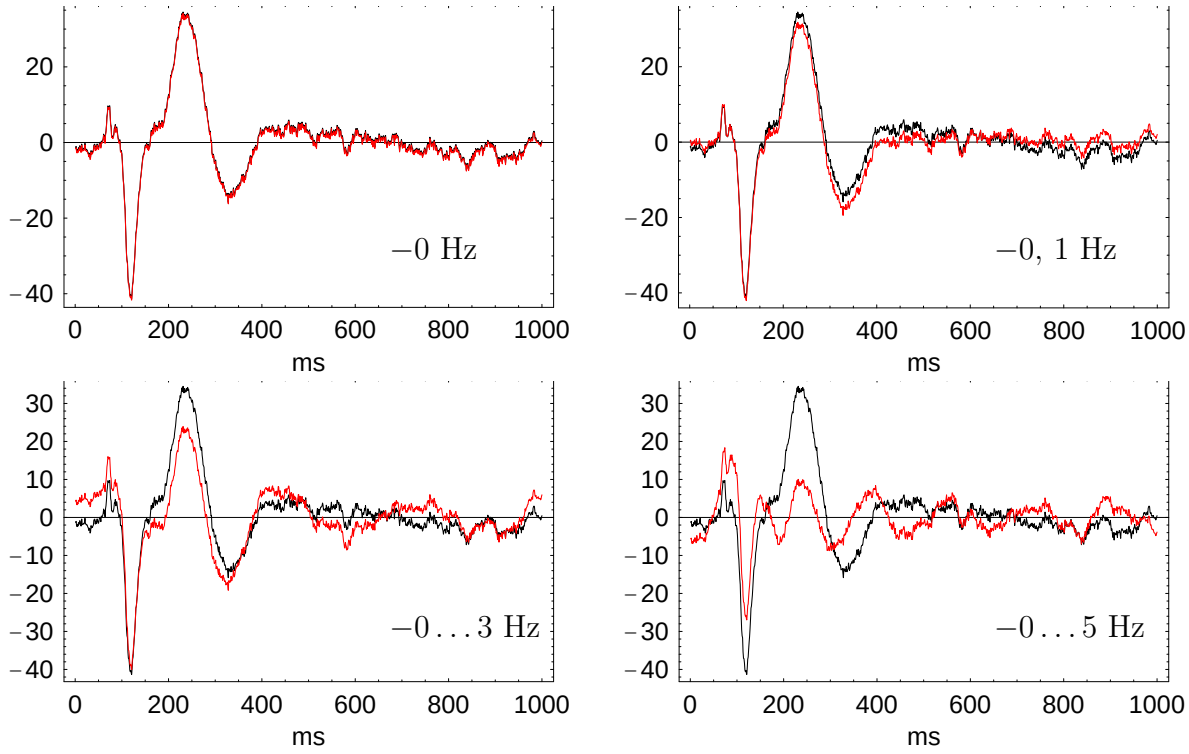


Abbildung 2.3: The high pass filtered cortical signal of fig. 2.2 (MA) with frequency 0, 0-1, 0-4, and 0-6 Hz cut out. The black curve is the original signal, the red one the filtered one.

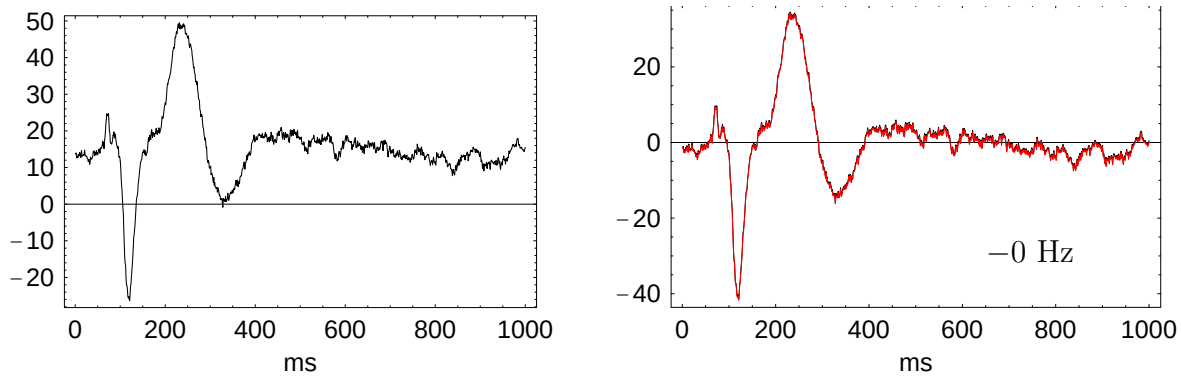


Abbildung 2.4: The cortical signal of fig. 2.2 (MA) with a constant contribution of 15 nAm added, below the high pass filtered signal with frequency 0 cut out. The original signal is black, the filtered one red.

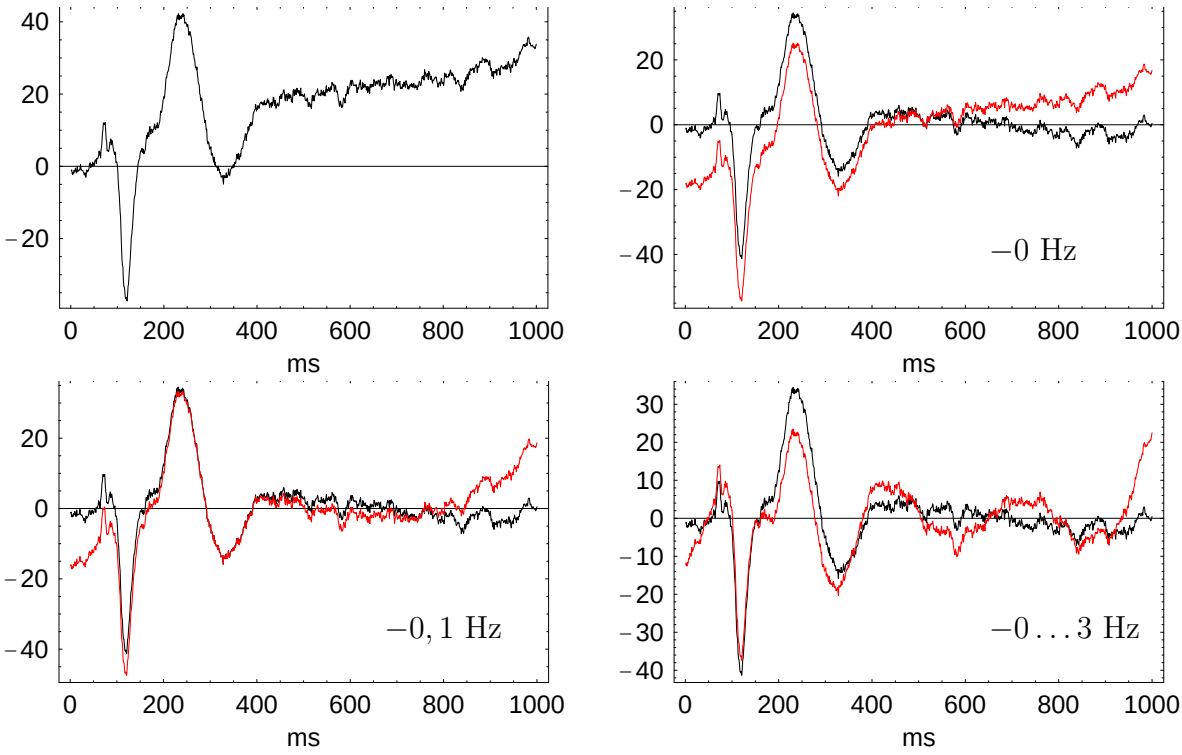


Abbildung 2.5: The cortical signal of fig. 2.2 with a linear rising contribution added, below the high pass filtered signal with frequency 0, 0-1, 0-3 cut out. The original signal is black, the filtered one red.

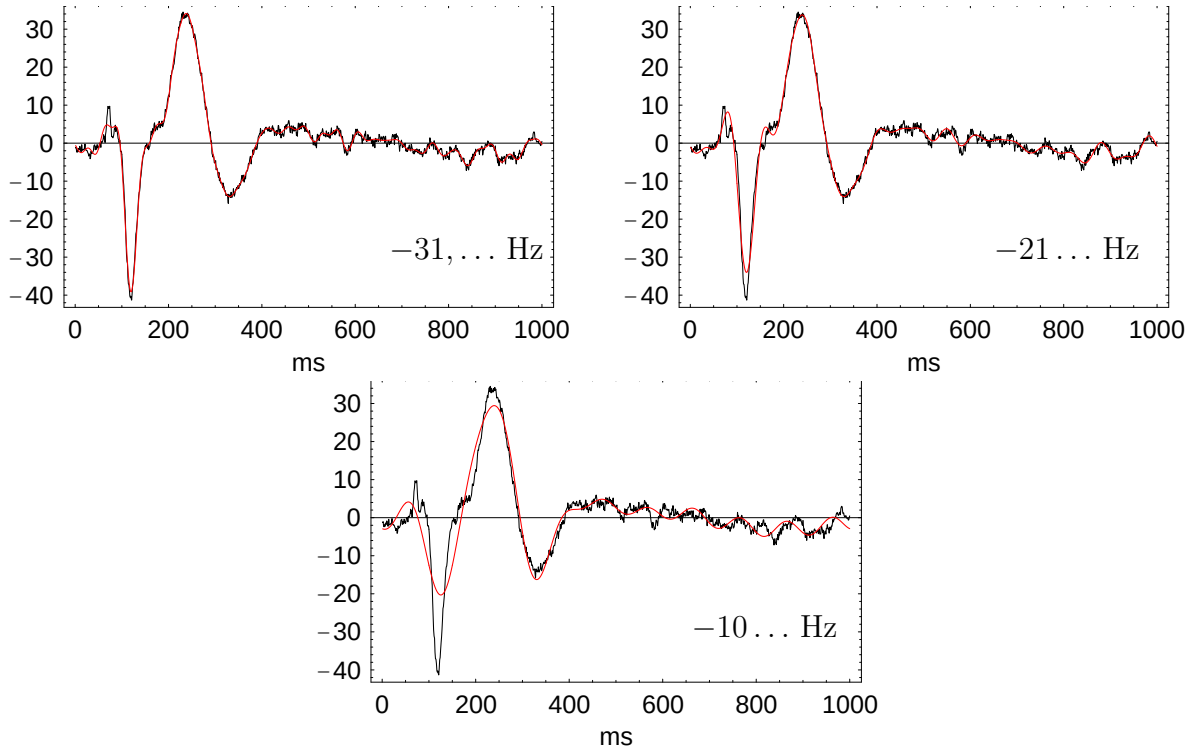


Abbildung 2.6: The low pass filtered signal of fig. 2.2 with frequencies above 30, 20 an 10 Hz cut out respectively. The original signal is black, the filtered one red.

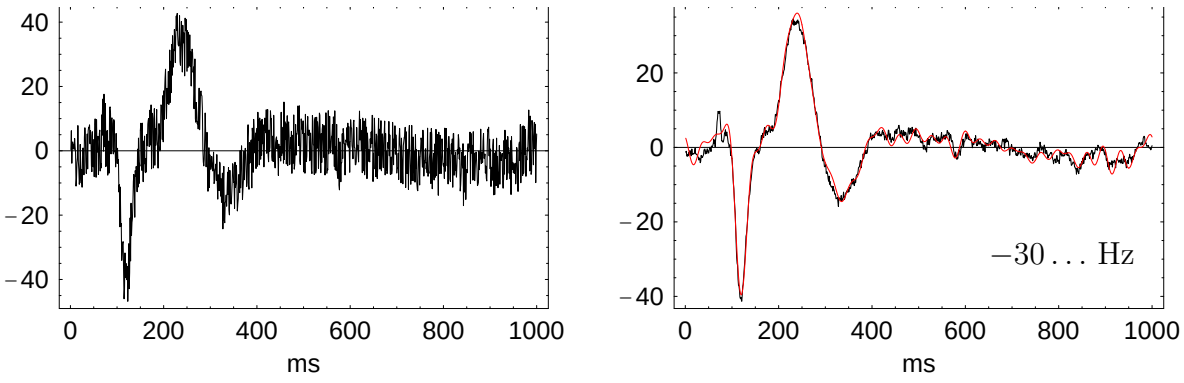


Abbildung 2.7: The cortical signal of fig. 2.2 with a noise added, and the low pass filtered signal with frequencies above 30 Hz cut out.

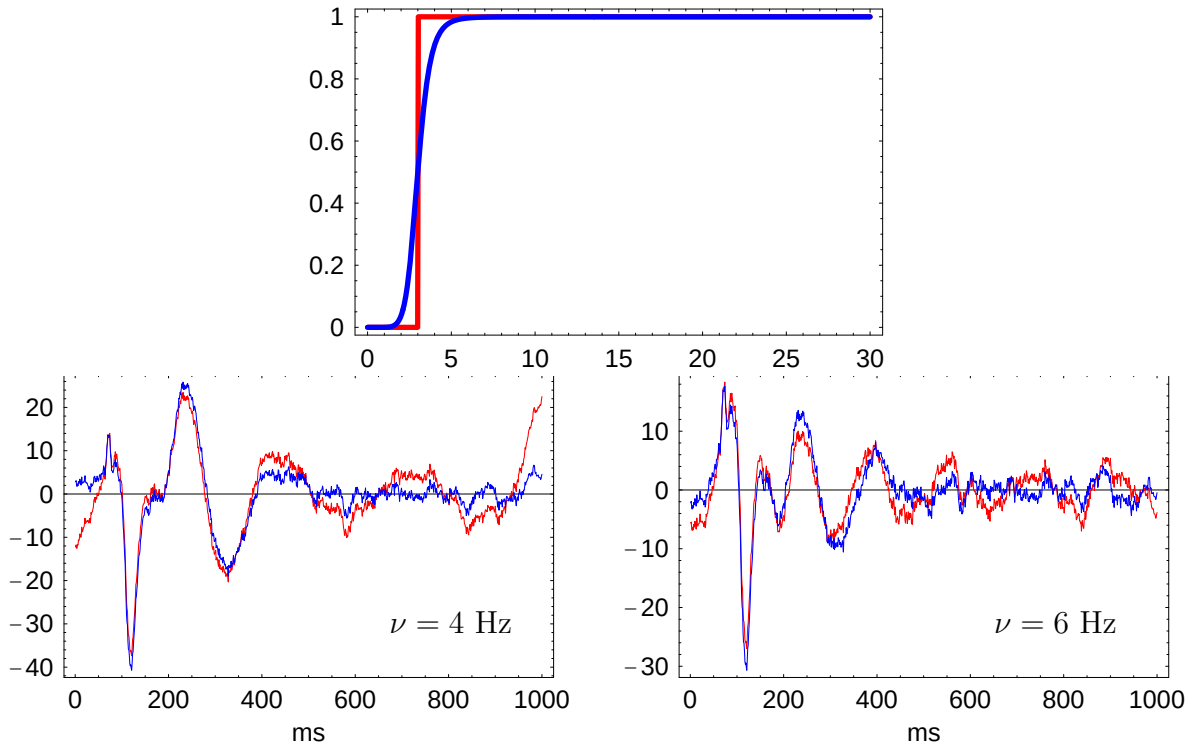


Abbildung 2.8: Hard (red) and soft (blue) high pass filter with limiting frequency $\nu = 3$ Hz. The signals of 2.2, MA high pass filtered with hard (red) and soft (blue) filter with limiting frequency $\nu = 4$ and $nu = 6$ Hz respectively.

Up to now we have filtered with a sharp cutoff, fig. 2.8, red. If we high pass filter with a soft cutoff (blue) the filter induces less oscillations.

We If the signal (ZH) (see fig. 2.2 is filtered, we see in figure 2.9 that high pass filtering even with only cutting out the 0-Hz component distorts the signal. High pass filetering with a cutoff at 30 Hz only smoothes the signal, fig. 2.10

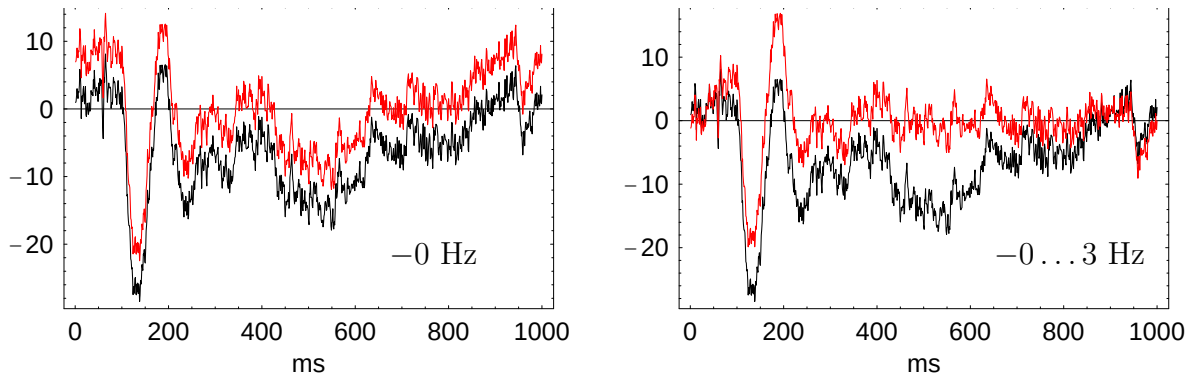


Abbildung 2.9: The signal (ZH) of fig. 2.2 high passed filtered with 0 Hz and 0 . . . 3 Hz cut out.

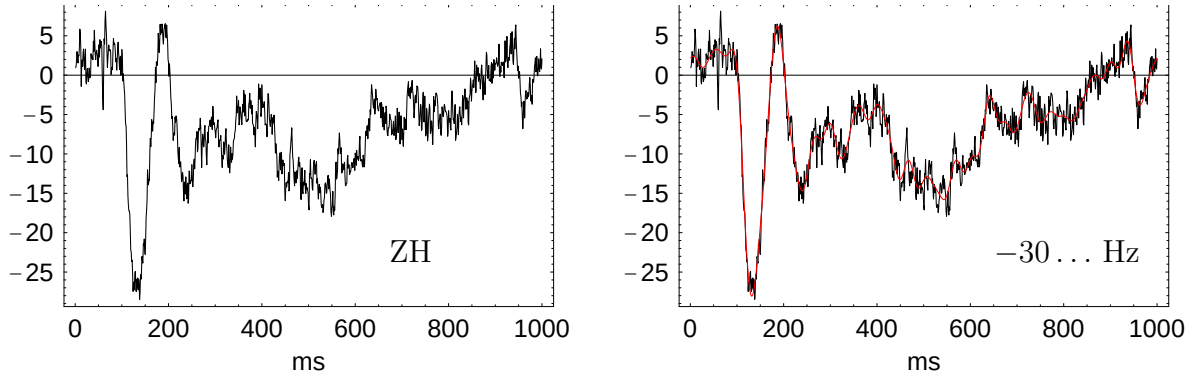


Abbildung 2.10: The signal (ZH) of fig. 2.2 low passed filtered with frequencies > 30 Hz cut out.

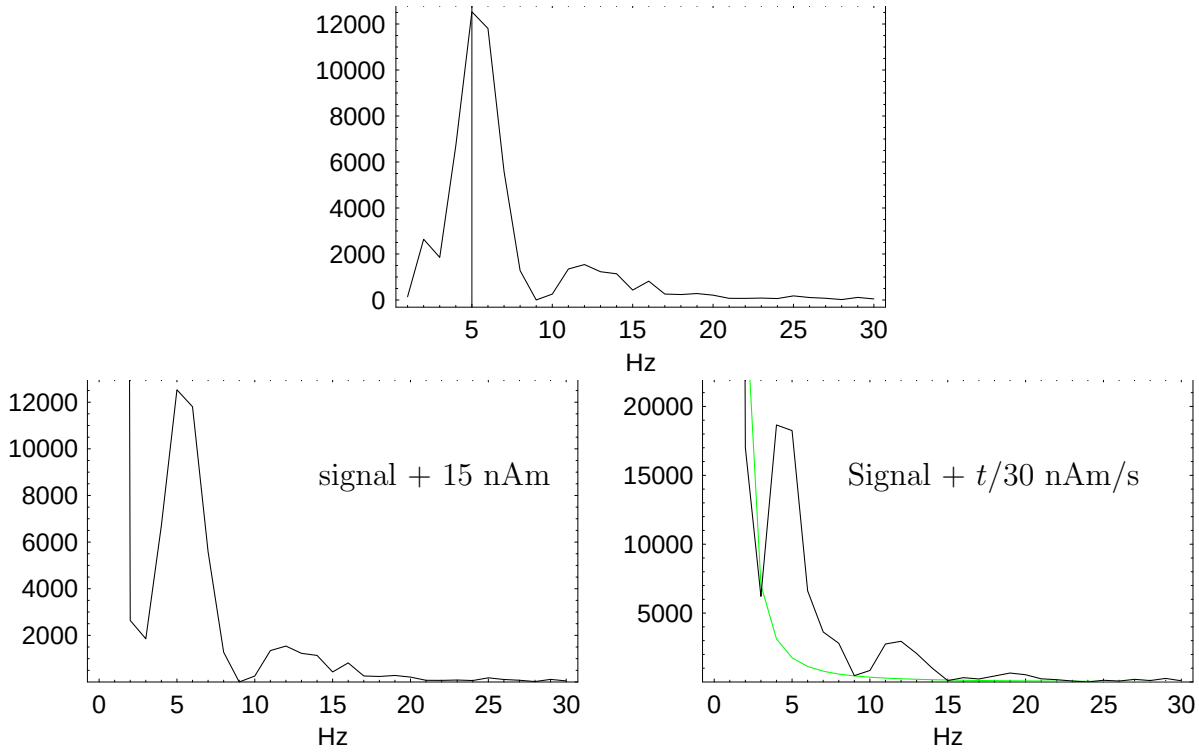


Abbildung 2.11: Upper row: the spectrum of the signal of fig. 2.2 (MA). Lower row: the same with a constant contribution of 15 nAm added and with a linear contribution $t/30$ nAm/s added

In order to obtain information on the filter distortions *a priori*, one has to inspect the spectrum. The power spectra for the original signal and the ones with a constant and linear background added are shown in fig. 2.11. We see, that the spectrum of the linear background overlaps with essential parts of the spectrum of the signal.

In figure 2.12 the spectra of the signals of individuals ZH and HD are shown. The large sutained field manifests itself in a strong increase of the spectrum at small frequencies. Therefore high pass filtering leads to a distortions of the signal. For HD the sustained field is less marked, therefore i a filtering would lead to similar results as for teh signal of MA.

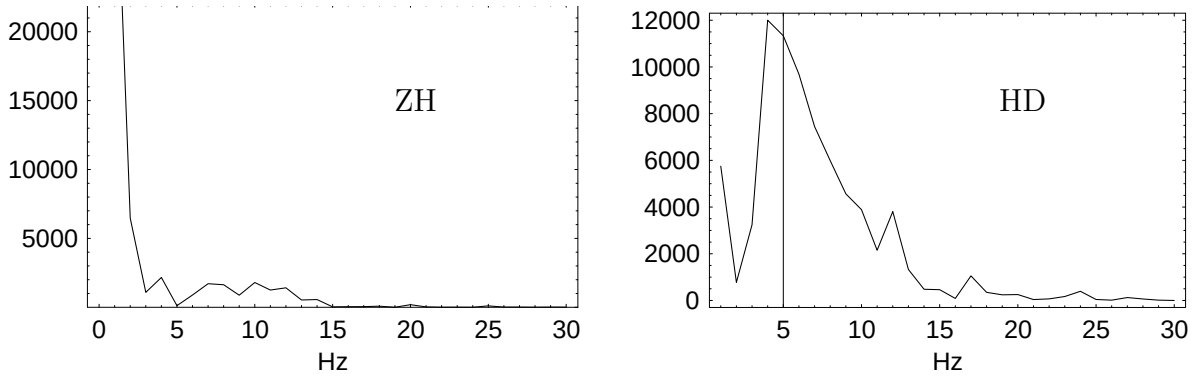


Abbildung 2.12: The spectra of the signals for ZH and HD of fig. 2.2

2.1.6 Direct filtering of the signal. Convolution

How to express the filtered signal $p_f(t)$ directly by the signal $p(t)$ and the time dependent filter function

$$\phi(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{\phi}(\omega) e^{-i\omega t}? \tag{2.23}$$

To answer that question we insert in 2.22 the expression for filtered spectrum 2.21 and express the spectra by the time dependent functions:

$$\begin{aligned}
p_f(t) &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \tilde{\phi}(\omega) \cdot \tilde{p}(\omega) && (2.24)\\
&= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \int_{-\infty}^{\infty} dt' \phi(t') e^{i\omega t'} \int_{-\infty}^{\infty} dt'' p(t'') e^{i\omega t''} \\
&= \int\int\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} dt'\, dt'' e^{-i\omega(t-t'-t'')} \phi(t')\, p(t'') \\
&\quad \text{using the miracle 2.17 one obtains} \\
&= \int\int\int_{-\infty}^{\infty} \delta(t-t'-t'') dt'\, dt'' \phi(t')\, p(t'') && (2.25)\\
&= \int_{-\infty}^{\infty} dt' \phi(t') p(t-t') && (2.26)\\
&= \int_{-\infty}^{\infty} dt'' \phi(t-t'') p(t'') && (2.27)
\end{aligned}$$

The last three equations are expressions for the *convolution* of the time dependent filter function $\phi(t)$ and the signal $p(t)$, which will be discussed in more detail in the following paragraph.

Convolution **Definition:** The convolution of two functions $f(t)$ and $g(t)$ is defined as:

$$\begin{aligned}
f * g\,(t) &= \int\int\int_{-\infty}^{\infty} \delta(t-t'-t'') dt'\, dt'' f(t')\, g(t'') && (2.28)\\
&= \int_{-\infty}^{\infty} dt' f(t')\, g(t-t') && (2.29)\\
&= \int_{-\infty}^{\infty} dt'' f(t-t'')\, g(t'') && (2.30)
\end{aligned}$$

Comparing 2.28 with 2.25 - 2.27 we therefore obtain for the filtered signal 2.22 :

$$p_f(t) = (\phi * p)\,(t) \tag{2.31}$$

More generally, since the derivation is valid for all functions $\phi(t)$ and $p(t)$ we have "proven" the following theorem:

The back-Fourier transform from a product of two spectra is the convolution of the signals.

$$\tilde{f}(\omega)\,\tilde{g}(\omega) \to (f * g)(t)$$

Since the Fourier transform 2.9 and the back-Fourier transform 2.6 are symmetric ($i\omega$ is replaced by $-i\omega$) we can also state the following

Theorem:The Fourier transform of a product is the convolution of the Fourier transforms of the products:

$$\tilde{(f \cdot g)}(\omega) = \tilde{f} * \tilde{g} \tag{2.32}$$

or explicitely:

$$\int_{-\infty}^{\infty} dt\, f(t) \cdot g(t) e^{i\omega\, t} = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \tilde{f}(\omega') \cdot \tilde{g}(\omega - \omega') \tag{2.33}$$

Exercise: Show that, following the lines of the derivation of 2.24.

Alternatively to the definition of a filter based on the spectrum, given in sect. **??**, we can also define filtering exclusively in the t-space:

A Filter is a function $\phi(t)$ and the filtered signal is the convolution of the signal with the filter:

$$p_f(t) = (\phi * p)(t) = \int_{-\infty}^{\infty} dt' \phi(t') p(t - t') = \int_{-\infty}^{\infty} dt' \phi(t - t') p(t') \tag{2.34}$$

Since convolution integrals play an important role in biophysics, we shall study in detail an example for a convolution:

$f(t) = t + \sin(5\,t); \quad g(t) = \frac{2}{\pi}\, e^{-4t^2};$

in Fig. 2.25 $f(t)$ and $g(t - \tau)$ for $\tau = 0$, 1, and 2 are displayed. The resulting expression:

$$\int_{-\infty}^{\infty} dt\, f(t)\, g(\tau - t)$$

is shown in figure 2.26. As we can see the effect of the convolution consists in the smoothing of the signal curve $f(t)$.

2.1.7 Examples of filters

For a real signal $p(t)$ one wants a filter which leads to a real filtered function $p_f(t)$, that is also the time dependent filter $\phi(t)$ must be real. Therefore the filter function $\tilde{\phi}(\omega)$ must fulfill the

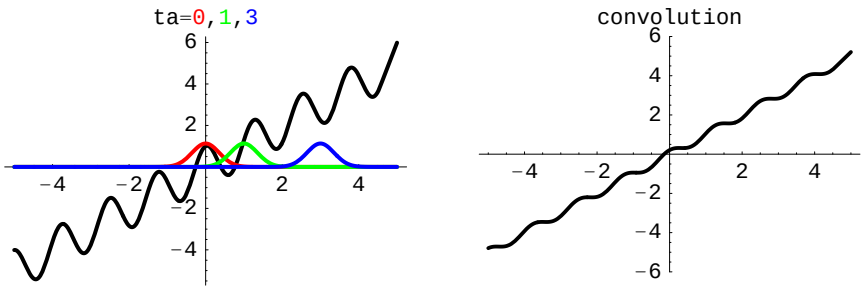


Abbildung 2.13: Left: The function $f(t) = t + \sin(5\,t)$ and $g(t) = \frac{2}{\pi}\, e^{-4(t-\tau)^2}$ for $ta = 1, 2, 3$, Right: The convolution $f * g$.

condition 2.12, i.e.:

$$\tilde{\phi}(\omega) = \tilde{\phi}^*(-\omega). \tag{2.35}$$

that is it is sufficient to define the filter for positive frequencies.

Since also the spectrum of a real signal has this property, we can express the filtered signal also as

$$p_f(t) = \int_0^\infty \frac{d\omega}{2\pi}\, \tilde{\phi}(\omega)\, \tilde{p}(\omega) \tag{2.36}$$

The most transparent filters are the ones with sharp cutoff, which we have already mentioned in sect. **??**.

Sharp cutoffs tend to induce oscillations, therefore quite often the cutoff is somewhat smoothed. Another reason for smooth filters is the fact, that before the digital age filtering had to be performed by electronics. Therefore the miltipole or Butterworth filters are still popular. A $2n$-pole high pass filter (Butterworth filter) with cutoff frequency $\omega_0 = 2\pi\nu_0$, which is also easy to program is given by:

$$\tilde{\phi}(\omega) = 1/(1 + (\omega/\omega_0)^{2n}) \tag{2.37}$$

In fig. 2.8 a high pass $2n = 8$ pole filter is shown, a low pass filter with $2n = 2$ and $2n = 16$, respectively, and a cutoff $\nu = 10$Hz is shown in figure 2.14

A similar soft filter edge is the Hann filter. For a limiting frequency g and an edge width w the high pass filter is given by, see fig. hann:

$$\phi_{\text{Hann}}(\nu) = \begin{cases} 0 & \text{for} \quad \nu < g - w \\ \sin\left(\frac{\pi}{2w}(\nu - g)\right)/2 + 1/2 & \text{for} \quad g - w \geq \nu \leq g + w \\ 1 & \text{for} \quad \nu > g + w \end{cases} \tag{2.38}$$

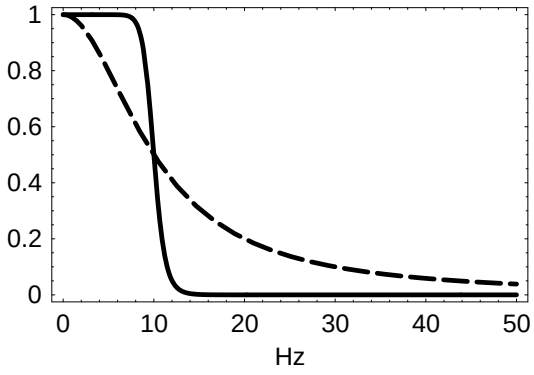


Abbildung 2.14: A Butterworth low pass filter $\tilde{\phi}(\omega) = 1/(1 + (\omega/\omega_0)^{2n}$ with $\nu_0 = \omega_0/(2\pi) = 10$ and with $2n = 2$ (dashed), and $2n = 16$ (solid)

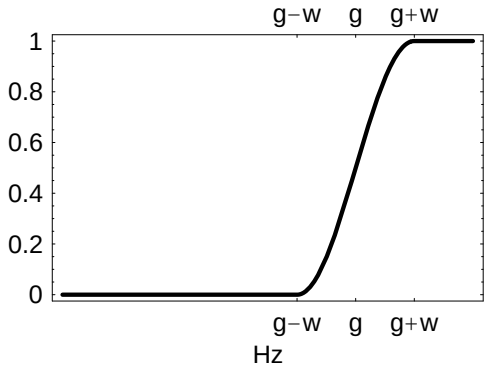


Abbildung 2.15: A Hann high pass filter

It is clear that in a oscillating function a high pass filter cuts out the slow oscillations and a low pass filter the fast oscillations. It is quite instructive to see, what are the effects of filtering for a non-oscillating function, as a square wall or a Gaussian.

Filtering of a Gaussian and a square wall For a Gaussian function $e^{-(t-c)^2/b^2}$ and a square wall the spectra can be calculated analytically:

For the Gaussian we show it with $c = 0$:

$$\int_{-\infty}^{\infty} dt e^{-t^2/b^2} e^{i\omega t} \;=\; \int_{-\infty}^{\infty} dt \, \exp\Big[-\frac{1}{b^2}(t^2 - 2i\omega b^2/2 - om^2 b^4/4) - \omega^2 b^2/4 \Big]$$

signal	spectrum
$p(t) = e^{-(t-c)^2/b^2}$	$\tilde{p}(2\pi\nu) = e^{i2\pi\nu c}\, b\sqrt{\pi} e^{-b^2(2\pi\nu)^2/4}$
$p(t) = \theta(t-c-b)\cdot\theta(c+b-t)$	$\tilde{p}(2\pi\nu) = e^{i2\pi\nu c}\, 2\frac{\sin(b2\pi\nu)}{2\pi\nu}$

Tabelle 2.1: Gaussian and square wall and their Fourier transforms

$$= \quad e^{-\omega^2 b^2/4} \underbrace{\int_{-\infty}^{\infty} dt\, \exp\Big[-\frac{1}{b^2}(t - i\omega b/2)^2\Big]}_{b\sqrt{\pi}}$$

For the square wall one obtains easily:

$$\int_{c-b}^{c+b} dt e^{i\omega t} = \frac{1}{i\omega}\Big[e^{i\omega(c+b)} - e^{i\omega(c-b)}\Big] = e^{i\omega c} 2\frac{\sin(\omega b)}{\omega} \tag{2.39}$$

We see from the formulae and also from the figure 2.16: The more square the signal, the more wavy the spectrum and the thinner the signal the thicker is the spectrum and *vice versa*.

We now filter the Gaussian and the square wall signals. In figure 2.17 the Gaussian and square wall are displayed. The black curve is the original signal, the red curve the filtered signal. We show a rectangular high pass filtering with lower frequency 2 Hz, and a low pass filtering with upper frequencies of 30 and 100 Hz.

A filter is called *zero phase filter*, if $\tilde{\phi}(\omega)$ is real. In that case the function $\phi(t)$ is symmetric: $\phi(t) = \phi(-t)$, that is a zero phase filter cannot be causal.

A special role in biophysics play *causal filters*. A filter is called *causal*, if $\phi(t) = 0$ for $t < 0$. In that case we have:

$$p_f(t) = \int_0^{\infty} dt'\, \phi(t')\, p(t - t') \tag{2.40}$$

that is the filtered signal $p_f(t)$ depends only on values of $p(t')$ with $t' \leq t$.

If $p(t)$ is the physical excitation of a system, e.g. energy input, the system will very often not accept the excitation immediately, but with some time delay. In that case the response of the system will be a causally filtered excitation. Example: A system responds on the average excitation of the last 5 ms, then the filter is a square wall, if the time is counted in ms, then

$$\phi(t) = \theta(t)\theta(5 - t) = \begin{cases} 1/5 & \text{if } 0 \leq t \leq 5\,\text{ms} \\ 0 & \text{otherwise} \end{cases} \tag{2.41}$$

If the excitation (signal) is given as a list of numbers, it would be silly to go back to the spectra. In that case it is very fast to perform the convolution.

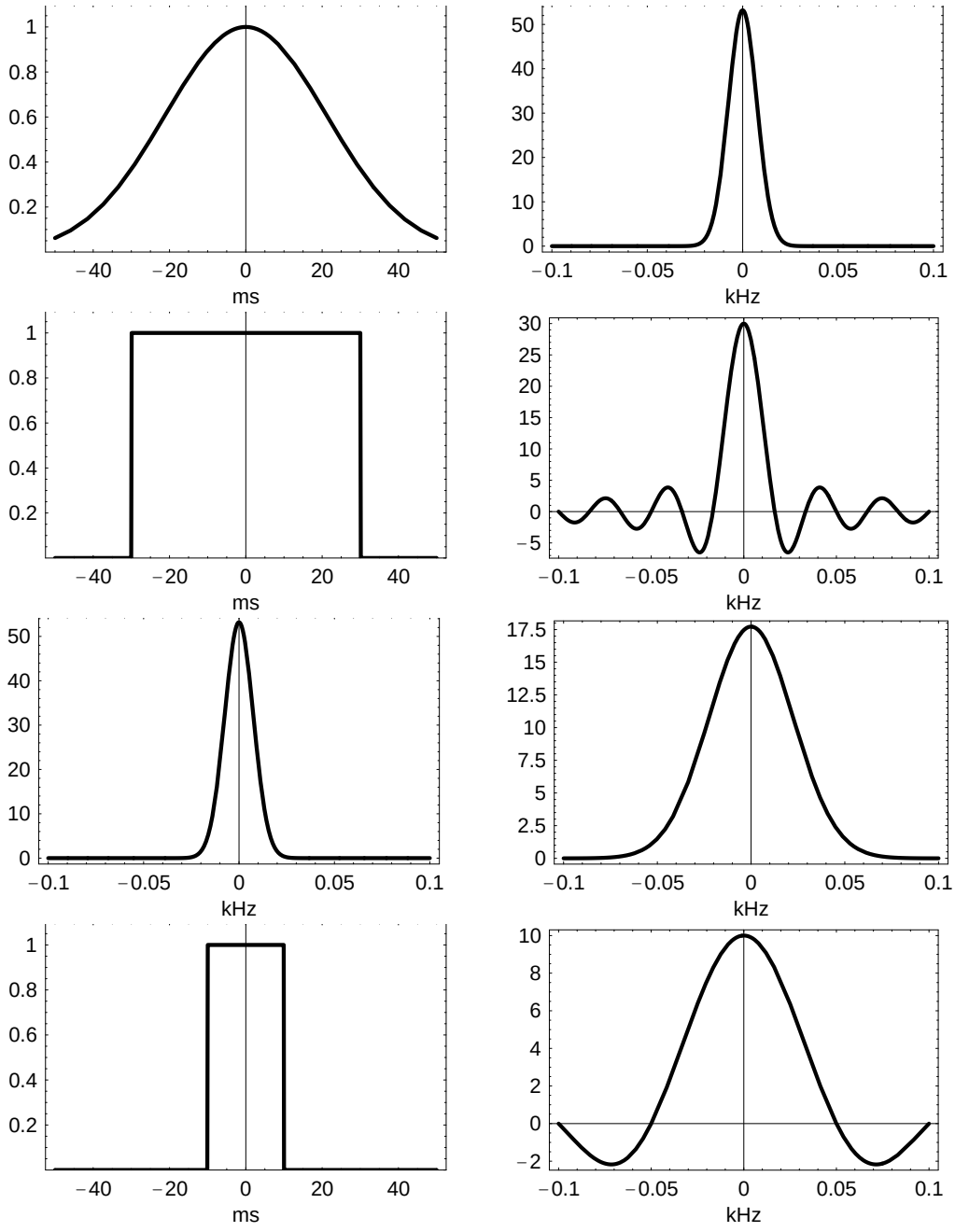
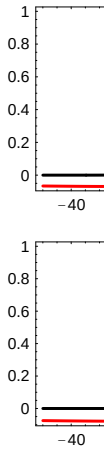


Abbildung 2.16: Left: The signal: Gaussian and square wall. Right: the corresponding spectra (Fourier transforms)



high

Abbildung 2.18: A rectangular filter bank with limits $z_i = \sum_{k=1}^{i}(2b_k)$

Be $\{p_i\}$ the list of numerical values of the signal at i=1,2, ... 1000 ms, then the filtered signal is:

$$p_{f\,i} = \frac{1}{5}\sum_{k=1}^{5} p_{i-k} \tag{2.42}$$

In this way noise can be partially eliminated.

Filter banks . A filter bank is s series of adjacent band pass filters.

The easiest example is a series of neighboring square wall filters with widths $2b_i$, see figure 2.18.

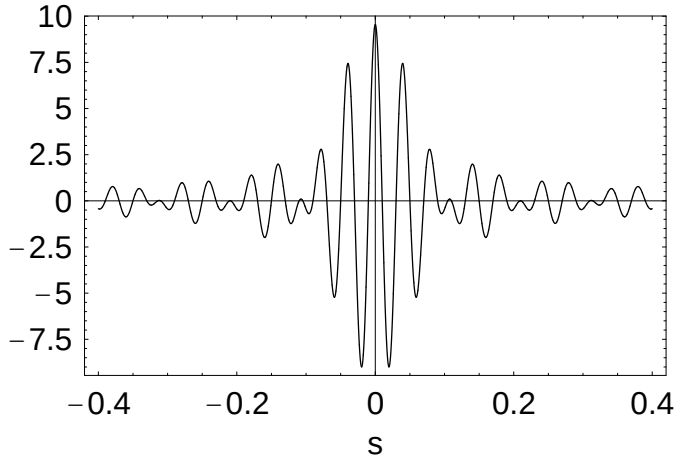


Abbildung 2.19: The time dependent filter function $\phi_i(t)$ of a rectangular band pass filter with central frequency $\nu_i = \frac{\omega_i}{2\pi} = 25$ Hz and $b = 30$ s^{-1}.

In that case the filters $\tilde{\phi}_i(\omega)$ have the form:

$$\tilde{\phi}_i(\omega) = \theta(\omega - z_{i-1}) \cdot \theta(z_i - \omega) \quad \text{with } z_i = \sum_{k=1}^{i} (2b_k) \tag{2.43}$$

and central frequencies $\omega_i = z_i - b_i$. The time dependent filter function is then:

$$\begin{aligned} \phi_i(t) &= \int_{z_{i-1}}^{z_i} \frac{d\omega}{2\pi} = \frac{1}{-2\pi i\, t}\Big(e^{-iz_i t} - e^{-iz_{i-1}t}\Big) \\ &= e^{-i\omega_i t}\Big(e^{-ib_i t} - e^{ib_i t}\Big) \end{aligned}$$

from which follows:

$$\mathrm{Re}\phi_i(t) = 2\cos(\omega_i t)\frac{\sin(bt)}{2\pi t} \tag{2.44}$$

The function $\phi_i(t)$ for $\nu_i = \frac{\omega_i}{2\pi} = 25$ Hz and $b = 30$ sec^{-1} is displayed in figure 2.19.

In fig. 2.20 a bank of Hann pass band filters with constant band width equivalent regular band width (ERB) $2b = 500$ Hz is displayed.

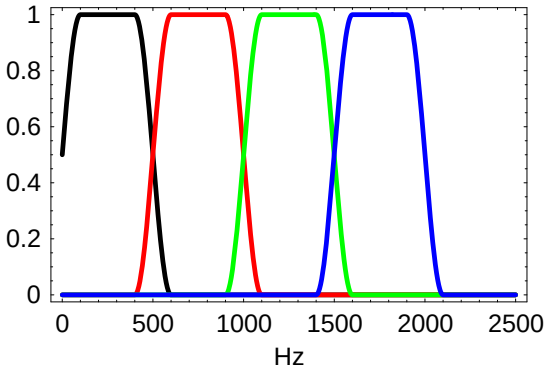


Abbildung 2.20: A bank of Hann band pass filters with equivalent rectangular bank width $2b = 500$ Hz

2.1.8 Fourier transform and filtering on the PC

As mentioned, the Fourier series 2.2 is the basis of numerical Fourier Transformation. The signal $p(t)$ is given by a list of real numbers, $\{p_n\} = \{p_1, p_2, \dots p_N\}$ of length N, the Fourier transform (`Fourier[list]` in Mathematica, `ifft[list]` in Mathlab, is given by

$$\tilde{p}_\ell = \frac{1}{Z} \sum_{n=1}^{N} p_n \, e^{2\pi i(\ell-1)(n-1)/N} \tag{2.45}$$

The back-Fourier (inverse Fourier) transformation is given by:

$$p_n = \frac{Z}{N} \sum_{\ell=1}^{N} \tilde{p}_\ell \, e^{-2\pi i(\ell-1)(n-1)/N} \tag{2.46}$$

The constant Z is chosen as $Z = \sqrt{N}$ in Mathematica, $Z = 1$ in Mathlab.

If the time interval corresponding to the list $\{p_n\}$ is T, that is

$$p_n = p\Big(\frac{(n-1)T}{N}\Big) \tag{2.47}$$

then the Fourier transformed list $\{\tilde{p}_\ell\}$ is the discretization of the Fourier transform $\tilde{p}(\omega)$:

$$\tilde{p}_{\ell+1} = \begin{cases} \frac{N}{TZ}\tilde{p}\left(2\pi\ell\frac{1}{T}\right) & \text{for } 0 < \ell < N/2 \\ \frac{N}{TZ}\tilde{p}^*\left(2\pi(N-\ell)\frac{1}{T}\right) & \text{for } N-1 > \ell > N/2 \end{cases} \tag{2.48}$$

That is: the Fourier transformed list gives the values of the spectrum in frequency steps of $1/T$ up to a maximal frequency of $\frac{N-1}{2\,T}$ (Nyquist):

$$\nu = \frac{\omega}{2\pi} = \left(0, \frac{1}{T}, \frac{2}{T} \ldots \frac{N-1}{2\,T}\right) \tag{2.49}$$

Proof: The signal to be analyzed will in practise always last only for a finite time, say from 0 to T. Then 2.9 reads:

$$\tilde{p}(\omega) = \int_0^T dt\, p(t)\, e^{i\omega t} \tag{2.50}$$

This integral is expressed as a finite sum:

$$\tilde{p}(\omega) \approx \sum_1^N p(T(n-1)/N) e^{i\omega T(n-1)/N} \frac{T}{N} \tag{2.51}$$

Comparison with 2.45 yields the relation 2.48 for $\ell < N/2$

Exercise: Show the relation for $\ell > N/2$. Hint: $e^{2\pi i k} = 1$ for any integer k.

Filtering Be $\tilde{\phi}_\ell$ a discretization of the filter $\tilde{\phi}(\omega)$, that is for positive values of ω:

$$\tilde{\phi}_{\ell+1} = \tilde{\phi}\left(2\pi\ell\frac{1}{T}\right) \tag{2.52}$$

Then the filtered signal can be expressed as, see 2.13

$$\begin{aligned} (p_f)_n &= \sum_{\ell=1}^{N} (\tilde{\phi}_\ell\, \tilde{p}_\ell\, e^{-2\pi i(\ell-1)(n-1)/N}) && (2.53)\\ &= \sum_{\ell=1}^{N/2} 2\,\mathrm{Re}\,(\tilde{\phi}_\ell\, \tilde{p}_\ell\, e^{-2\pi i(\ell-1)(n-1)/N}) && (2.54) \end{aligned}$$

If the time dependent filter function $\phi(t)$ is given with the discretization:

$$\{\phi_m\} = \{\phi(0), \phi(1/T) \ldots \phi((M-1)/T)\}$$

then the filtered signal is given by:

$$(p_f)_m = \frac{T}{M} \sum_{n=1}^{M} \phi_n\, p_{m-n} \tag{2.55}$$

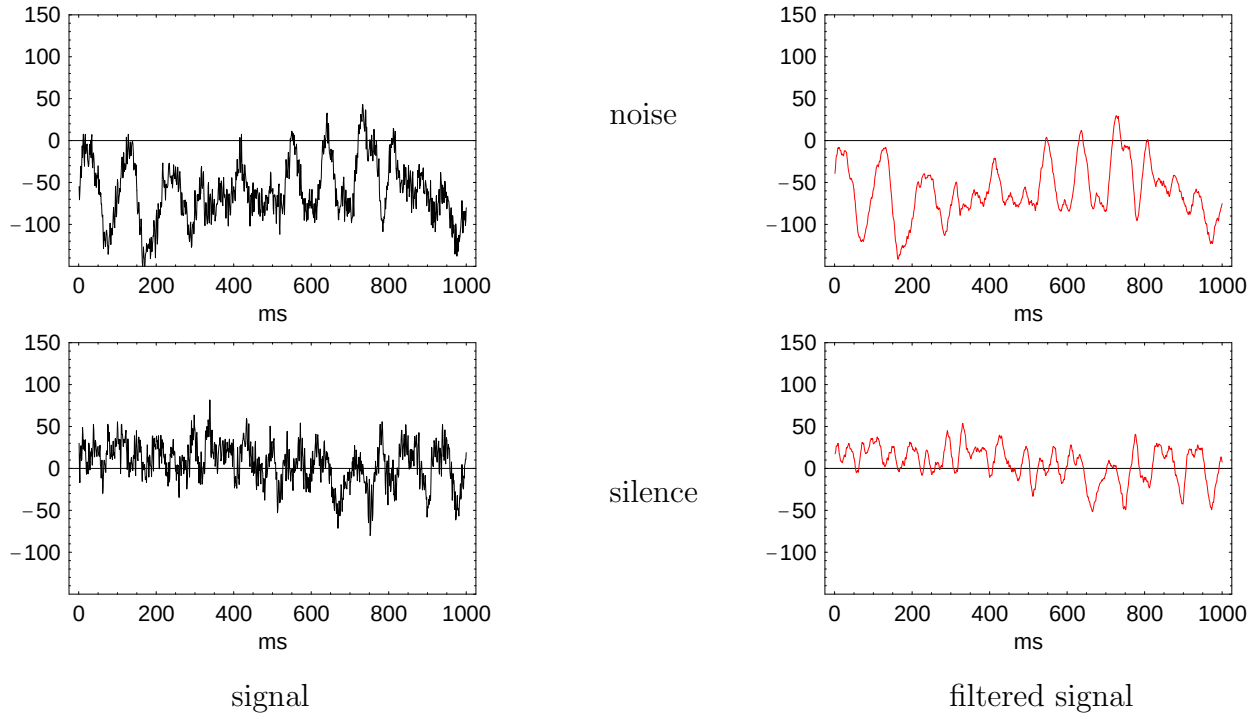


Abbildung 2.21: The original activity (black) and filtered (smoothed)(red) activity of the auditory cortex when the individual is exposed to a homogeneous noise or to silence

with the conventions
$p_n = 0$ for $n < 1$ and $n > N$ and $\phi_m = 0$ for $m < 1$ and $m > M$

The performance of the Fourier transform using the program Mathematica is given in `demofourier.nb` on the home page of the lecture www.thphys.uni-heidelberg.de/$\tilde{}$dosch/vorl10_11.

Application: Noise in the brain We shall illustrate the filtering by convolution on a neurophysiological example. In fig. 2.30 the activity of the auditory cortex, which is not related to a special event, is recorded. In the first row the individual is exposed to a continuous noise, in the second row there was no stimulation at all (silence). In both cases a period of 1000 ms is reported. The left hand side shows the original measurements, the right hand shows the signal filtered by averaging (convoluting) over 10 ms. A periodic component with a frequency of about 10 Hz is visible.

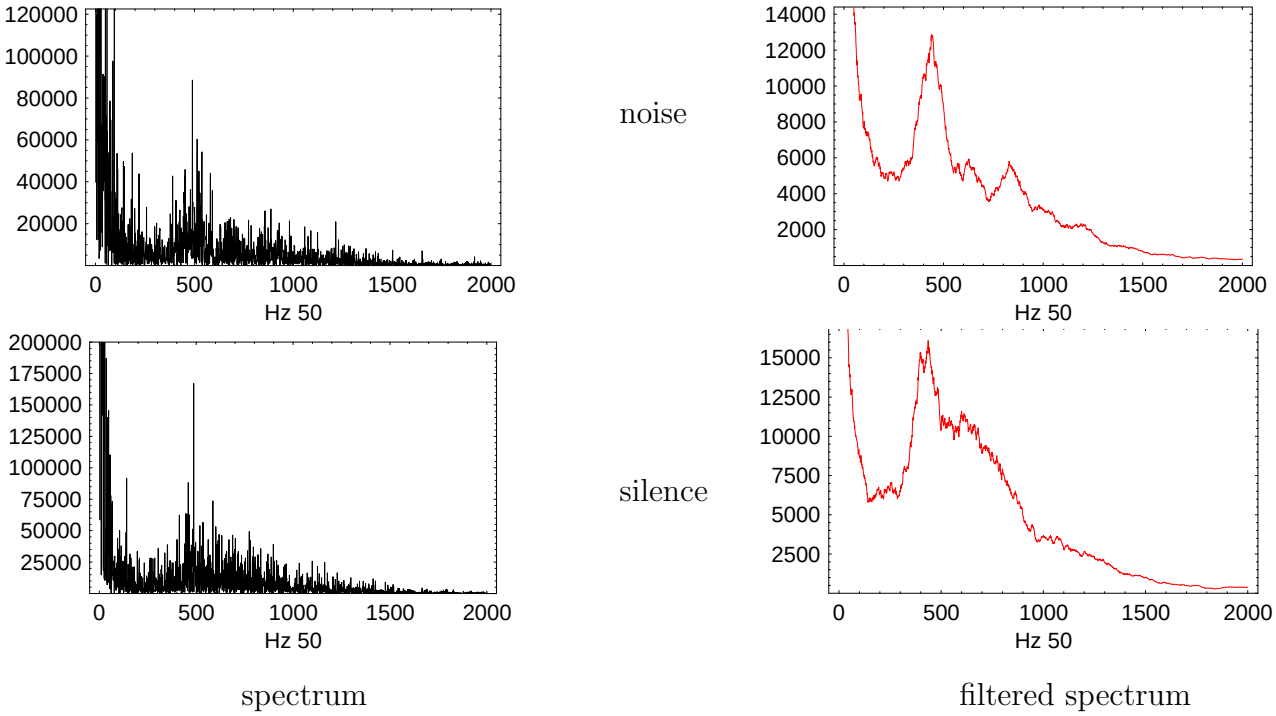


Abbildung 2.22: The original power spectrum(black) and filtered (smoothed)(red) power spectrum of the activity of the auditory cortex when the individual is exposed to a homogeneous noise or to silence. The time interval of the Fourier integration is 50 s.

In fig. 2.31 the power spectra, obtained from a Fourier series over an integral of 50 s (corresponding to 50 000 data points) are displayed. Since the time interval over which was integrated in the Fourier transform is 50 sec, the spectrum is in steps of 1/50 Hz. Again to the left the original spectrum is shown (cut below 2 Hz), on the right the spectrum is filtered by averaging over ±50 data points, corresponding to convolution interval of ±1 Hz.

In both conditions a band of signals around a frequency of 10 Hz is clearly visible.

2.1.9 The Hilbert envelope

The Hilbert envelope is the result of a special filtering, namely one which does not lead to a real filtered signal, that is the filter $\tilde{\phi}(\omega)$ does not meet the requirement:

$$\tilde{\phi}(-\omega) = \tilde{\phi}^*(\omega)$$

. The filter leading to the Hilbert envelope has the form:

$$\tilde{\phi}(\omega) = \theta(\omega) \tag{2.56}$$

the filtered signal of $p(t)$ with spectrum $\tilde{p}(\omega)$ is thus:

$$p_f(t) = \int_0^\infty \frac{d\omega}{2\pi}\, \tilde{p}(\omega) \tag{2.57}$$

The *Hilbert envelope* is defined as:

$$p_{HE}(t) = |2p_f(t)| \tag{2.58}$$

One can show, using integration in the complex plane (for details see the file `hilbert.pdf`, that

$$p_f(t) = p(t) - \frac{i}{\pi}\mathrm{P}\int_0^\infty dt'\, \frac{p(t')}{t-t'} \tag{2.59}$$

If you do not know the definition of the principal value integral, keep to the definition 2.57. Just to explain the name , the principal value on the rhs of **??** is the *Hilbert transform* of $p(t)$. For the important special case of a slowly modulated complex tone

$$p(t) = \theta(t)\, H(t) \sum_i\, c_i \cos(\omega_i t + \delta_i) \tag{2.60}$$

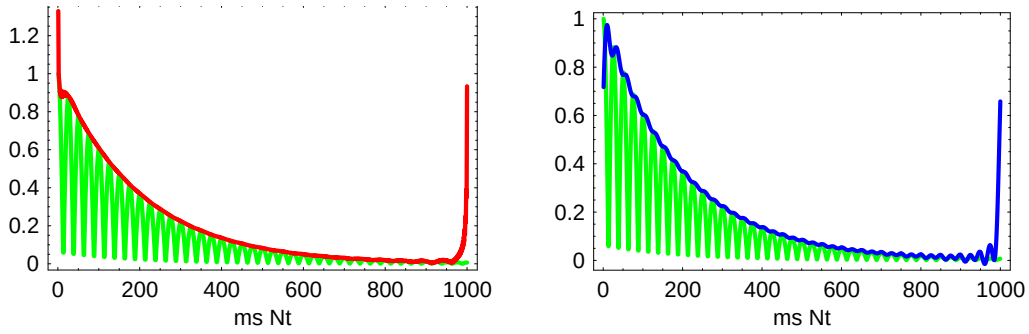


Abbildung 2.23: Left: Absolute value of the signal, $p(t) = \exp[-t\,\lambda]\cos(2\pi\,\nu\,t)$ (green),its Hilbert envelope (red, left), and its low pass filtered (40 Hz) Hilbert envelope (blue,right). $\lambda = 5,\ \nu = 20$.

where the enveloping function H is slowly varying with respect to the oscillations $2\pi\nu_i$, one obtains for the Hilbert envelope:

$$|2p_a(t)| \approx \left|\sum_j c_j e^{-i\,\delta_j - \omega_j\,t}\right| |H(t)|. \tag{2.61}$$

If the signal $p(t)$ contains only a single frequency or only a narrow range of frequencies, the Hilbert envelope is approximately equal to the enveloping function H. If however the signal contains many components with different frequencies, the interference terms in 2.61 make the modulus factor in front of H vary strongly.

Some examples are displayed in figures 2.23 - **??**. In Fig. 2.23 an exponentially damped sinusoidal tone $p(t) = \exp[-t\,\lambda]\cos(2\pi\,\nu\,t)$ is shown. Apart from the overshooting near $t = 0$ and at the end the Hilbert envelope describes very well the physical envelope, the exponential function. The overshooting is due to a singularity of the principal value at $t = 0$ and at the end, since the signal starts and ends with a value unequal zero. By low pass filtering of the Hilbert envelope this effect vanishes.

In figure 2.24 the signal is $p(t) = \exp[-t\,\lambda]\Big(\sin(2\pi\,\nu\,t) + \sin(4\pi\,\nu\,t)\Big)$. It starts with zero, but the interference between the two sinusoidal tones (see 2.61) leads to a oscillations in the Hilbert envelope and it is far from the representation of the slowly varying envelope. These oscillations are reflected in the spectrum of the Hilbert envelope, fig. 2.24, right). By low pass filtering, with a cutoff frequency of 20 Hz, this structure can be removed and the filtered Hilbert envelope is a fair description for the physical envelope.

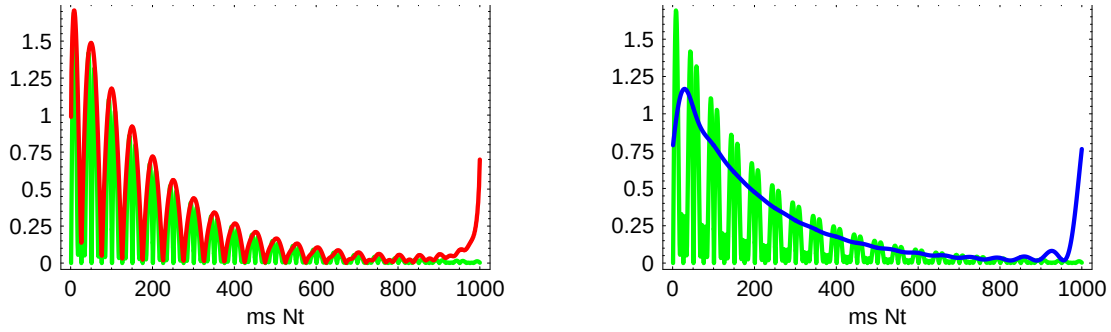


Abbildung 2.24: Absolute value of the signal $p(t) = \exp[-t\,\lambda]\Big(\sin(2\pi\,\nu\,t) + \sin(4\pi\,\nu\,t)\Big)$ (green),its Hilbert envelope (red,left), and its low pass filtered (15 Hz) Hilbert envelope (blue,right). $\lambda = 5,\ \nu = 20$. Nt=1

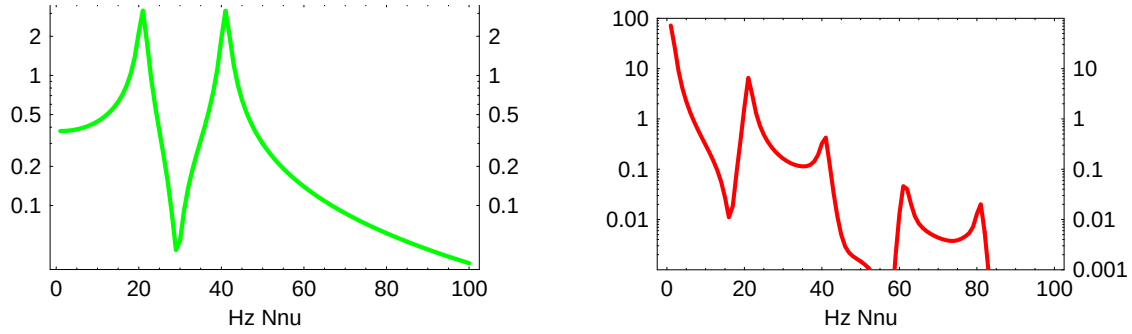


Abbildung 2.25: Spectrum of the signal (green, left) and of the Hilbert envelope (red,right), Nnu=1.

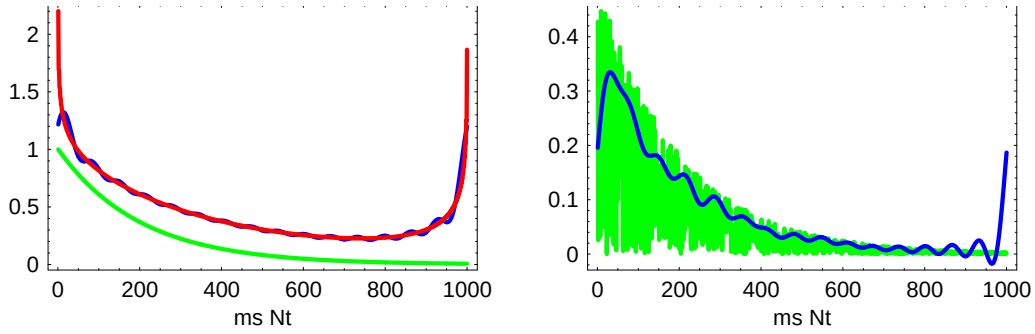


Abbildung 2.26: Left: Absolute value of the signal $|p(t)| = |\exp[-t\,\lambda]|$ (green),its Hilbert envelope (red), and its low pass filtered (15 Hz) Hilbert envelope (blue). $\lambda = 5$. Right: An exponentially damped random noise (green) and its low passed filtered (20 Hz) Hilbert envelope (blue).

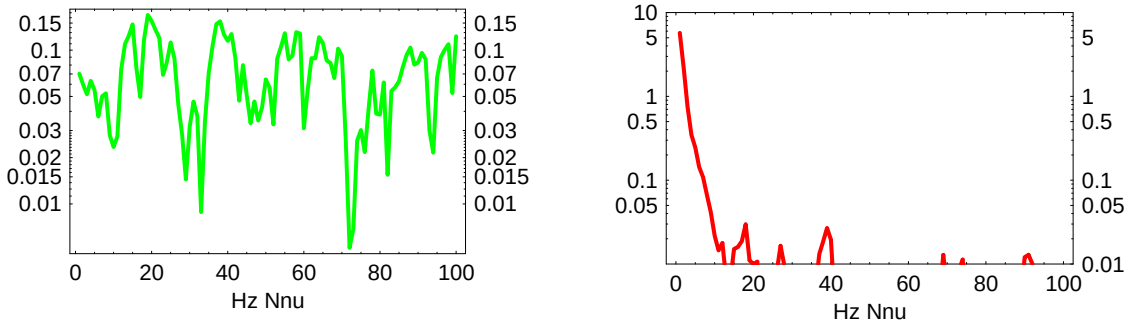


Abbildung 2.27: Spectrum of the signal of 2.26(green, left) and of its Hilbert envelope (red,right), Nnu=1.

For non-oscillating functions the Hilbert envelope is not used and its application would be misleading, as can be seen from fig. 2.26, left. The envelope of an exponentially damped random noise is quite well reproduced by its low pass filtered (20 Hz) Hilbert envelope. A smoother filtering would improve the result.

The filtered Hilbert envelope is most useful for following the envelope e.g of human speech. In fig. 2.28 a speech signal, the word "hallo?" processed. Finally I give a realistic example. The sound "hallo ?" had been recorded, changed into a text file, and then Hilbert envelope and the filtered Hilbert envelope had been calculated using Fourier transforms. The length of the signal is approximately one second and contains 9634 data points.

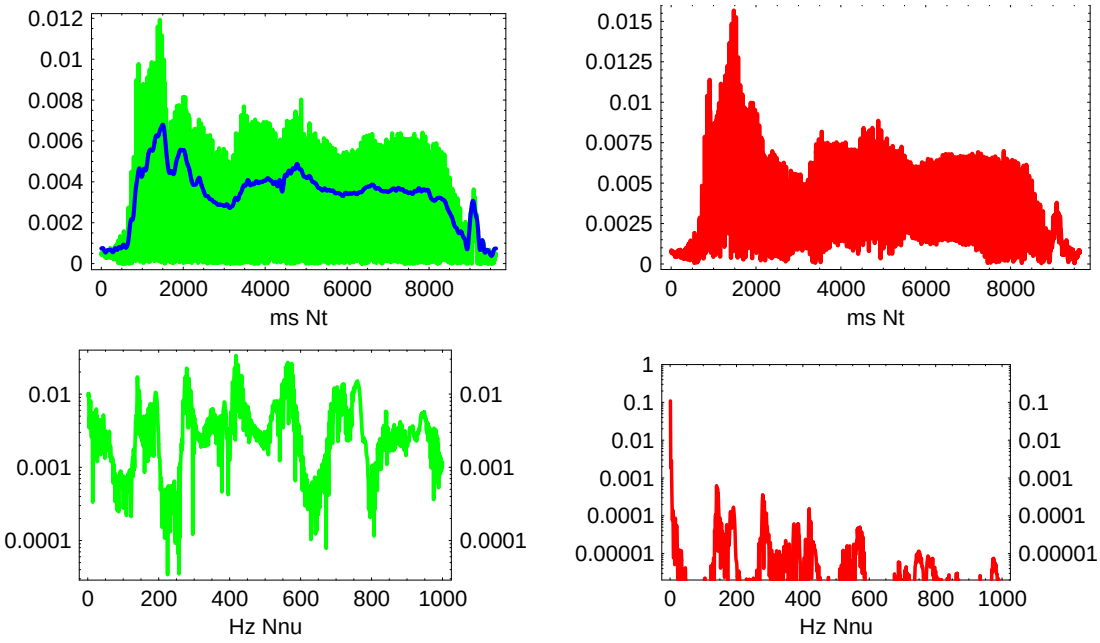


Abbildung 2.28: Top: Left: The absolute value of the speech signal “hallo?” (green) and its low pass filtered (100 Hz) Hilbert envelope (blue) Right: the unfiltered Hilbert transform of the speech sinal. Bottom: Spectrum of the speech signal (green, left) and of its Hilbert envelope (red, right).

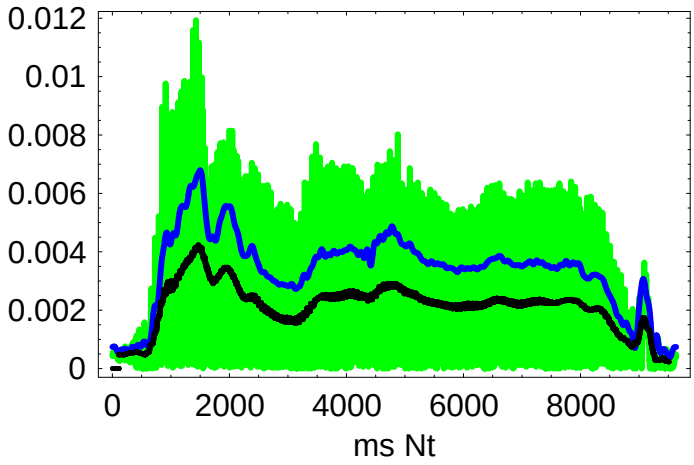


Abbildung 2.29: The speech signal (green), the low passed Hilbert transform (blue) and a curve obtained by averaging the absolute value of the signal over ±10 ms (black).

One can also easily obtain the envelope by filtering in the time region, for instance by averaging over a time interval. In fig. 2.29 the time averaged (20 ms), corresponding to 200 data points is diplayed (black), together with the absolute value of the signal (green) and the low pass filtered (100 Hz) Hilbert envelope.

In the file `demohallo.nb` a demonstration for the calculation of a Hilbert envelope using mathematica is given.

Noise in the brain We shall illustrate the filtering by convolution on a neurophysiological example. In fig. 2.30 the activity of the auditory cortex, which is not related to a special event, is recorded. In the first row the individual is exposed to a continuous noise, in the second row there was no stimulation at all (silence). In both cases a period of 1000 ms is reported. The left hand side shows the original measurements, the right hand shows the signal filtered by averaging (convoluting) over 10 ms. A periodic component with a frequency of about 10 Hz is visible.

In fig. 2.31 the power spectra are displayed. The time interval over which was integrated in the Fourier transform is 50 sec, that is the spectrum is in steps of 1/50 Hz. Again to the left the original spectrum is shown (cut below 2 Hz), on the right the spectrum is filtered by averaging over ±50 data points, corresponding to ±1 Hz.

In both conditions a band of signals around a frequency of 10 Hz is clearly visible.

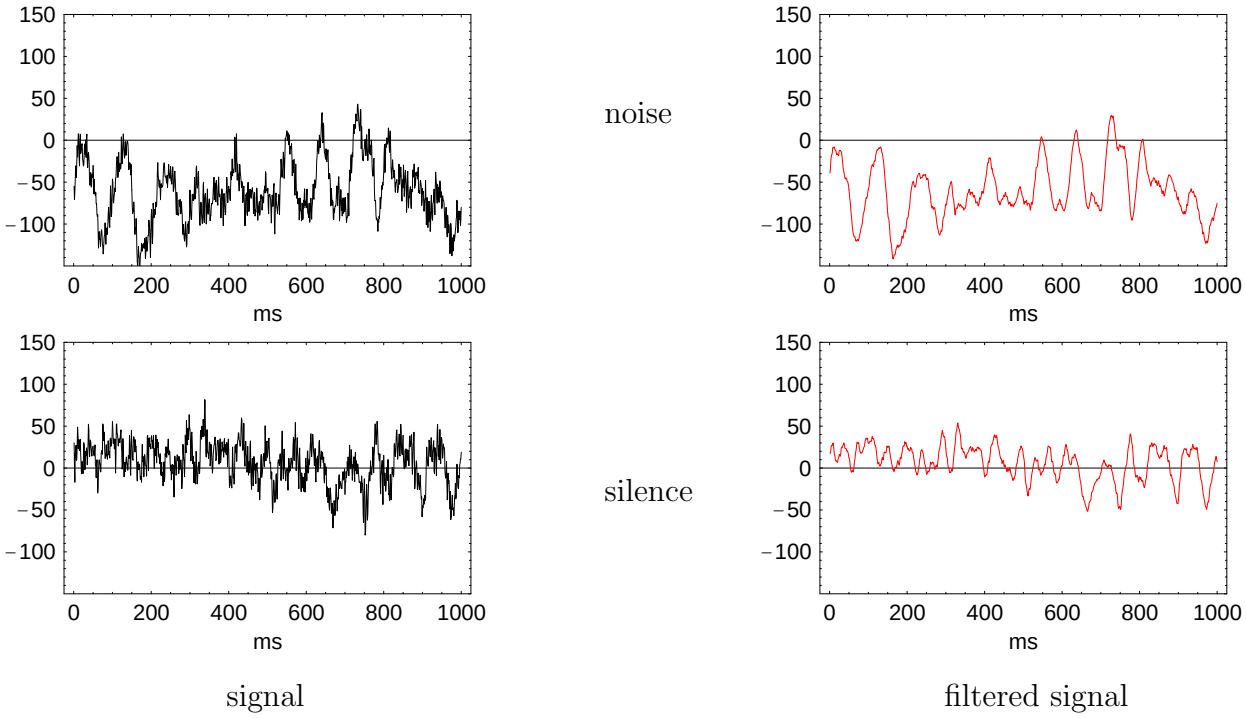


Abbildung 2.30: The original activity (black) and filtered (smoothed)(red) activity of the auditory cortex when the individual is exposed to a homogeneous noise or to silence

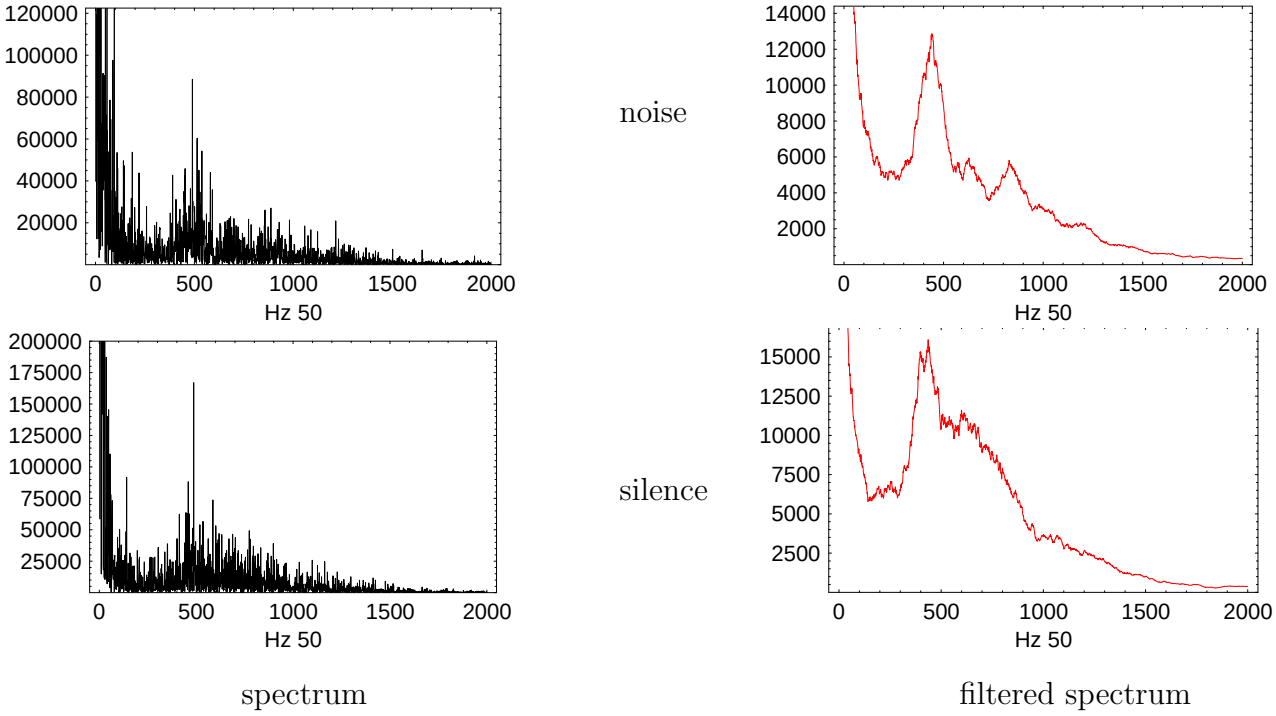


Abbildung 2.31: The original power spectrum(black) and filtered (smoothed)(red) power spectrum of the activity of the auditory cortex when the individual is exposed to a homogeneous noise or to silence2

Kapitel 3

Physics of the ear

3.1 Einige psychophysische Grundlagen

In der Psychophysik, hier speziell der Psychoakustik, werden Sinneseindrücke mit physikalischen Grössen korreliert.

3.1.1 Tonhöhe

Die vielleicht am längsten bekannte psychophysische Tatsache ist die Korrelation von Tonhöhe mit Frequenz des Tones. Schon die jüngeren Pythagoräer vermuteten einen Zusammenhang zwischen Frequenz und Tonhöhe, quantitativ wurde dieser im 17. Jahrhundert (Mersenne, Harmonie Universelle, 1637; Sauveur) bestimmt. In Abb. 3.1 ist der Frequanzbereich des menschlichen Gehörs dargestellt. Die Empfindlichkeit ist bei einigen Kiloherz am grössten, hier spielt auch die Verstärkung durch das Mittelohr eine Rolle. Im Alter wird die Hörschwelle bei hohen Frequenzen progressive höher, wie aus Abb. 3.1, rechts zu sehen.

In Abb. 3.3 ist der Tonumfang auf einer Klaviertastatur mit den Frequenzen der einzelnen Töne dargestellt. Helmholtz hat den im Vergleich zum Auge sehr viel höheren Frequnzumfang an einer eigenhändigen Zeichnung Abb. 3.2 dargestellt.

Allerdings ist der Zusammenhang zwischen Tonhöhe und Frequenz nicht so ganz einfach, wie an folgenden Tonbeispielen zu erkennen ist. Tonbeispiel `praats/pitch-sensation`. In Abb. 3.4 ist sowohl der zeitliche Verlauf als auch das (Energie-)Spektrum dieser Töne dargestellt. Was die Tonhöhe bestimmt, ist immer noch nicht vollständig geklärt, wahrscheinlich spielen

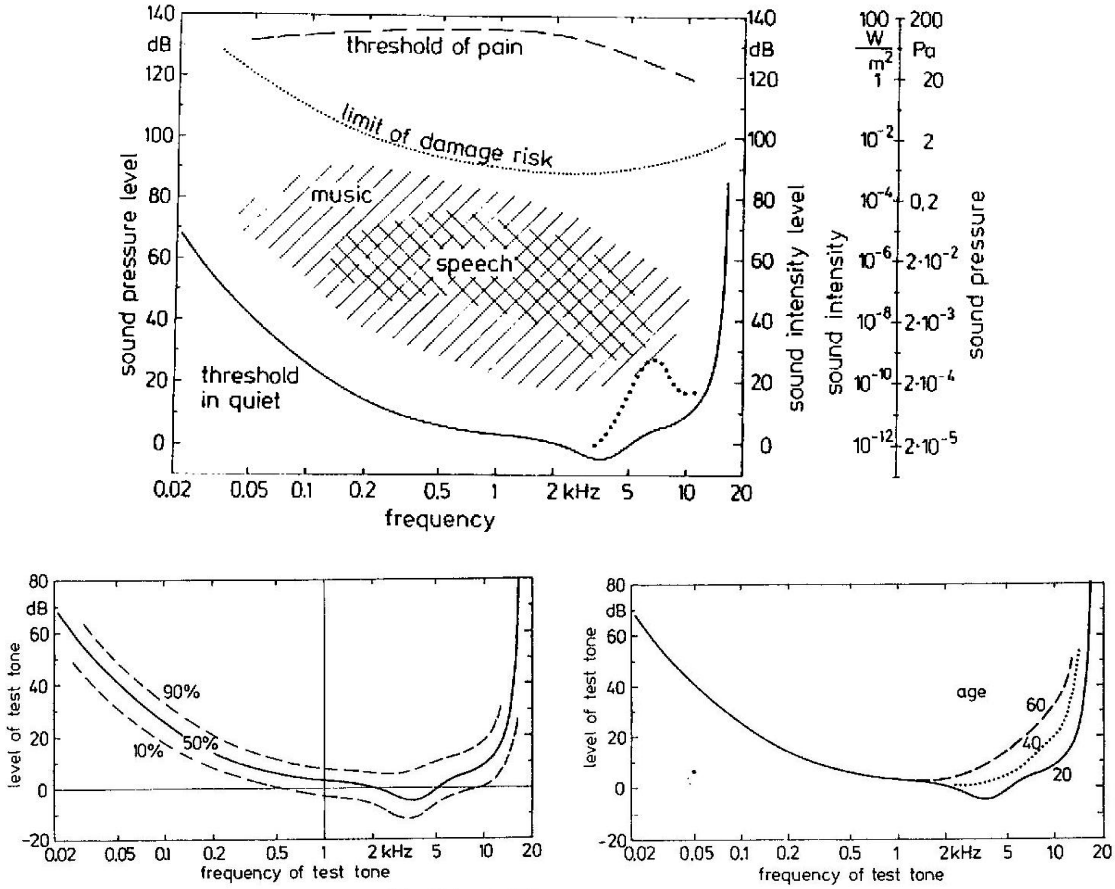


Abbildung 3.1: Hörschwelle des Menschen, aus [**?**] **Links unten:** Indivduelle Schwankungen. **Rechts unten:** Altersabhängigkeit,

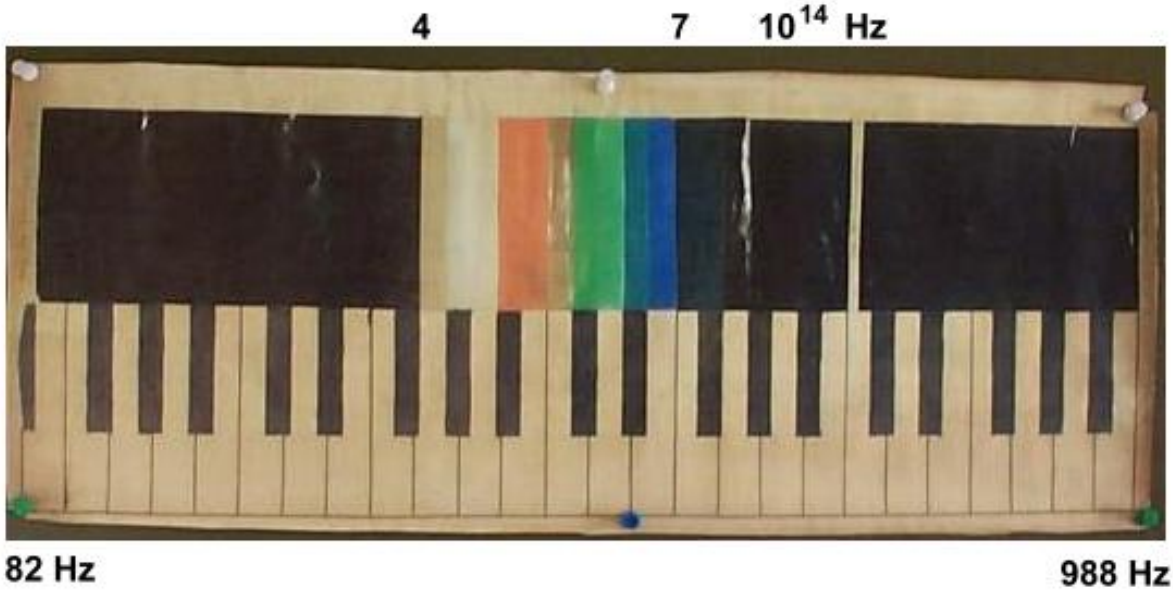


Abbildung 3.2: Farbklavier. Eigenhändige Zeichnung von H. Helmholtz, um den weit grösseren Frequenzbereich des Ohres im Vergleich mit dem Auge klarzumachen; der dargestellte Tonumfang entspricht etwa dem der menschlichen Stimme

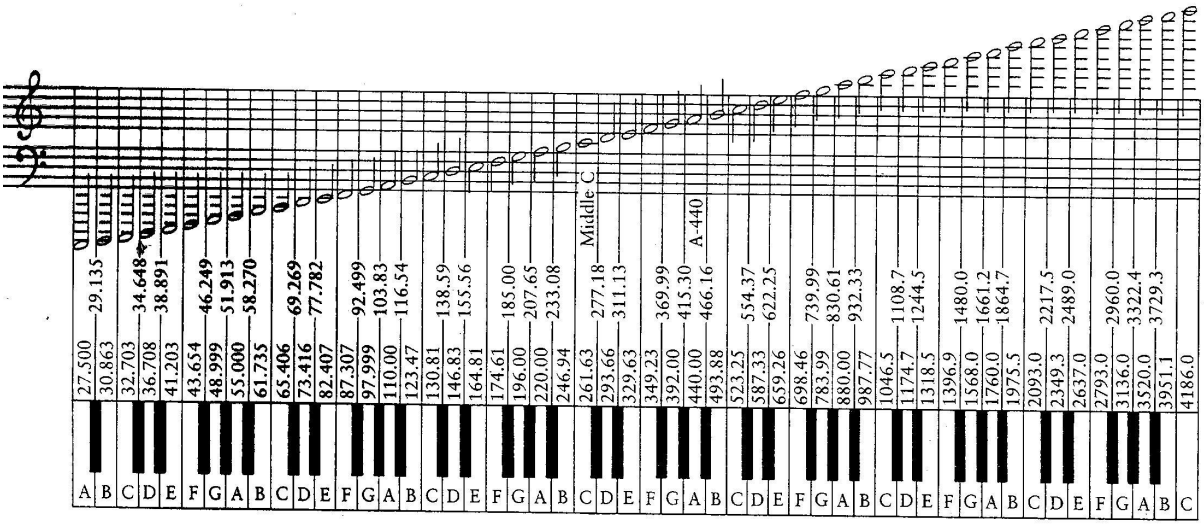


Abbildung 3.3: Musikalisch benutzter Tonumfang

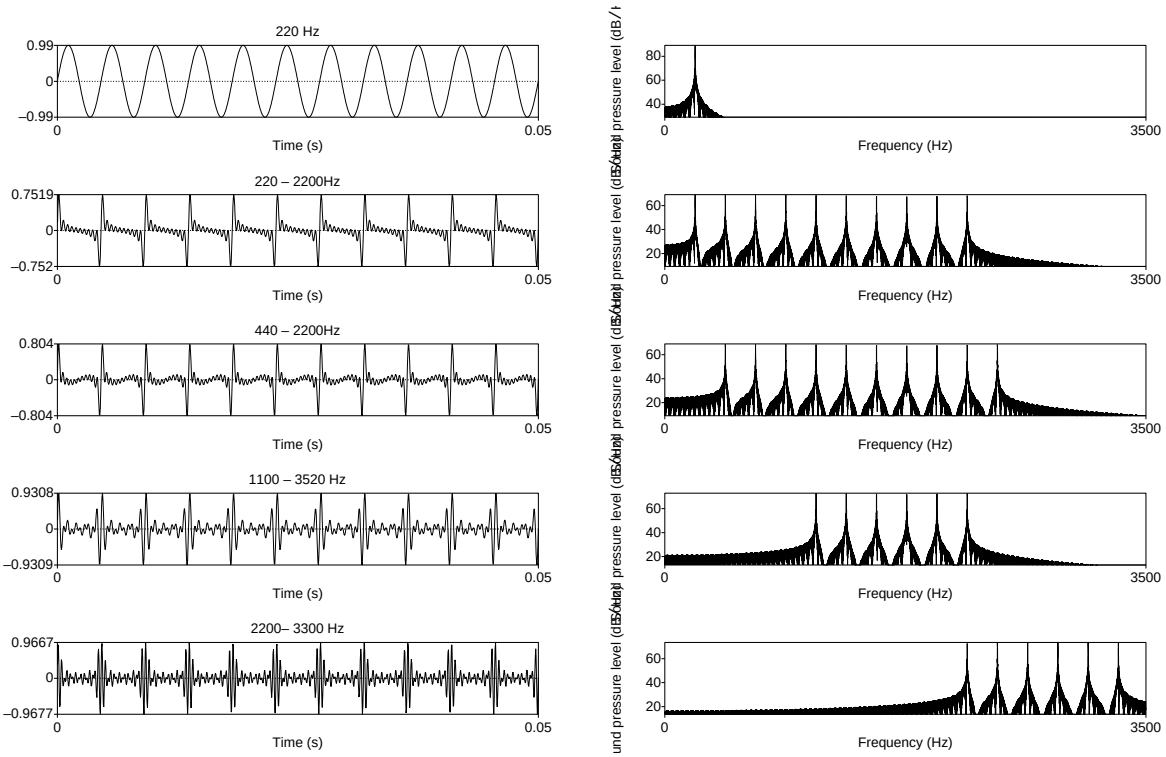


Abbildung 3.4: Zeitverlauf und Spektren der in Bsp. `praats/pitch-sensation` abgespielten Töne

sowohl das Spektrum, das sich ja direkt auf der Cochlea abbildet, als auch die Periodizität eine entscheidende Rolle.

Wenn wir zwei Töne hintereinander hören, nehmen wir schon Tonhöhenunterschiede von weniger als einem Prozent wahr (1/2-Ton $\sim 6\%$). Bsp. `praats/zweiton`, 880 und 900 Hz. Wenn wir diese beiden Frequenzen zusammenhören, nehmen wir dagegen nur **einen** rauhen Ton war. Damit wir zwei Töne deutlich getrennt hören, muss das Verhältnis der Frequenzen schon etwa 1.2 (etwa kleine Terz) sein. Wir werden später sehen, dass dies eng mit der Mechanik der Cochlea zusammenhängt. Erst wenn die Anregungen durch die zwei verschiedenen Frequenzen auf der BM weit genug voneinander entfernt sind (ca 1 mm), dann hört man zwei getrennte Töne.

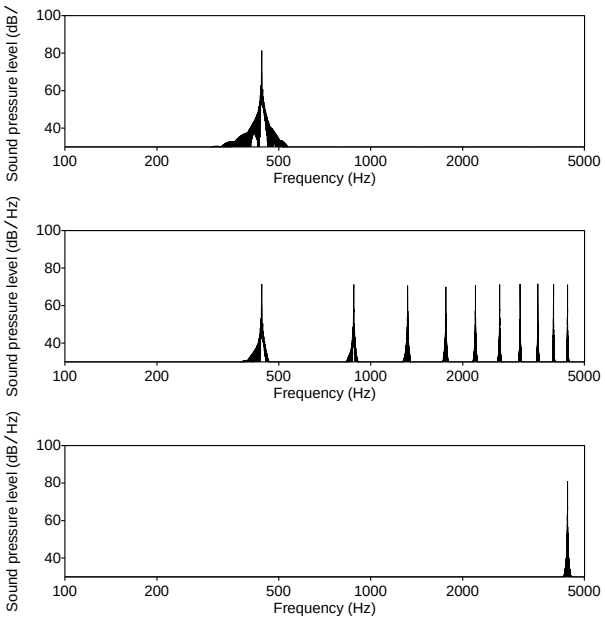


Abbildung 3.5: Verschieden Töne gleicher Energie aber verschiedener Lautheit. `praats/loudeness`

3.1.2 Lautstärke

Die empfundene Lautstärke ist ungefähr proportional dem Logarithmus der Amplitude des Signals. Deshalb wird i. A. die Lautstärke in Einheiten von deciBell (dB) angegeben: Zwanzigmal der dekadische Logarithmus des Verhältnisses der Amplitude des Schalldruckes des Signals zur Hörschwelle ist die Lautstärke in dB. Der Faktor 20 rührt daher, dass der Energiefluss proportional dem Quadrat der Schalldruckkurve ist.

Allerdings sind auch hier die Verhältnisse, wie bei der Tonhöhe, komplizierter, es kommt ganz wesentlich auf die spektrale Verteilung an, Bsp. `praats/cb-noise`. Im Beispiel ist der Energiefluss immer der gleiche, aber die spektrale Breite des Rauschsignals wächst an, wenn die spektrale Breite einen gewissen Wert überschreitet, dann wird das Signal, bei gleichem Energiefluss, lauter. Viel eindrucksvoller ist das Beispiel, bei dem ein sinus-Ton mit einem harmonischen Ton gleicher Energie verglichen wird. Bsp. `praats-loudeness`, s. Abb. 3.5. Dieser Effekt, da daher rührt, dass eine grössere Region der BM angeregt wird, wird beim Orchstercrescendo ausgenützt.

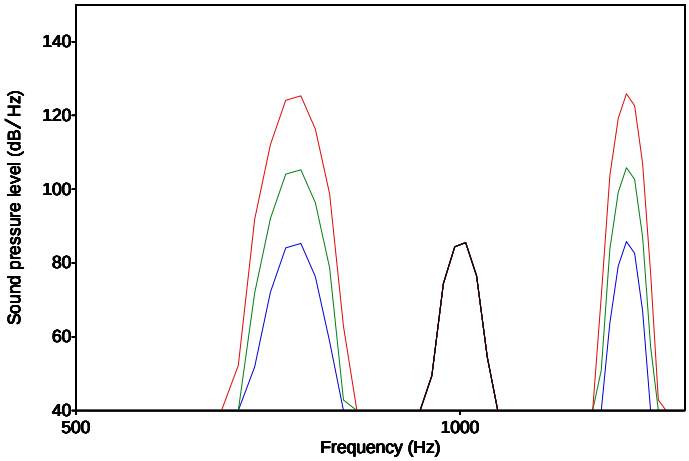


Abbildung 3.6: Prinzip des Maskierungs-Experimentes, ein Testton (schwarz) sowie Maskierer tieferer **oder** höherer Frequenz werden gleichzeitig dargeboten. Die Versuchsperson hat zu entscheiden, wann sie den Testton nicht mehr hört.

3.1.3 Maskierung

Die Nichtlinearität bei der Lautstärkenwahrnehmung führt zur Maskierung: Ein lauter Ton kann einen leiseren Ton benachbarter Frequenz vollständig unhörbar machen. Dies wird z. B. bei MP3 ausgenutzt. Die Makierung spielt aber auch in der Psychophysik eine entscheidende Rolle. Wichtig sind die sogenannten ,,Psychoakustischen Tuning-Kurven". Man hat einen Testton bestimmter Lautstärke uns sucht, ab welcher Lautstärke ein Ton benachbarter Frequenz ihn maskiert, höre Bsp `praats/maskierung3`. Das Prinzip ist in Abb. 3.6 dargestellt. Der Testton (schwarz) wird jeweils mit Maskierungstönen verschiedener Intensität und Frequenz gleichzeitig angeboten. Das (wohl ingenieursmässig etwas geschönte) Resultat solcher Untersuchungen ist in Abb. 3.7 dargestellt. Man sieht, dass die Kurve asymmetrisch ist. Maskierer von tieferer Frequenz als der Testton sind effektiver als solche von höherer Frequenz. Ausserdem wird die Kurve bei höheren Schallpegeln deutlich flacher.

In diesem Abschnitt beschreiben wir den anatomischen Aufbau des Ohrs und schildern die wichtigsten Funktionen.

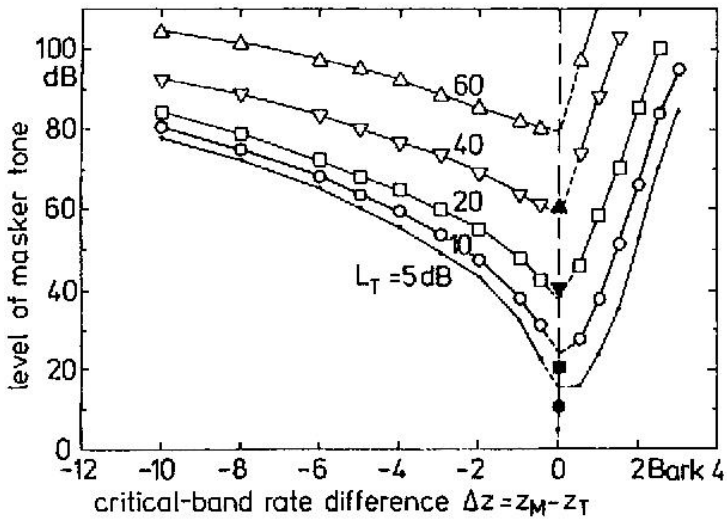


Abbildung 3.7: Psychoakustische Tuning-Kurve. Bei einer Frequenz des Testtons von 1000 Hz enspricht ein Bark etwa einem Frequenzverhältnis von 1.16. Die ausgefüllten Symbole geben die Lautstärke des jeweiligen Testtons an. Man beachte die Asymmetrie und den flacheren Verlauf bei hohen Schallpegeln, aus [**?**].

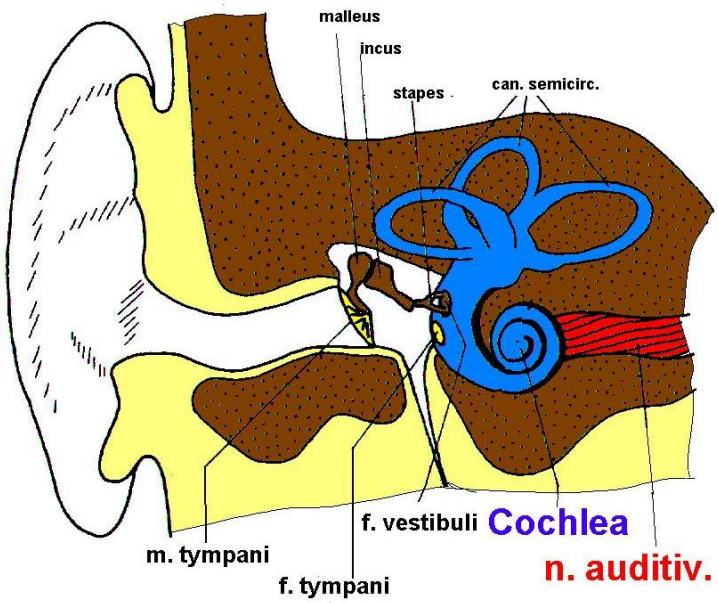


Abbildung 3.8: Das menschliche Ohr

3.2 Anatomie des Ohres

[?, ?, ?] Man teilt nach Funktion und Anatomie das menschliche Hörorgan in 3 Teile, die in der Tabelle

3.2.1 Das Aussenohr

Das Aussenohr ist der Teil von der Ohrmuschel bis zum Trommelfell (*membrana tympani*). Die Ohrmuschel wirkt als Schalltrichter, z. B. zur Differenzierung zwischen vorn und hinten. Der Hörgang als Kanal von ca 2.5 cm Länge und ca 0.75 cm Durchmesser, ein Ende offen, ein Ende (fast) geschlossen hat eine Resonanz bei $\lambda/4 = 2.5$ cm, d.h. bei einer Frequaenz $\nu = c_L/\lambda \approx 3300$ Hz. Deshalb leitet er bei dieser Frequez den Schall besonders gut und hier ist bei Normalhörern die Hörschwelle am niedrigsten. In diesem Bereich gibt es daher viele

	Lage	Funktion
Aussenohr	Ohrmuschel bis Trommelfell	Schalltrichter, Hilfe bei Richtungshören, Schalleiter
Mittelohr	vom Trommelfell bis zum ovalen Fenster der Cochlea	Impedanzausgleich, Schall-Leiter
Innenohr	Cochlea (Schnecke) im Felsenbein des Schädels, gefüllt mit Perilymphe (Wasser).	Umwandlung der Schallwellen in Nervenimpulse

Tabelle 3.1: Die Teile des Ohres und ihre hauptsächlichen Funktionen

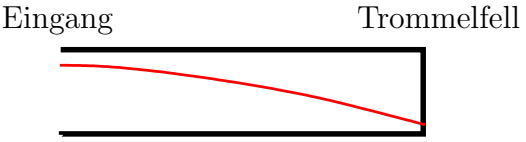


Abbildung 3.9: Die $\lambda/4$-Resonanz im Aussenohr.

Hörschäden (Disko-Ohr).

3.2.2 Das Mittelohr

Das Mittelohr reicht vom Trommelfell bis zum ovalen Fenster (*fenestra tympani*) der Cochlea. Die Schallwellen bewegen das Trommelfell (*m. tympani*), die Bewegung des Trommelfells wird durch die drei Gehörknöchelschen Hammer (*malleus*), Amboss (*incus*), Steigbügel (*stapes*) auf das **ovale Fenster** (*f. tympani*) der Cochlea übertragen. Die Fläche des Trommelfells (ca 0.5 cm^2) ist etwa einen Faktor 20-30 mal grösser als die des ovalen Fensters , dies erhöht den Druck um einen Faktor 20-30, die Gehörknöchelschen haben eine Hebelwirkung, die die Geschwindigkeit etwa um den Faktor 3 verringert.

Dies führt zu einem **Impedanzausgleich**. Die Moleküle in Luft bewegen sich schnell bei geringen Druck, in Wasser langsam bei hohem Druck. Daher werden Schallschwingungen in der Luft von einer Wasseroberfläche fast vollständig reflektiert. Der Reflexions- bzw. Transmissionskoeffizient ist bestimmt durch Impedanzunterschied. Die Impedanz ist definiert als:
R= spezifische akustische Impedanz = Druck/Teilchengeschwindigkeit.

Für ebene Wellen gilt: $R = \rho c$, Dichte mal Schallgeschwindigkeit.

Für Luft gilt: $R_{\text{Luft}} = 43$ g cm^{-2}s^{-1}

Für Wasser: $R_{\text{H}_2\text{O}} = 1.510^5$ g cm^{-2}s^{-1}

Wir werden spaäter sehen, dass die für das Hören relevanten Wellen im der Cochlea (Innenohr) keine Schall-, sondern Wanderwellen sind, so dass die Impedanz etwa einen Faktor 10 kleiner ist, d.h. $R_{\text{Cochlea}}/R_{\text{Luft}} \approx 360$, durch Flächenunterschied und Hebelwirkung wird im Mittelohr die Impedanz um etwa einen Faktor 60-90 erhöht, es findet also kein vollständiger Impedanzausgleich statt, aber doch eine erhebliche Impedanzangleichung.

3.2.3 Das Innenohr, die Cochlea

Der entscheidende Teil des Ohres ist das Innenohr, in dem die Schallschwingungen in Nervenimpulse umgewandelt werden. Wir werden uns damit ausfürlich im nächsten Abschnitt beschäftigen, und hier nur den antomischen Aufbau und einige mechanische Eigenschaften beschreiben.

Das Innenohr besteht aus der Cochlea (Schnecke) und den drei Gängen des Gleichgewichtorgans. Sie sind anatomisch verbunden, aber funktionell vollkommen getrennt.

Die Cochlea besteht aus zweieinhalb Windungen, aufgerollt ist sie etwa 3.5 cm lang, ihr Querdurchmesser ist von der Grössenordnung 2 mm. In Abb. 3.10 ist ein Querschnitt schematisch dargestellt.

Der Gang der Schnecke ist mit Lymphe (hydrodynamisch Wasser) angefüllt, und durch die **Basilarmembran** in 2 Kammern (Skalen) getrennt: Die *scala vestibuli*, die mit dem ovalen Fenster beginnt und am Apex der Schnecke endet, und die *scala tympani*, die durch das elastische runde Fenster mit dem Mittelohr verbunden ist. Die beiden Skalen sind am Apex durch das *helicotrema* verbunden. Von der*scala vestibuli* ist durch die Reisnersche Membran die *scala media* abgetrennt. Hydrodynamisch ist dies ohne belang, wohl aber elektrophysiologisch. Die Lymphe in der scala media ist mit 80 mV gegnüber der in den anderen Kammern positiv geladen. In Abb. 3.10, rechts, ist ein Photo des Querschnitts durch den Gang dargestellt.

Eine besondere Rolle spielt die Basilarmembran, die die scala vestibuli von der scala typmpani trennt. Sie geht entwicklungsgeschichtlich auf die Seitenlinie der Fische zurück. Auf ihr endet der Hörnerv. Der Schalldruck wird durch den *stapes* über das ovale Fenster in die Lymphe übertragen, das runde Fenster dient dem Druckausgleich. Dieser findet sowohl über das Helicotrema statt, als auch über die elastiche Basilarmembran. Wird ein Ton einer festen Frequenz eingestrahlt, so findet der Druckausgleich über die Basilarmembran in einem bestimmten Gebiet statt, das umso näher am ovalen Fenster liegt, je höher die Frequenz ist. Dort wird dann

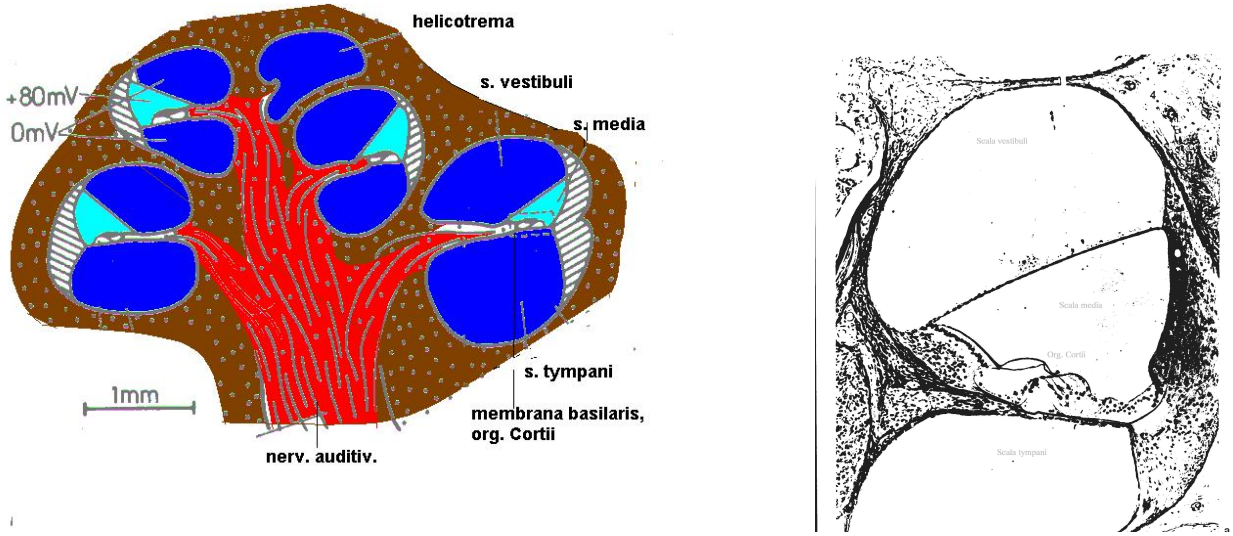


Abbildung 3.10: Querschnitt durch die Schnecke, schematisch und Photo vom Querschnitt durch die *scala vestibuli*

die Basilarmembran besonders stark ausgelenkt. Dies wurde erstmals durch von Bekesy an sehr sorgfältig präparierten frischen anatomischen Präparaten gefunden, s. Abb. 3.11, links. Inzwischen wurde dies auch in vivo Experimenten an Tieren beobachtet, teilweise mit optischen Methoden, teilweise mit Hilfe des Mösbauereffektes, s. Abb. 3.11, rechts. Es fällt auf dass die Resonanzkurve bei den in vivo-Experimenten sehr viel schärfer ist als bei denen, die an anatomischen Präparaten vorgenommen wurden. (Man beachte, dass links eine lineare, rechts eine logarithmische Skala für die relative Auslenkung gewählt ist, die Einheit deciBell (dB) ist ein Zehntel des dekadischen Logarithmus des Verhältnisses, d.h. 10 decibel bedeuten einen Faktor 10.)

In Abb. 3.12 ist die Resonanzkurve, d.h. Amplitude der BM über Amplitude des Trommelfells in den in vivo Experimenten für verschiedene Lautstärken dargestellt. Es wird sehr deutlich, dass für kleinere Lautstärken die Resonanzkurve sehr viel schärfer ist. Diese Dämpfung bei grossen Lautstärken, oder wahrscheinlich besser die ,,Entdämpfung" bei kleinen Lautstärken, wird uns während der ganzen Vorlesung beschäftigen, es ist eines der noch nicht vollständig geklärten Phänomene der Hörphysiologie.

In Abb. 3.13 ist das Mittel und Innenohr nocheinmal sehr schematisch dargestellt, wie es auch in der Modellierung im nächsten Kapitel benutzt wird.

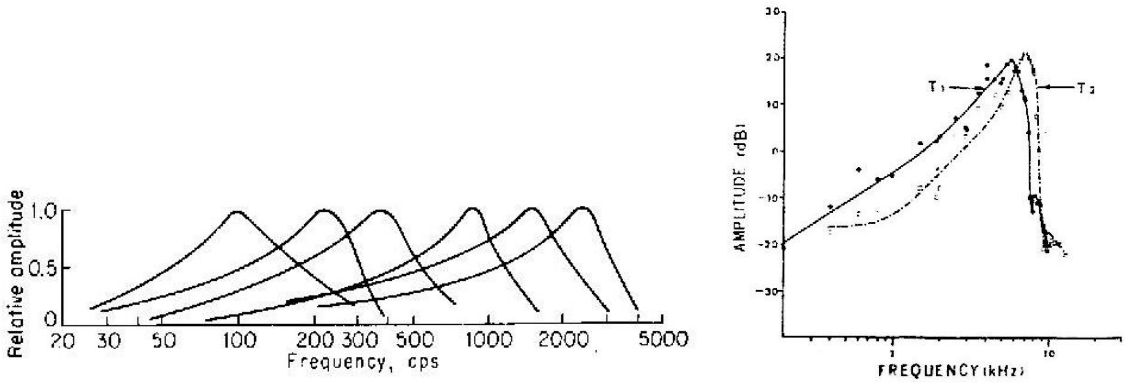


Abbildung 3.11: **Links**: Resultate optischer Untersuchungen an der frisch präparierten menschlichen Cochlea (post mortem, v. Bekesy). Gezeigt ist die Auslenkung der BM an verschiedenen Stellen jeweils als Funktion der eingestrahlten Frequenz, bei sehr hohen Schallpegeln, nach [**?**]. **Rechts:** Resultate von Rhode und Mitarbeitern an Pinseläffchen, in vivo. Auf zwei verschieden Stellen der BM wurde ein radioaktives Präparat aufgebracht, dessen Bewegung in Abhängigkeit von der eingestrahlten Frequenz mit Hilfe des Mössbauereffektes untersucht wurde. Dargestellt ist der dekadische Logarithmus des Verhältnisses (dB) der Amplitude der Auslenkung der BM zur Auslenkung des Hammers (Trommelfells), nach [**?**]

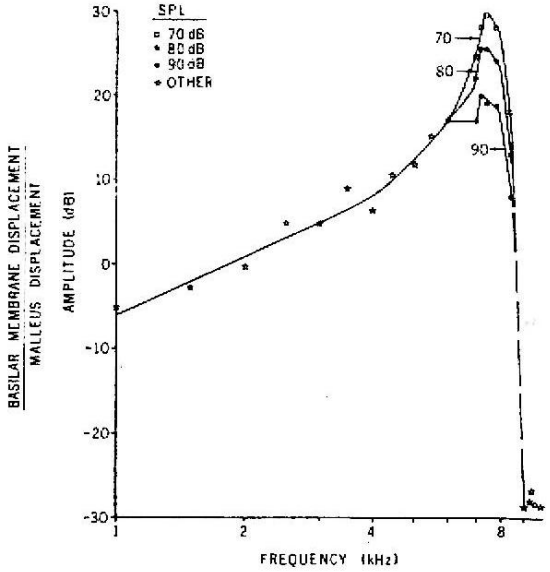


Abbildung 3.12: Die relative Auslenkung der BM für verschiedene Schallpegel, nach [**?**]

3.3 The cochlea

3.3.1 Qualitative considerations

If the movement due to a sound wave arrives at the oval window, the perilymph (hydrodynamically: water) in it will be agitated. Part of this agitation shall create a sound wave in it. For a frequency of 10 Hz, which is near the upper limit of perception, the wave length is 15 cm, that is large compared with the cochlea. Therefore it is a good approximation to assume the perilymph to be an incompressible liquid. So, if the BM were rigid, the pressure would pass practically instantaneously from the entry, the oval window, to the exit, the round window, passing through the helicotrema. But the BM is not rigid, but elastic. Therefore the pressure can make a shortcut, the pressure difference between the scala vestibuli (entry) and the scala tympani (exit) takes place not at the helicotrema, but earlier, somewhere through the BM. A simple enrgetic considerations shows, that high frequencies have the will take the short cut near the windows, low frequancies near the helicotrema. For a high frequency the internal movement in the perilymph is fast, therefore it is advantageous to move as little perilymph as possible, even if it costs some energy to shift the BM. For low frequencies the internal movement is slower, therefore mor mass can be moved and the pressure compensation happens later. But there

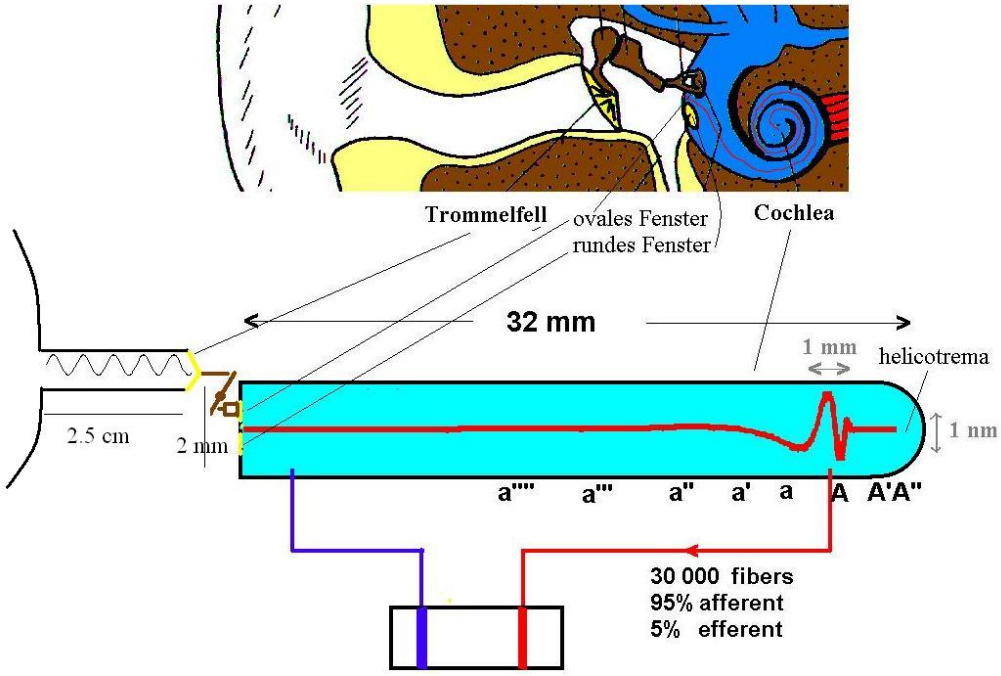


Abbildung 3.13: Schematische Darstellung des Mittelohrs und der Cochlea

is also another, presumably more important mechanism: Since the BM is elastic, one expects that the pressure passes most easily through the BM where it has a resonance frequency which is equal to the frequency of the sound wave. At any rate, we see that the ear is a kind of filter bank: The part with frequencies passes through the high band pass filter, near the windows, low frequencies through the low band pass filter near the helicotrema. This is what has been observed by Bekesy and the Moessbauer experiments.

In Abb. 3.13 ist das Mittel und Innenohr nocheinmal sehr schematisch dargestellt, wie es auch in der Modellierung im nächsten Kapitel benutzt wird.

3.3.2 Hydrodynamics of the cochlea

Here I shall only review some essential steps of the most popular hydrodynamical models of the cochlea. For more details and literature I refer to `cochleadynamics.pdf`. I start with some general formulae.

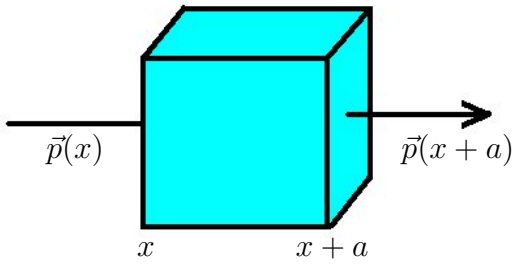


Abbildung 3.14: Forces in x-direction on a small volume element of the perilymph.

By $\vec{p}(x,y,z,t)$, $\vec{v}(x,y,z,t)$ and $\rho(x,y,z,t)$ I denote the pressure, local velocity and density in the perilymph. Since for the acoustically most relevant frequencies the wave length of sound in the lymph is large as compared to the dimension of the cochlea[1], we can consider the liquid as incompressible, $\rho(x,y,z,t) = \rho_0$. The forces in x-direction on a small volume element are shown in fig. 3.16, from which follows, for an ideal liquid without internal friction, the equation of motion:

$$\rho_0 a^3 \frac{d}{dt}\vec{v}_1 = F_1 = \left(p(x+a,y,z) - p(x,y,z)\right) a^2 \tag{3.1}$$

From which follows in the limit $a \to 0$:

$$\rho_0 \frac{d}{dt}\vec{v}_1 = \frac{d}{dx} p(x,y,z) \tag{3.2}$$

from which we finally obtain:

$$\vec{\partial} p \approx \rho_0 \partial_t \vec{v} \tag{3.3}$$

Since matter is conserved and the liquid is incompressible, we have the continuity equation:

$$\vec{\partial}\, \rho_0 \vec{v} = 0 \tag{3.4}$$

from which finally we obtain the important Laplace equation for the pressure:

$$\vec{\partial} \cdot \vec{\partial} p = \vec{\partial}^{\,2} p = \partial_t \vec{\partial}\, \rho_0 \vec{v} = 0 \tag{3.5}$$

or in coordinates:

$$\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2} \right) p(x,y,z) = 0 \tag{3.6}$$

[1]For $\nu = 5000$ Hz, the wave length is given by $\lambda = c_{H_20}/nu = 30$ cm

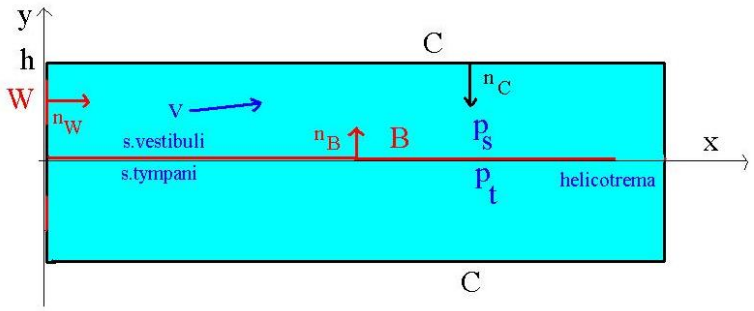


Abbildung 3.15: Two dimensional cartesian version of the cochlea

This equation is essential for the buildup of a travelling wave in the cochlea, since a change of the pressure e.g. in x direction is necessarily related t a change e.g. in y direction.

One can actually show, that the pressure p is mirrored at the BM, if it is $p_0 + p(x, y, z)/2$ in the scala vetibuli, it is $p_0 - p(x, -y, z)/2$ in the scala tympani. In the following we shall not consider the constant part p_0 of the pressure.

There are still two ingredients missing:

1) what are the boundary conditions?
2) How reacts the BM on the pressure $p(x, 0, z)$?

We first consider the boundary conditions:
1a) On the oval window the movement of the lymph is determined by the movement of the stapes
1b) at the walls of the cochlea the vomponent of the local velocity orthogonal to the wall is zero.
1c) At the BM the local velocity of the lymph is the same as that of the BM, $v_2(x, 0, t) = V_y(x, t)$.
1d) at the helicotrema the pressure is zero: $p(x_h, 0, z) = 0$.

We have already idealized the cochlea by unwinding it, we shall go further by making it rectangular. This is of course not essential, but makes the formulae more transparent. Then we can furthermore reduce the problem to a two-dimensional one by separating the third spatial coordinate. A picture of the reduced cochlea is given in fig. 3.15.

I shall not go into details of the derivation of the equations and its solutions, but give only some simple steps which lead to the solution.

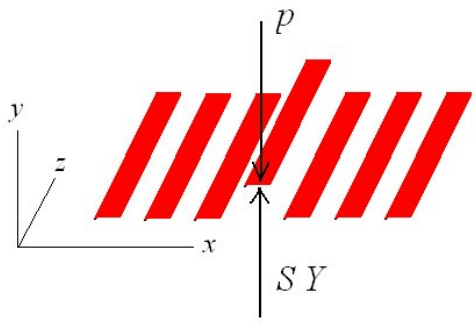


Abbildung 3.16: Forces on the BM, it is here assumed that the cochlea has no longitudinal, but only transverse stiffness (like a grate of bars, Lattenrost).

We make the ansatz, that near $x \approx x_0$ the pressure $p(x,0,t)$ can be described like

$$p(x,0,t) \approx e^{ik(x_0)\,(x-x_0)} \tag{3.7}$$

with the *local wave vector* $k(x_0)$. For pedagogical reasons I shall confine to the case mostly treated, namely the case $k(x_0)h \ll 1$[2] In that case we obtain from (3.5) and the boundary condition 1c):

$$-\rho_0 \partial_t v_2(x,0,t) = -\rho_0 \partial_t^2 Y(x,t) = -h\partial_x^2\, p(x,0,t) \tag{3.8}$$

where $Y(x,t)$ is the deviation of the BM from its rest position at x, hence $V_y = \partial_t\, Y$.

Next we consider a simple model for the BM, we make the simple ansatz for its movement as a series of independent driven damped harmonic oscillators see fig. **??**:

$$\mu\, \partial_t^2 Y = -S\, Y - R\, \partial_t Y - p \tag{3.9}$$

Where μ is the mass of the BM per area, S proportional to the restoring force (elasticity) and R proportional to the damping constant. p is the value of the pressure at the position of the BM.

We furthermore perform a Fourier transform in t, or equivalently, we assume that all quantities have the time dependence $e^{-i\omega t}$, the general case can, if all equations are linear, be obtained by the inverse Fourier transform, that is integrating over all frequencies ω.

[2]This yields the so called shallow water equation: The wave length of the water waves is $\lambda = 2\pi/k$, h is the depth of the water, the condition $kh \ll 1$ thus corresponds to the condition that the wavelength is large compared to the water depth. Because of its extreme long wave length, a Tsunami is well described through the shallow water approximation.

In that case (3.9) reads:

$$-\mu\,\omega^2\tilde{Y} = -S\,\tilde{Y} + iR\,\omega\,\tilde{Y} - \tilde{p}$$

or

$$\tilde{Y}(x,\omega) = \frac{1}{S(x) - \mu\omega^2 - i\omega R(x)}\tilde{p}(x,0,\omega) \tag{3.10}$$

and we obtain the final equation (in the shallow water approximation):

$$\partial_x^2\tilde{p}(x,0,\omega) = \frac{-\rho_0\omega^2}{S(x) - \mu\omega^2 - i\omega R(x)}\tilde{p}(x,0,\omega) \tag{3.11}$$

This equation can be easily integrated numerically with the boundary conditions

$$\partial_x\tilde{p}(0,0,\omega) = i\omega\rho_0\tilde{v}_1(0,0,\omega) \text{ and } \quad \tilde{p}(x_h,0,\omega) = 0. \tag{3.12}$$

A good analytical approximation is the so called WKB approximation[3], it is:

$$\tilde{p}(x,0,\omega) = A\frac{1}{\sqrt{g(x)}}\exp\left[i\int_0^x dx'\,g(x')\right] \tag{3.13}$$

with

$$g(x,\omega) = \omega\sqrt{\frac{\rho_0}{S(x) - m\omega^2 - i\omega R(x)}} \tag{3.14}$$

The local wave vector, introduced earlier is thus given by:

$$k(x,\omega) = g(x,\omega) \tag{3.15}$$

3.3.3 Some properties of the solutions to the hydrodynamical equations

Numerical values, largely taken from deBoer for the linear model of for human cochlea are:

$$\begin{aligned} S(x) &= C_0e^{-\alpha x} - a, \quad C_0 = 10^9, \quad \alpha = 3, \quad m = 0.05,\ R = R_0e^{-\alpha x/2} \\ \delta &= \frac{R_0}{\sqrt{C_0\,m}}; \quad h = 0.1, \quad \rho = 1 \quad \text{all in CGS units} \end{aligned} \tag{3.16}$$

[3]The WKB solution (3.13) is a very good approximation to the exact (numerical) solution in the resonance region, but is quite poor near the oval window, it can be improved considerably by replacing the exponential by a Hankel function H_1.

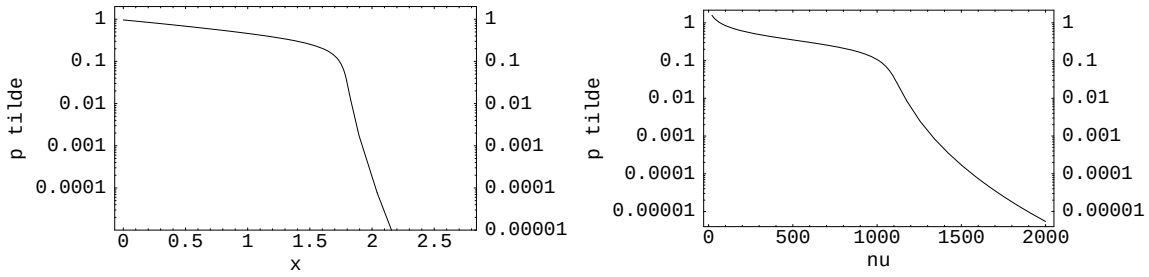


Abbildung 3.17: **Left:** Envelope of the pressure on the BM for a tone 1500 Hz as function of x **Right :** Envelope of the pressure on the BM at 2 cm as function of ν

with δ varying from 0.2 to 0.005.

As long as $S(x)$ is larger than $m\omega^2$, the exponential in (3.13) oscillates and its envelope, that is $|p|$ varies little, but as soon as $m\omega^2 > S(x)$, the exponent obtains an imaginary part and therefore the pressure falls of exponentially. In fig. 3.17 the (ebvelope of the) pressure as function of x for fixed $\nu = 1500$ Hz and ν for fixed $x = 2$ cm are displayed.

Physiologically, the velocity of the BM is the relevant quantity, it is given by (3.10) and (3.13):

$$\tilde{V}_y(x,\omega) = -i\omega\tilde{Y} = \frac{-2i\,\omega}{S(x) - m\omega^2 - i\omega R(x)}\tilde{p}(x,0,\omega) \tag{3.17}$$

$$= \frac{-2i\,\omega}{S(x) - m\omega^2 - i\omega R(x)}\frac{1}{\sqrt{g(x)}}\exp\left[i\int_0^x dx'\,g(x')\right] \tag{3.18}$$

The velocity has large amplitudes in the resonance region $S(x) \approx m\omega^2$ and is strongly damped for $S(x) < m\omega^2$. The resonance frequency ν at which a part x of the BM has its largest amplitude is called *characteristic frequency*. We shall also call the region of the BM, which for a given frequency has maximal amplitude of velocity, the characteristic region of that frequency.

In figure 3.18 the envelopes of the velocity of the BM for two tones of 500 (red) and 1500 Hz (blue) as a function of x for positions of $x = 2.3$ (red) and 1.4 cm (blue) as a function of ν are displayed. In `einton.mov` and `zweiton.mov` the animation of the cochlear movement, grossly exagerated in the y direction, is displayed.

The traveling wave along the BM has the phase velocity:

$$v_g(x) = \frac{\omega}{\mathrm{Re}[k(x,\omega)]} = \frac{\omega}{\mathrm{Re}[g(x,\omega)]} \tag{3.19}$$

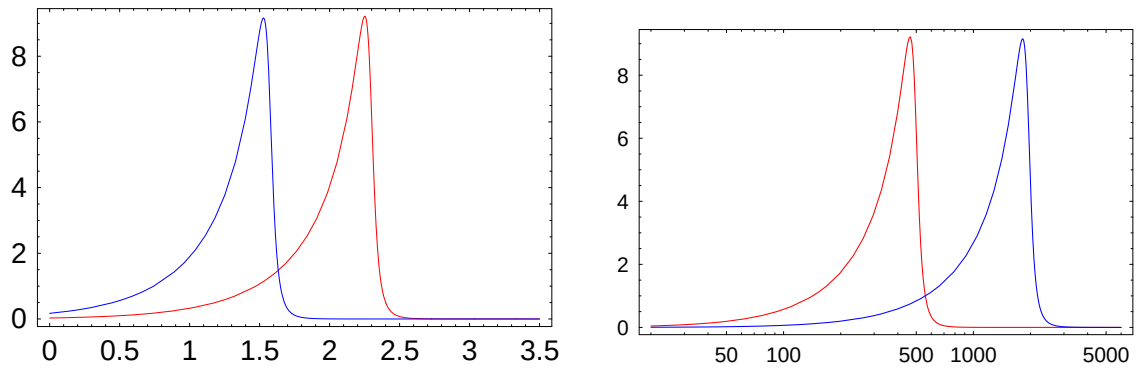


Abbildung 3.18: **Left:** Envelope of the velocity of the BM for two tones of 500 (red) and 1500 Hz (blue) as a function of x **Right :** Envelope of the velocity of the BM for positions of 2.3 (red) and 1.4 cm (blue) as a function of ν

. The time T which needs a wave with cycle frequency ω to reach the position x_0 is given by:

$$T(x_0,,\omega) = \int_0^{x_0} dx\, \frac{1}{v_g(x,,\omega)} = \int_0^{x_0} dx\, \frac{\mathrm{Re}\, g(x,,\omega)}{\omega} \tag{3.20}$$

.

Signals with low frequency, which excite the BM near the helicotrema need of course a longer time to excite the nerves than signals of high frequency, see fig. 3.19. This can be checked with chirps, that is signals which start with a low frequency and end with a high one. Optimal chirps are those, which excite all places at the BM at the same time, see fig. 3.20.

General solutions The general time dependent solution for the velocity is given by:

$$V_y(x,t) = \int_0^{\infty} \frac{d\omega}{2\pi}\, \mathrm{Re}\Big(\mathrm{A}(\omega)\, \tilde{\mathrm{V}}_{\mathrm{y}}(\mathrm{x},\omega)\mathrm{e}^{-\mathrm{i}\omega\mathrm{t}}\Big) \tag{3.21}$$

where V_y is given by (3.17) and $A(\omega)$ is determined by the boundary condition for the pressure at the oval window, see (3.12).

For a fixed frequency, $p(0,t) \sim e^{-i\omega_0 t}$ one has $A(\omega) = A\delta(\omega - \omega_0)$ and in this case

$$V_y^0(x,t) = \mathrm{Re}[\tilde{V}_y(x,\omega_0)e^{-i\omega_0 t}]$$

For a click the pressure is approximately $p(x,t) \approx A\,\delta(t)$ and therefore $A(\omega) = C$ =const.

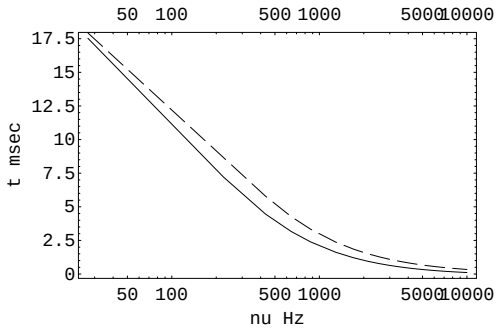


Abbildung 3.19: Time which elapses until the signal reaches the characteristic position of the cochlea, i.e. where the amplitude of the velocity is maximal. Solid line according to the linear cochlea model, 3.20, according to the gamma-tone, see sect. 3.3.4, (3.41)

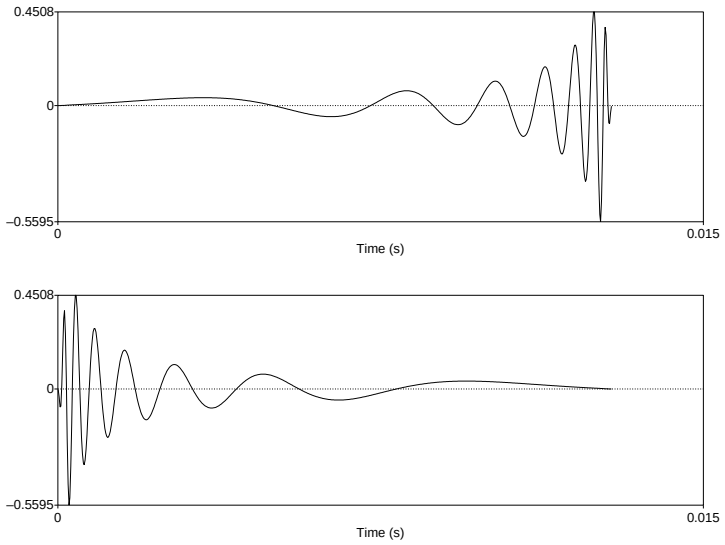


Abbildung 3.20: An up-chirp and a down chirp with increasing resp. decreasing frequency

In that case we have

$$V_y^w(x,t) = \int_0^\infty \frac{d\omega}{2\pi}\,\mathrm{Re}\Big(\mathrm{C}\,\tilde{\mathrm{V}}_\mathrm{y}(\mathrm{x},\omega)\mathrm{e}^{-\mathrm{i}\omega\mathrm{t}}\Big) \tag{3.22}$$

In Figure 3.21 the time behaviour at different positions x of the BM in response to a click (3.22) is shown. The vertical grid line shows the arrival time of the signal as calculated from the phase velocity (3.20).

In figure 3.22 snapshots of the BM with 2 ms time difference are shown, evoked by a single tone of 100 Hz and a double tone of 100 and 1000 Hz, all starting at t=0

The acoustical frequency component of a signal is encoded on the BM both through the position and through the frequency of the BM oscillations. In case of an isolated frequency the BM oscillates just with that frequency. If two frequencies have characteristic regions which overlap, the oscillation is furthermore modulated by the difference of the two frequencies, see fig. 3.23. also programm cochlea-animation-kurz.mb

Nonlinear model We have seen that the sensibility of the ear covers a tremendous dynamical range. The hearing threshold at 3 kHz is at an energy flow of about 10^{-16} W/cm2, the lower limit of damaging risk at 10^{-6} W/cm2 and the pain threshold at 10^{-4} W/cm2. This range can only be understood if nonlinear effects are present. There are also strong other indications for such nonlinear effects, as otacoustic emissions.

A nonlinear model proposed recently has the following relation between Y and p:

$$p(x,0,t) = \alpha(\omega_r(x) - i\partial_t)Y + B|Y|^2\,Y \tag{3.23}$$

In a nonlinear model the general solution cannot be obtained as linear superposition of solutions to fixed frequencies. Therefore only the investigation of pure tones is simple and can proceed along the same lines as discussed above. In Figure **??** and 3.25 the response of the BM to tones with a difference of SPL of 80 are shown. Note that the amplitude evoked by the 80 dB louder tone is only less than 20 db stronger than that of the weaker one. On the right the excitation of the BM at 1.57 cm is shown for two tones with 20 dB difference of sound pressure level, the red curve being 20 dB louder than the red one, the curve are normalized to equal sound pressure. The resemblance with Figure **??** is striking.

For complex tones one has to go back to the time dependent partial differential equation.

There are very good arguments for the assumption that the nonlinear behaviour of the cochlea is due to a dynamical "dedamping" for which the outer hair cells are responsible. We come back to this later.

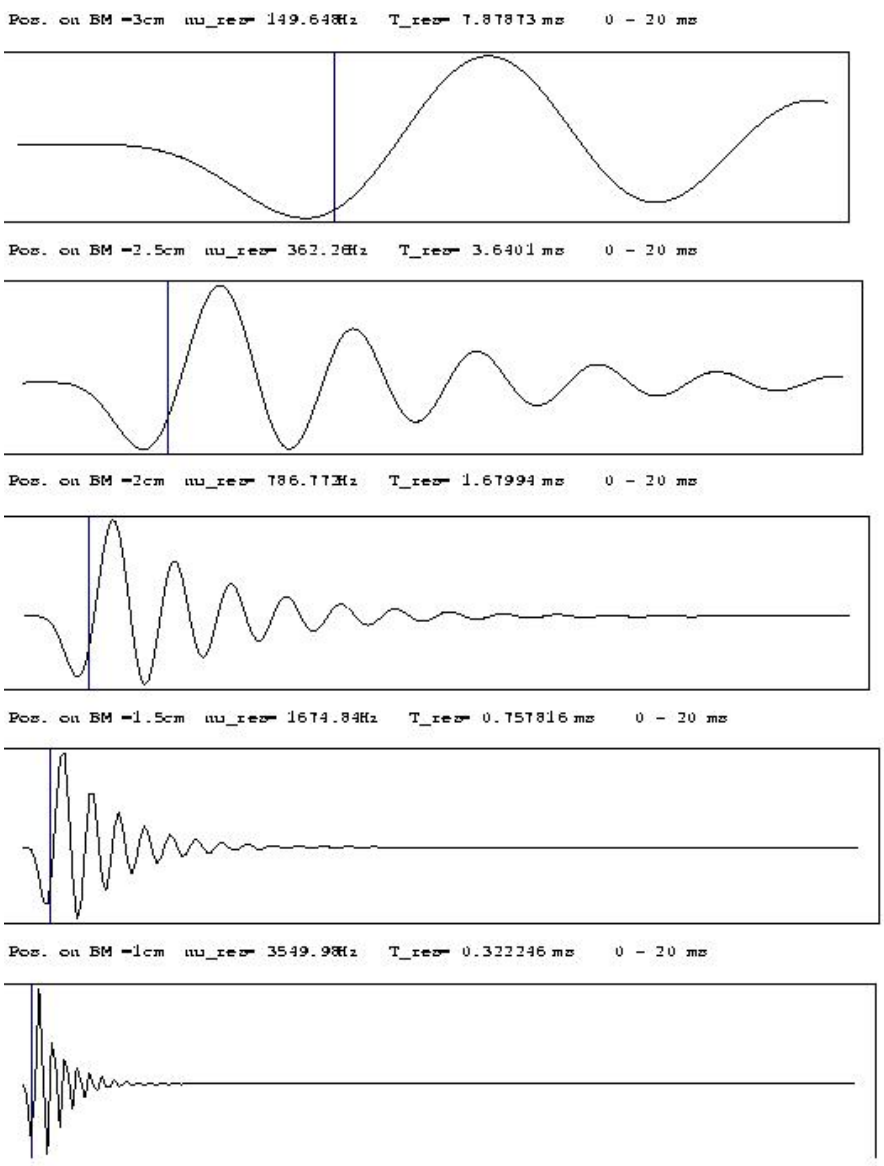


Abbildung 3.21: The amplitude Y of the BM as a function of time (0 to 20 ms) at positions $x = 2.5$, 2, 1.5 and 1 cm as a response to a click. The vertical grid line shows the arrival time of the signal as calculated from the group velocity.

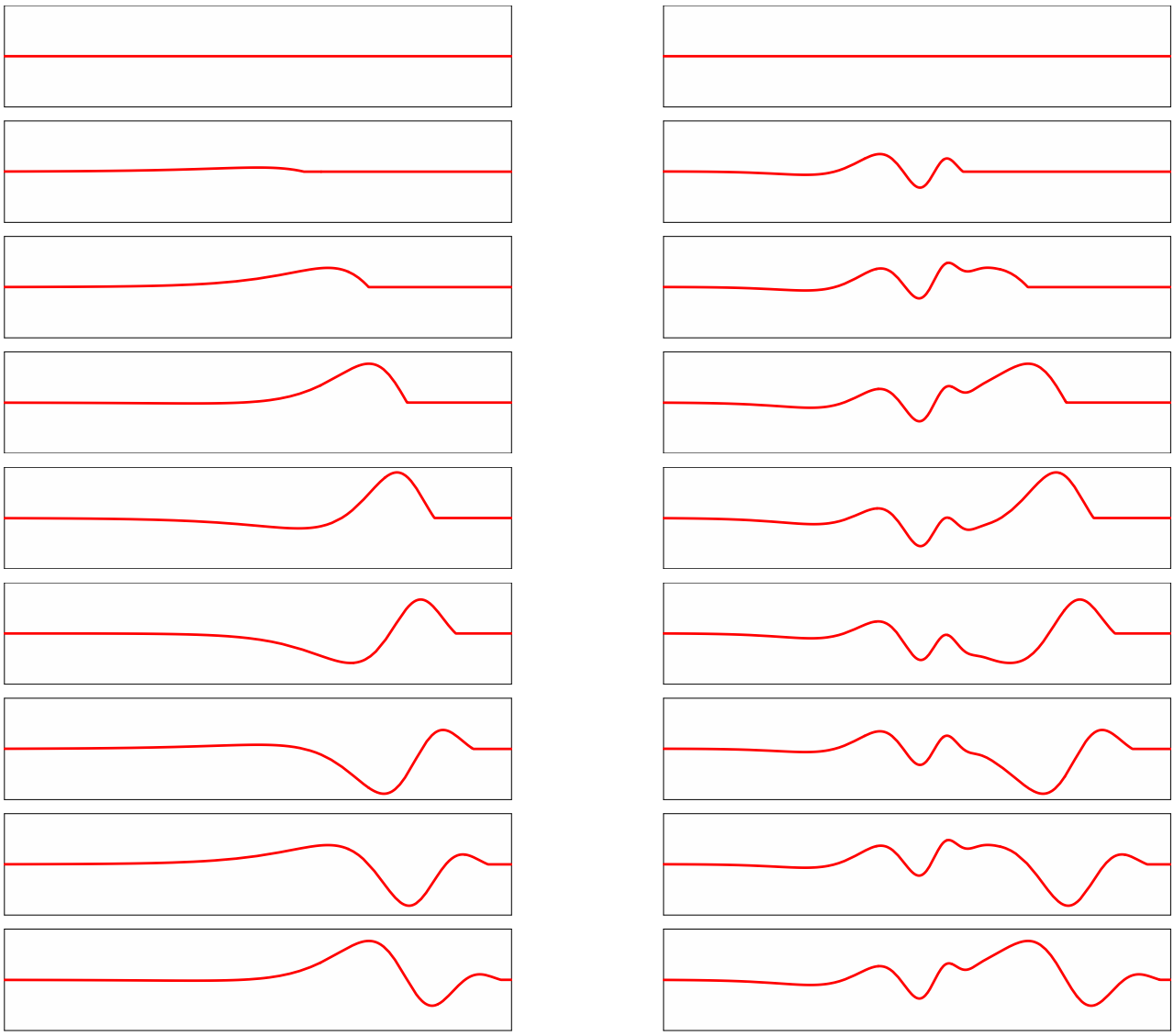


Abbildung 3.22: Snapshots of the BM with 2 ms time difference. **Left:** Single tone of 100 Hz starting at t=0; **Right:** Double tone of 100 and 1000 Hz.

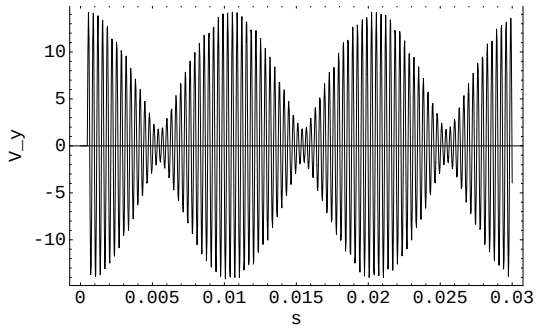


Abbildung 3.23: Oscillation of the BM at 1.2 cm for a complex tone with frequencies 3000 and 3100 Hz.

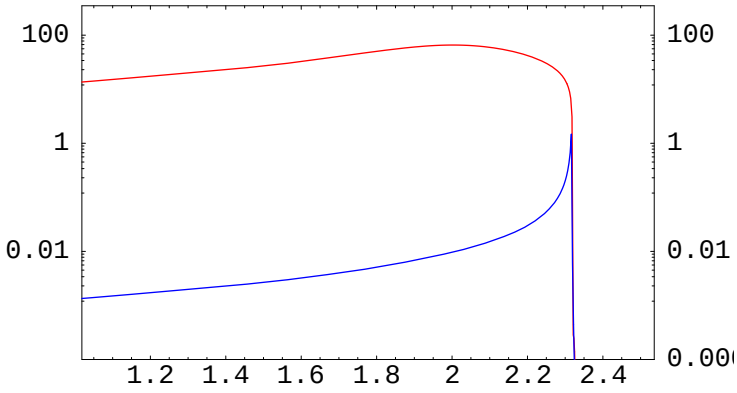


Abbildung 3.24: Maximal amplitude of The BM as function of x for a tone of 500 Hz with two different intesities in a nonlinear model. The tone in red has a SPL which is 10 000 times higher than the tone in blue
, but the amplitude of the BM is only 100 times larger.

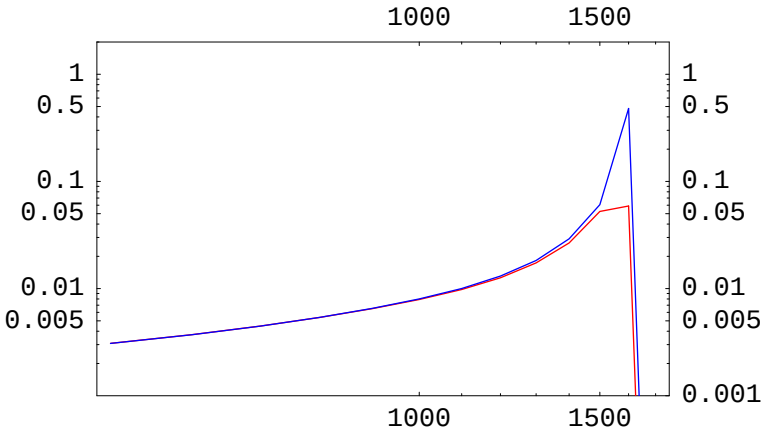


Abbildung 3.25: Maximal amplitude divided by the sound pressure level of the BM as function of ν for a fixed position $x = 1.57$ of two different intensities in a nonlinear model. The tone in red has a SPL which is 100 times higher than the tone in blue.

3.3.4 The cochlea as a filter bank, critical band width

One reads sometimes that the ear performs a Fourier analysis of the acoustic signal. This is only partially true. It is more appropriate to say that the ear acts as a Filter bank. Each section of the BM creates a band pass filter, as can be seen in fig. 3.26. There the envelope of the basilar membrane velocities V_y as function of x for different frequencies $\nu_k \left(\frac{4}{3}\right)^{(k-1)}$, $k = 1, 2 \ldots 25$ are displayed.

We can see from that figure that the excitation regions of two frequencies overlap strongly, if the frequencies are separated by less than a fourth. This is the physiological reason for the psychophysical concept of the critical band width. The critical band width is the frequency interval, where the characteristic positions are close enough to interfere with each other. Very roughly speaking one can say
1. This is the region inside which tones mask each other
2. Inside this region the perceived loudened does not increase linearly, but rather logarithmically, with the excitation energy.
3. It is the region where interference effects occur and can create roughness.

The following simple formula for the bandwidth as function of the central frequency describes

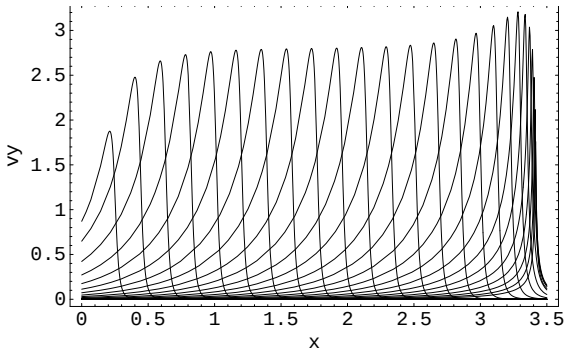


Abbildung 3.26: The curves for $|V_y(x)|$ for frequencies of 16.5 Hz in steps of Fourths, i.e. $16,5\left(\frac{4}{3}\right)^{(k-1)},\ k = 1, 25$.

several psychophysical findings quite well:

$$bw(\nu) = 25 + 0.108\nu \tag{3.24}$$

If we calculate the musical interval corresponding to the bw, we have to form the ratio

$$\frac{\nu + bw(\nu)}{\nu} \tag{3.25}$$

We see that for low frequencies this interval is about an octave, for high frequencies it approaches 1.08, a bit more than a minor second (1.06). Example `praats-bsp/critical-bw`. In fig. 3.27 the bw, expressed in half tones is displayed. Normally narrow chords in the bass are avoided, see fig. 3.28, Chopin, but also used deliberately as "Bizzarerie".

There seems however to be a general tendency in classical music to be at the edge of dissonance. Fig. 3.29 gives an analysis of a string quartet: The the black lines indicate the frequency of intervals. There are 25 % of the intervals larger than the upper thick black line, 50 % larger than the middle and 75 % larger than the lowest thick line. Most of the intervals are indeed in the region between consonance and dissonance.

Making the plausible assumption that the critical band width corresponds to a region with fixed spatial width Δ on the BM, one can determine this region and at the same time find a formula for the characteristic position of a frequency ν: $x(\nu)$.

According to the assumption the critical bw corresponds to a fixed distance Δ on the BM:

$$bw(\nu) = \nu(x - \Delta) - \nu(x) \approx -\Delta \frac{d\nu}{dx} \tag{3.26}$$

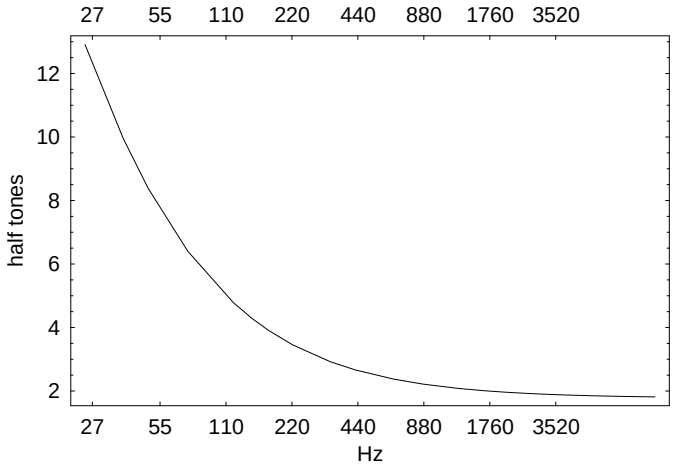


Abbildung 3.27: Critical bw in terms of half tones

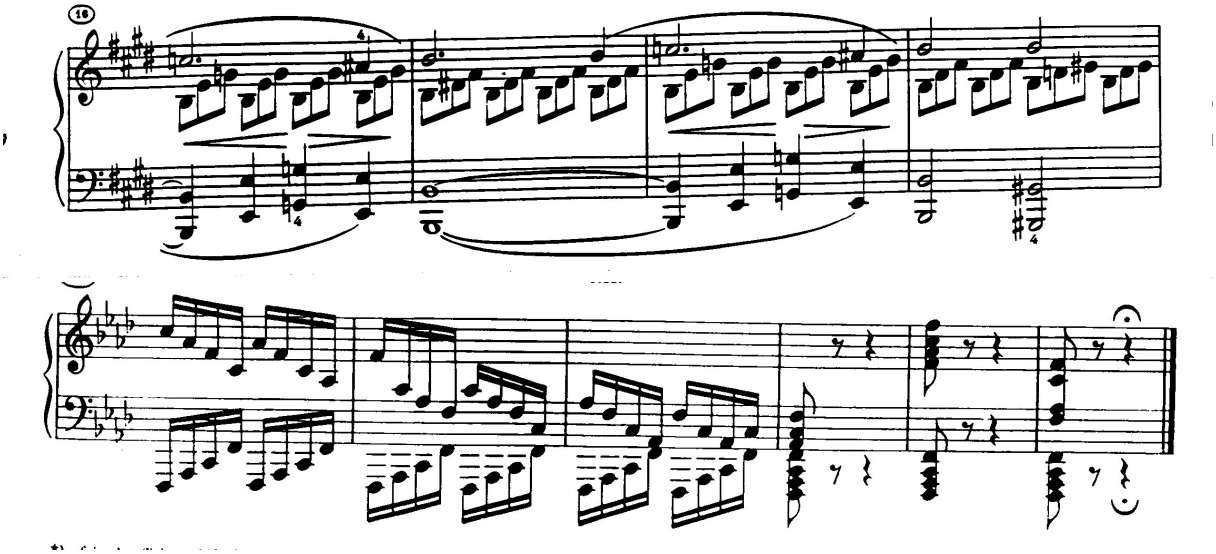
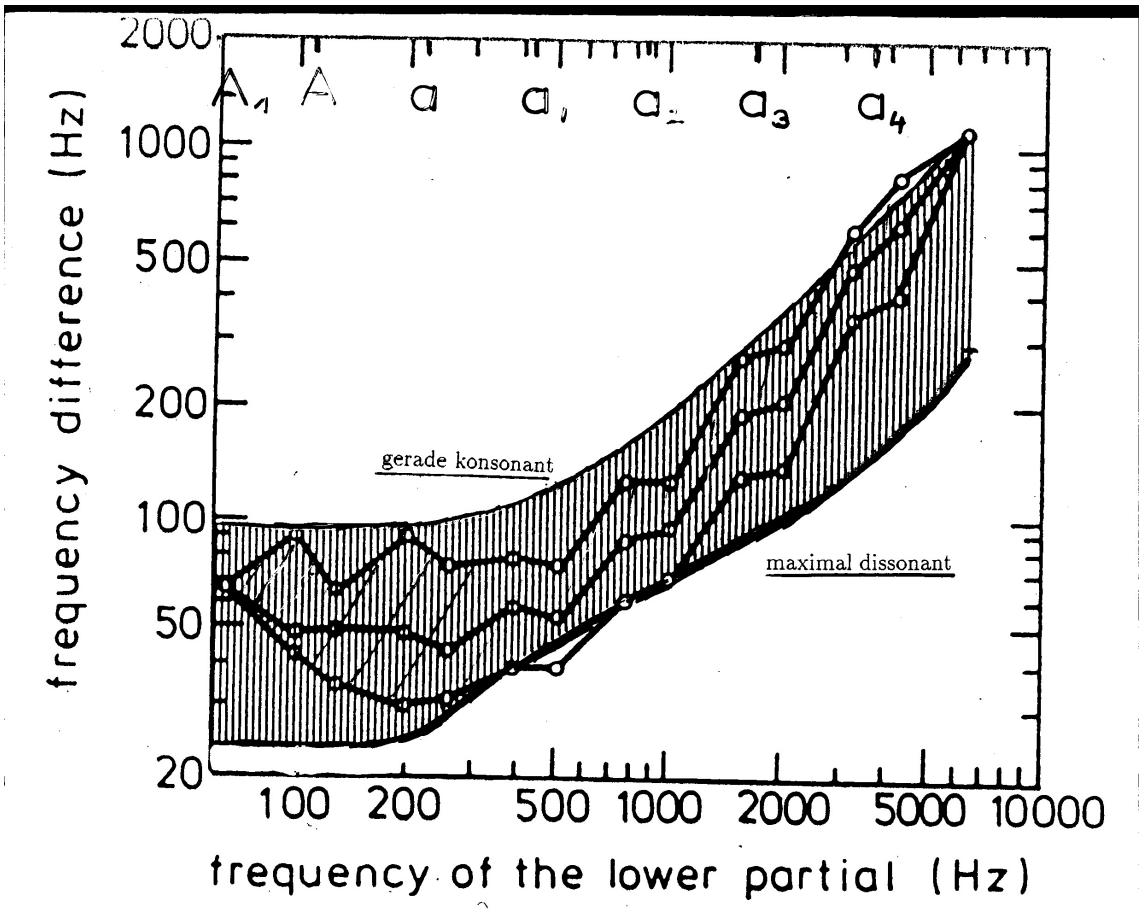


Abbildung 3.28: A line from a piano sonata of Chopin (above) and Beethoven (below)



Statistische Analyse der Akkorde von A. Dvoŕák,
Streichquartett E_b Dur, (op. 51) 3. Satz.
Nach Plomp und Levelt 1965

Abbildung 3.29: Statistical ananlysis of a string quartet of Dvoŕak. There are 25 % of the intervals larger than the upper thick black line, 50 % lagre than the middle and 75 % larger than the lowest thick line

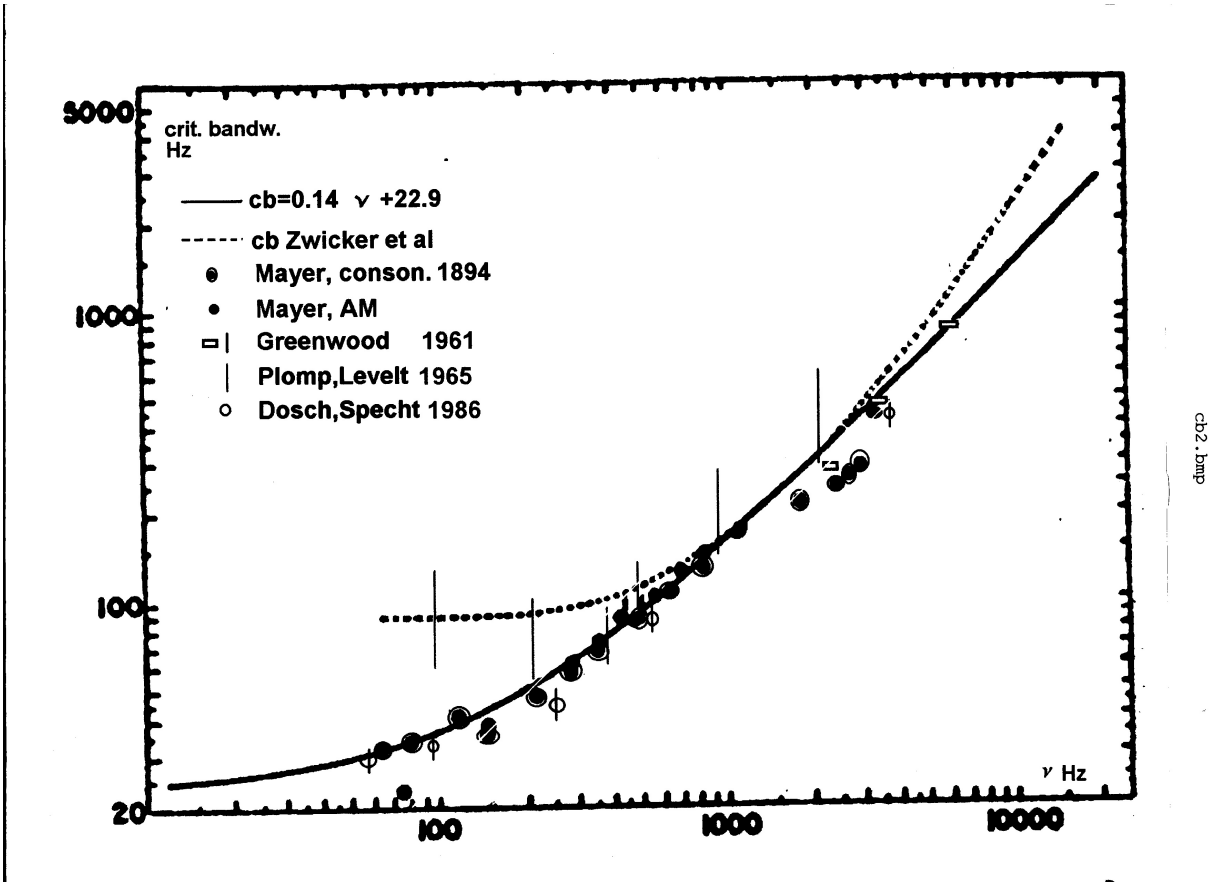


Abbildung 3.30: Several curves for the critical band width and results of experiments using roughness as criterion.

This yields

$$\frac{dx}{d\nu} = \frac{1}{bw(\nu)} \quad \text{or} \quad x(\nu) = c - \Delta \int^{\nu} \frac{d\nu'}{bw(\nu)} \tag{3.27}$$

The cochlea has a length of approx. 3.5 cm, the lowest frequencies of ca 30 Hz are situated at the helicotrema, the highest at the oval window. This yields the conditions:

$$x(30) = 3.5; \quad x(20000) = 0 \tag{3.28}$$

and we finally obtain, using (3.24) : $\Delta = 0.09$ cm.

$$x(\nu) = c - \frac{\Delta}{0.108} \log[25 + 0.108\nu] \tag{3.29}$$

3.3.5 Windowed Fourier Transform. Notes and wavelets

We have already stated that the ear does not perform a Fourier Transform, but rather acts a filter bank. Because of the finite width of the band pass filters the cochlea transmits as well spectral as temporal information: If a acoustical signal is modulated by a function, the time variation of which is slow as compared to the frequencies in the band than this temporal variation modulates the amplitudes of the BM-vibrations. This is of outmost importance for communications. The fast variations determine the quality of the signals, e,g, pitch, whereas the slow variations are necessary in order to form syllables, sentences and melodies. The mathematical tool to treat this analysis is the windowed Fourier analysis, with the special case of wavelet analysis.

Bei einer Fourier Transformierten wird über t von $-\infty$ bis ∞ integriert, die gesamte Information über das Zeitverhalten ist im Spektrum codiert. Bei der Windowed F A wird nur über ein endliches Zeitfenster $g(\tau)$ integriert, wobei das Zeitfenster $g(\tau)$ kontinuierlich über das Signal geschoben wird. Wir definieren die WFA $p(t) \to \hat{p}(t,\omega)$

$$\hat{p}(t,\omega) := e^{-i\omega t} \int_{-\infty}^{\infty} dt'\, g(t-t') f(t') \exp[-i\omega t'], \tag{3.30}$$

wobei $g(\tau) \to 0$ für $|\tau| \to \infty$, der Faktor $e^{-i\omega t}$ dient zum Anschluss an die Wavelet-Transformation.

Ein mögliches Zeitfenster ist ein Rechteck:

$$g(\tau) = \theta(\tau)\theta(\Delta - \tau). \tag{3.31}$$

Analytisch bequemer ist ein Gauss'sches Fenster:

$$g(\tau) = \frac{1}{\Delta\sqrt{\pi}} \exp[-\frac{\tau^2}{\Delta^2}]. \tag{3.32}$$

Hier ist das Fenster im Prinzip nicht endlich, aber die Gauss Funktion fällt so schnell ab, dass wir sie praktisch schon bei einigen Bandbreiten null setzen können.

Die WFT (Fouriertransformierte im Zeitfenster) lässt sich als schreiben als eine Faltung des Signals mit einer ,,Note" g_ω:

$$g_\omega(\tau) = g(\omega,\tau) \exp[-i\omega\tau] \tag{3.33}$$

$$\hat{f}(t,\omega) = \int_{-\infty}^{\infty} dt' f(t') g_\omega(t-t') \tag{3.34}$$

Hierbei kann das "Fenster" g durchaus auch von der Frequenz abhängen, z.B. die Bandbreite Δ in (3.31) oder (3.32).

Durch die WFT wird sowohl die Zeit als auch die Frequenz-Auflösung verschmiert.

Ein Klick, bei dem $p(t) = \delta(t)$, mit (theoretisch) unendlich scharfer Zeitauflösung, wird bei WFT mit dem Gauss'schen Fenster 3.32:

$$p(t) = \delta(t) \;\rightarrow\; \hat{p}(t,\omega) = \frac{1}{\Delta\sqrt{\pi}} \exp[-\frac{(t)^2}{\Delta^2}] \tag{3.35}$$

d.h. die scharfe Zeitauflösung ist verschmiert zu einer Gaussfunktion mit der Fensterbrteite Δ.

Für einen reinen Ton gilt:

$$p(t) = \cos(\omega_0 t) \;\rightarrow\; \hat{p}(t,\omega) = \cos\Big((\omega-\omega_0)t\Big) e^{-\frac{1}{4}\Delta^2(\omega-\omega_0)^2} \tag{3.36}$$

d.h. die scharfe Frequenz ist zu einer Gaussfunktion mit der Breite $4\Delta^2$ verschmiert.

Die WFA information enthält weniger Information über das Spektrum, codiert in der Variablen $\omega = 2\pi\nu$ von $\hat{f}(t,\omega)$, da dieses verschmiert ist, aber dafür wiederum Information über die langsam zeitveränderliche Modulation des Spektrums, codiert in der Variablen t von $\hat{p}(t,\omega)$. Diese Abhängigkeit von 2 Variablen wird dargestellt im Spektrogramm, einer 3-dimensionalen Darstellung, in der i. A. die y-Achse die Frequenz ν und die x-Achse die Zeit$t$ anzeigt; der Grad der Schwärzung indiziert die Grösse des Quadrat des Absolutwertes der Funktion, $|\hat{p}(t,\omega)|^2$. In Abb. 3.31 ist Spektrum und Spektrogramm für einen Ton, der zwischen 440 Hz (a') und 466 Hz (b') wechselt angegeben (`praats/alternate-440-460`).

Ein spezielles Spektrogramm ist das Cochleogramm, bei dem auf der y-Achse die Auslenkung der BM als Funktion der Stelle x dargestellt wird. Auch hier bibt der Grad der Scwärzung die Stärke der Auslenkung an. In fig. 3.33 the cochleogramm for ma_3, falling and rising pitch, and ma_4, rising pitch is given. In the cochleogramm increasing grayness denotes increasing amplitudes, 5 Bark is the characteristic region for ca 440 Hz, 18 Bark for ca 4400 Hz. Below the cochleogramms the pitch, as calculated by a pitch extractor is shown.

Nimmt die Fensterbreite linear mit der Frequenz ab, so hat die Note immer gleichviele ,,wiggles". Setzen wir z. B. im Gauss'schen Fenster$\Delta = \alpha/|\omega|$ erhalten wir die Note:

$$g_\omega(\tau) = \frac{|\omega|}{\alpha\sqrt{\pi}} \exp[-\frac{(\omega\tau)^2}{\alpha^2} - i\omega\tau] \tag{3.37}$$

. In diesem Falle spricht man von einem ,,wavelet".

Allgemein ist die wavelet-transformierte definiert durch:

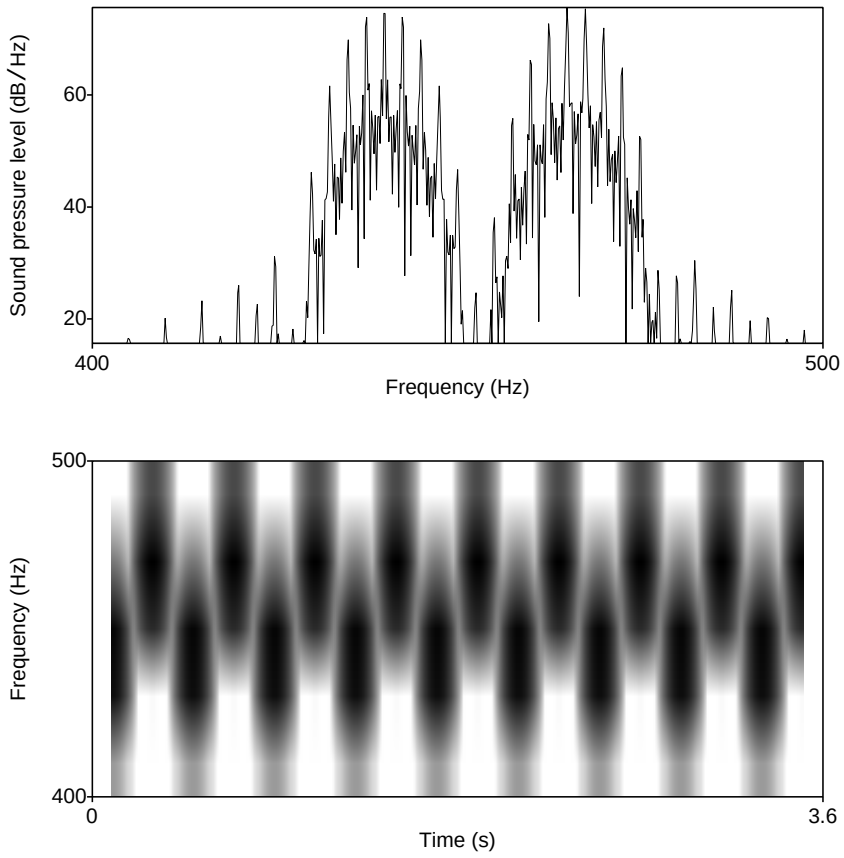


Abbildung 3.31: Spektrum und Spektrogramm eines zwischen 440 und 466 Hz wechsenden Tones

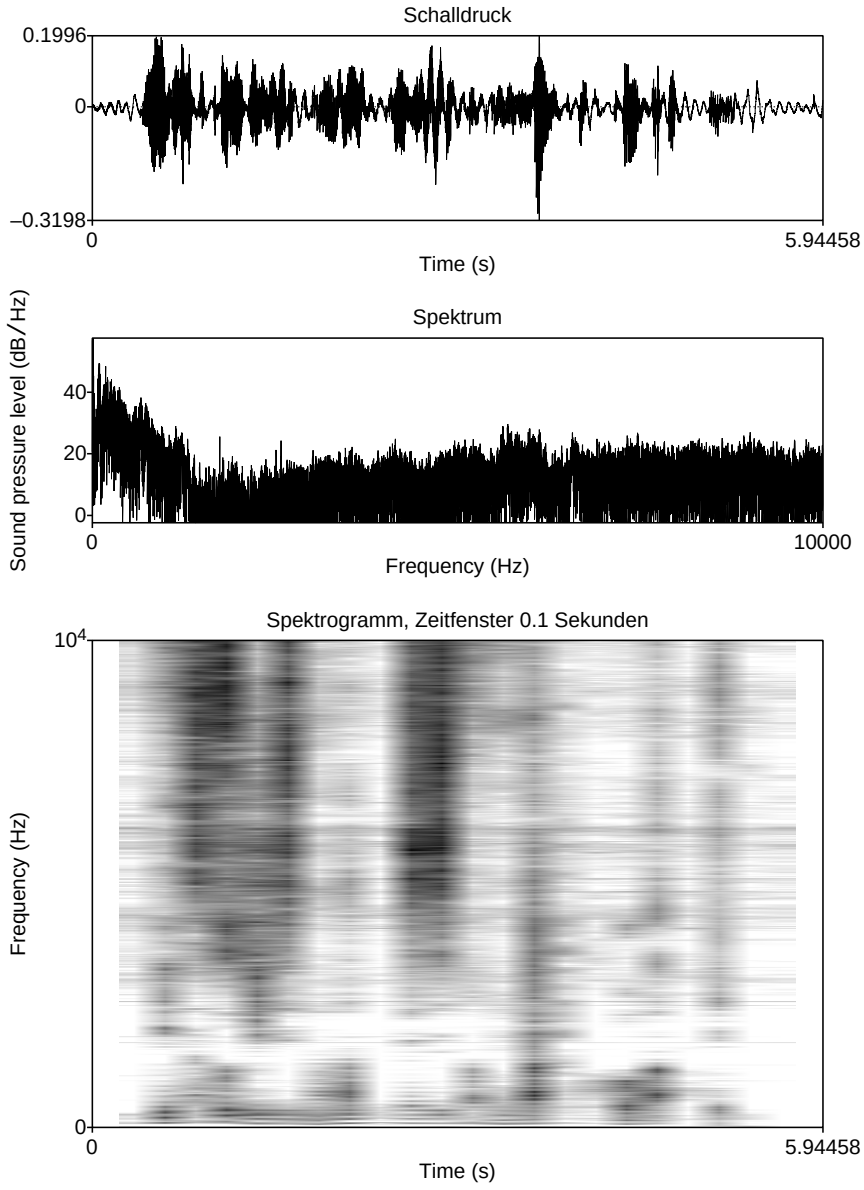


Abbildung 3.32: Signal, Spektrum und Spektrogramm des Satzes „Jedes askustische Signal lässt sich auf eine Schalldruckkurve zurückführen”.

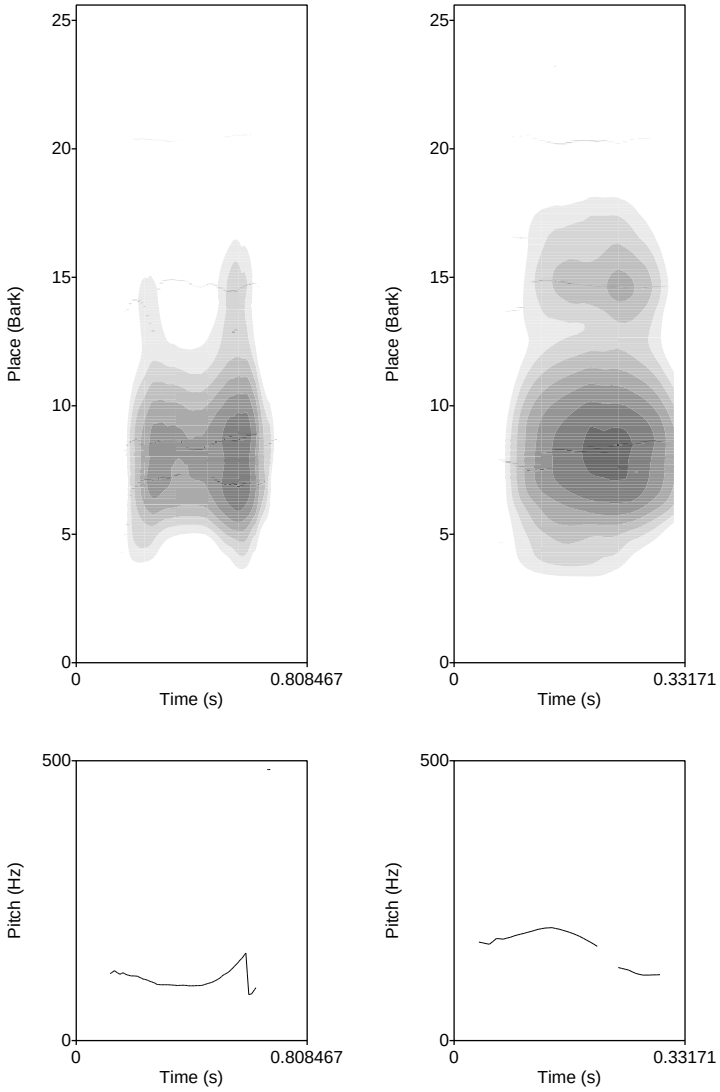


Abbildung 3.33: Cochleogram and pitch development of the chinese syllables ma_3 (left) and ma_4 (right). 4 bark correspond to ca 440 Hz, 18 Bark to 4400 Hz.

$$p_w(t,s) = \int_{-\infty}^{\infty} dt'\, f(t') \frac{1}{|s|^p}\, h\left(\frac{t-t'}{s}\right) \tag{3.38}$$

Wir sehen, dass die Note 3.37 in der Tat eine Wavelet-Transformation ist, mit

$$p = 1; \quad \omega = 1/s; \quad h(x) = e^{ix - x^2/\alpha^2} \tag{3.39}$$

.

Die Geschwindigkeiten der BM an der Stelle x, gegeben durch (3.22) und dargestellt in fig. 3.21 sind Noten (beinahe wavelets), die die windowed FA der BM vermitteln. Die Frequenz der Note ist durch die charakteristische Frequenz der betreffenden Stelle gegeben, die Breite durch die Dämpfung auf der BM. Haben wir ein Signal $p_s(t)$, so ist das zeitliche Verhalten aud der BM an der Stelle x gegeben durch[4]:

$$V_y^s(x,t) \approx \int_0^{\infty} dt'\, V_y^w(x,t') p(t-t') \tag{3.40}$$

Da die Noten 3.22 mathematisch etwas unbequem sind, benutzt man häufig mathematisch etwas einfachere, die sog. Gamma-töne.

Ein Gamma-Ton ist ein kausal exponentiell gedämpfter Ton mit Potenzanstieg.

$$\gamma(t;\gamma,\nu,a,\phi_0) = t^{\gamma-1}\ \exp\Big[-2\pi\,\mathrm{bw}(\nu)\,t\Big]\cos[2\pi\nu t + \phi_0]\ \theta(t) \tag{3.41}$$

s scale factor (1.5). Die Grösse bw ist die äquivalente rechteckige Bandbreite (ERB) des Filters im Frequenzraum, z. B. gegeben durch (3.24). In Praat is für den Gammaton der etwas compliziertere Ausdruck:

$$\mathrm{bw}(\nu) = 6.23\ \ 10^{-6}\nu^2 + 93.39\ \ 10^{-3}\nu + 28.5, \tag{3.42}$$

. Die exponentielle Dämpfung ist der Cochlea angepasst, siehe fig. 3.34. ϕ_0 ist die Anfangsphase.

Für die Fourier-Transform benutzen wir :

$$\int_0^{\infty} dt\ t^x e^{-\alpha t} e^{i\omega t} = \frac{\Gamma(1+x)}{(\alpha + i\omega)^{1+x}} \tag{3.43}$$

[4]Ich überspringe hier meherere zusätzliche Faktoren, die durch die Umwandlung von der Bewegung des ovalen Fensters herrühren

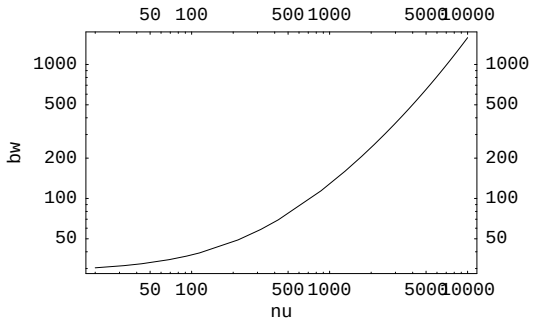


Abbildung 3.34: Die Funktion $bw[\nu]$ im Gamma-ton 3.41

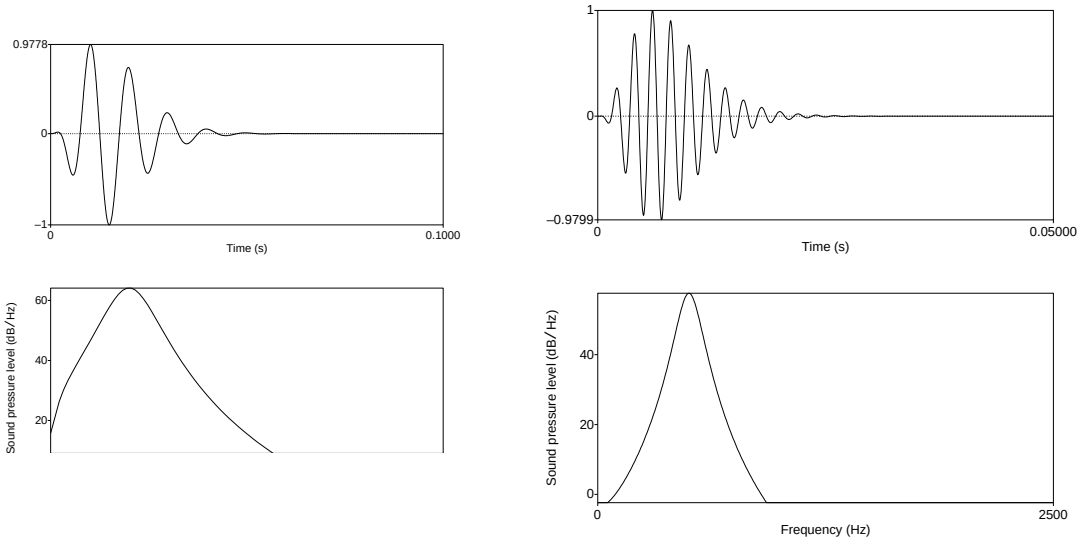


Abbildung 3.35: Impuls und Spektrum des Gamma-Tones 3.41 mit $\nu_0 = 100$ und 500 Hz, $\gamma = 3$

für $\mathrm{Re}\, x > 0$.

Daraus folgt:

$$\tilde{\phi}(2\pi\nu) = \frac{1}{2}\Gamma(\gamma)\left(\frac{1}{2\pi\Big(\mathrm{bw}(\nu_0) - i(\nu-\nu_0)\Big)^{\gamma}} + \frac{1}{2\pi\Big(\mathrm{bw}(\nu_0) + i(\nu+\nu_0)\Big)^{\gamma}}\right) \tag{3.44}$$

Kapitel 4

Von der mechanischen Bewegung zur Wahrnehmung

4.1 Movement of the BM and firing rate of the auditory nerve

Auf der Basilalmembran (BM) befindet sich das Corti'sche Organ, s. Abb. 4.1. Eine entscheidende Rolle spielen die Haarzellen, s. Abb. **??**. Ihre Bewegung werden in Nervenimpulse übersetzt. Dies geschieht wahrscheinlich durch einen mechano-elektrischen Schaltmechanismus, wie er in Abb. 4.2, rechts, bildlich dargestellt ist. Werden die Haarzellen nach rechts (zu den grösseren hin) gekippt, öffnet sich eine Klappe und positive Ionen aus der scala media strömen in die Zelle, dies erzeugt neuronale Signale. Dieser Mechanismus führt zu der beobachteten Gleichrichterwirkung der Signale, nur Bewegungen nach rechts, d.h. bei der BM von der s. vestibuli zur s. tympani hin, führen zu erhöhter neuronaler Aktivität. Allerdings sind die Verhältnisse nach neueren Untersuchungen etwas komplizierter. Bleibt die Haarzelle längere Zeit (mehr als ca 25 ms) gekippt, wandert die Öffnung nach unten und die Klappe schliesst sich wieder; dies führt zu einer Adaption.

Die Haarzellen sind sehr empfindlich. Die thermische Bewegung in Querrichtung an der Spitze beträgt etwa 3 nm, doch erzeugen durch Zeitmittelung periodische Bewegungen mit einer Amplitude von etwa 0.3 nm noch relevante Signale.

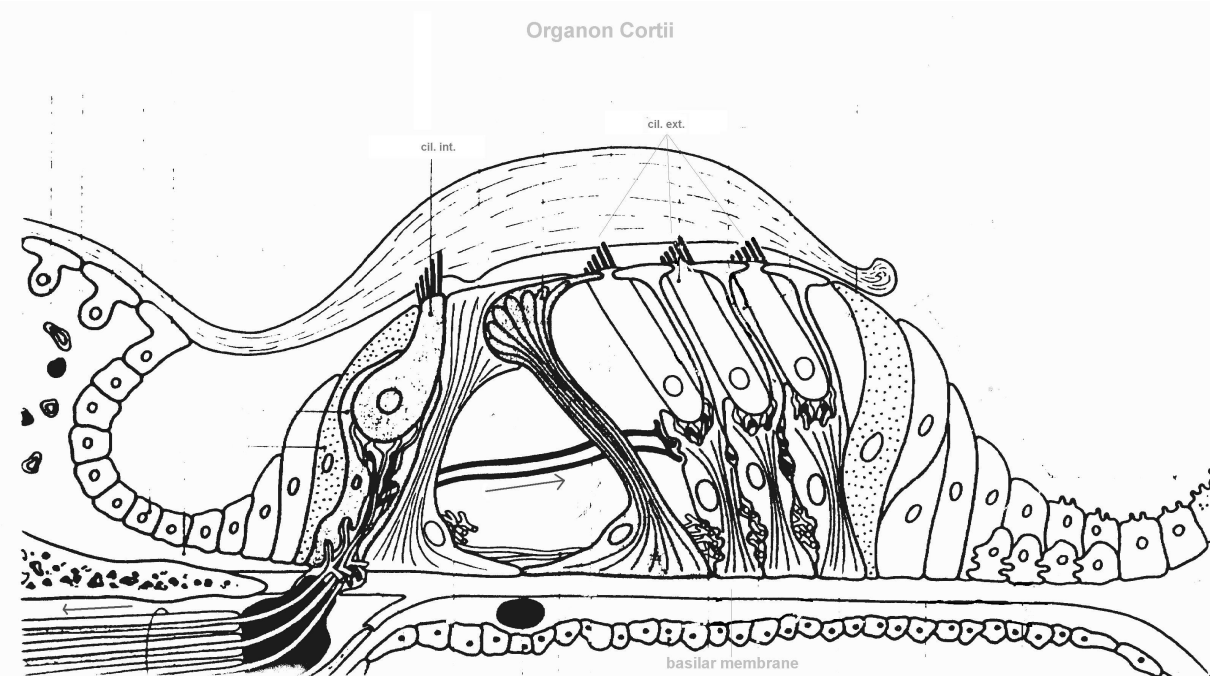


Abbildung 4.1: Das Corti'sche Organ auf der BM

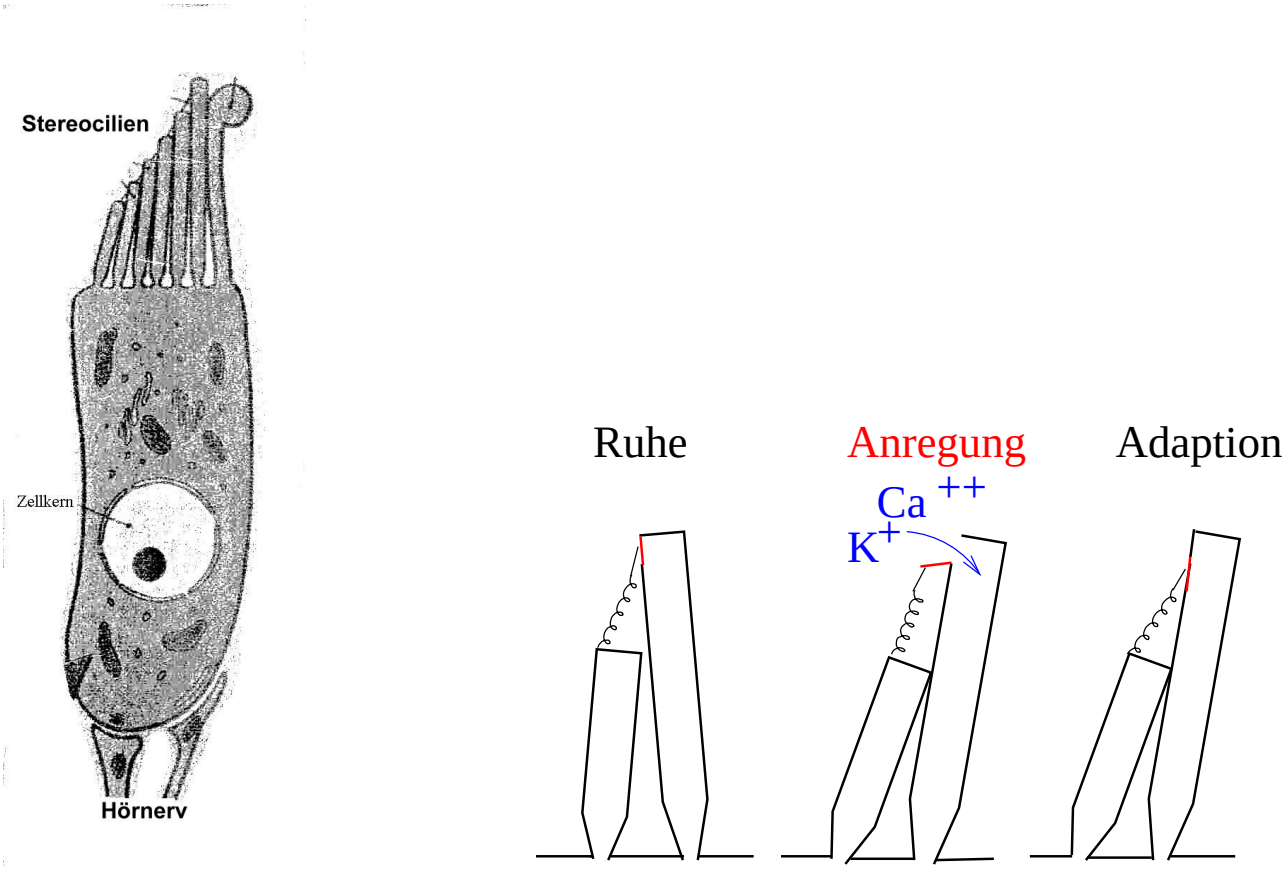


Abbildung 4.2: Haarzellen auf der BM, rechts ist der wahrscheinliche Schaltmechanismus graphisch dargestellt, nach [**?**]

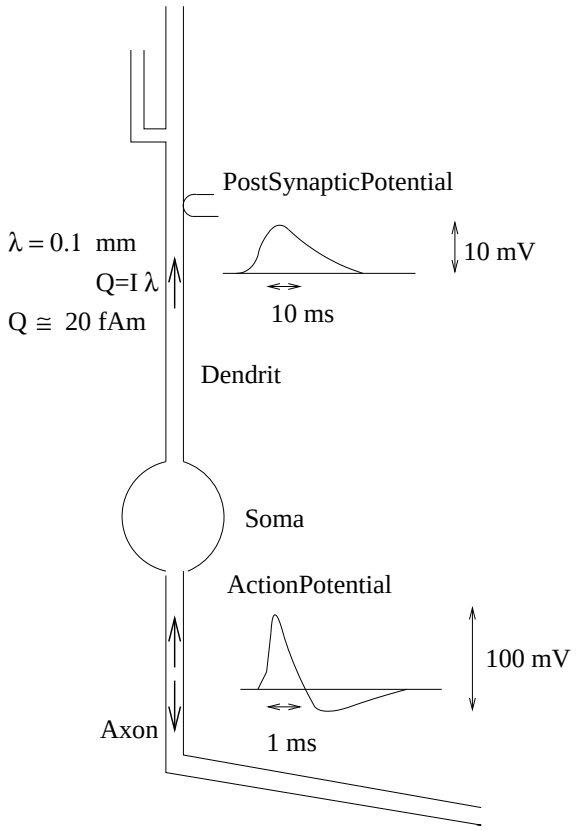


Abbildung 4.3: Schematic picture of a neuron with action potential and postsynaptic potential

4.1.1 Kodierung

Der Signaltransport geschieht entlang der Axone eines Neurons (Nervenzelle) durch die Aktionspotentiale, die Übermittlung zwischen verschiedenen Nervenzellen über die Synapsen. Die Aktionspotentiale können direkt durch Sonden gemessen werden, die sind von der Grösse etwa 100 mV und pflanzen sich mit einer Geschwindigkeit von einigen zehn m/s in den Axonen fort. Zwischen den Nerven geschieht der Austausch durch Synapsen, die postsynaptischen Potentiale sind etwa 10 mV. In App. 4.3 ist dies schematisch dargestellt.

Die Information wird durch eine Modulation der Frequenz der Aktionspotentiale kodiert, entweder durch eine gegenüber der spontanen Emissionsrate erhöhte oder eine erniedrigte Aktivität.

Es stehen im auditorischen System insgesamt 4 Kodierungen zur Verfügung, die wahrscheinlich

alle gemeinsam zur Signalveratrbeitung genutzt werden:

1. Labeled Line Code: Ein bestimmtes Neuron transportiert eine ganz spezifische Information, z.B. die Tonhöhe. Dieser Code spielt sicher im auditorischen System eine wichtige Rolle, da verschiedene Neuronen an Haarzellen an verschiedenen Stellen der BM koppeln und diese Information bis zum Hörkortex parallel verarbeitet wird (tonotope Organisation der Hörbahn)

2. Rate Code. Die Information ist in der quantitativen Erhöhung (oder Erniedrigung) der Feurungsrate enthalten. Dieser Code spielt sicher bei der Lautstärke eine Rolle.

3. Temporal Code. Die Nerven feuern in Phase mit dem Signal ganze Salven (volleys) von Aktionspotentialen. Auch diese Kodierung spielt höchstwahrscheinlich bei der Tonhöhenerkennung eine Rolle (und sicher beim Richtungshören).

4. Ensemble Code. Die Information wird von einem vernetzen Ensemble von Neuronen übermittelt, nicht nachgewiesen, aber wahrscheinlich.

4.1.2 Orts- und Zeitabhängigkeit

Werden die durch einen Sinusston evozierten Aktionspotentiale eines Neurons gemessen, so ergibt sich eine maximale Erhöhung der Feurungsrate bei einer ganz bestimmten Frequenz, der sogenannten charakteristischen Frequenz (CF) der Nerveneinheit. In Abb. 4.4 ist die Erhöhung der Feurungsrate einer Nerveneinheit als Funktion der eingestrahlten Frequenz für verschiedene Schallpegel dargestellt. Wir beobachten die schon bei der Messung der Auslenkung der BM beobachtete, und in nichtlinearen Modellen auch berechnete, Verbreiterung (stärkere Dämpfung) bei hohen Schallpegeln. In Abb.4.5 ist eine neuronale Tuning Kurve mit einer physiologischen der BM verglichen. Es ist der Schallpegel eingetragen der bei einem Neuron eine gerade noch messbare Erhöhung der Feurungsrate bewirkte, bei der mechanischen Kurve ist die gerade noch messbare Auslenkung der BM an einer Stelle, die der CF des Neurons entspricht aufgetragen. Man sieht, dass die Feurungsraten die mechanische Bewegung der BM sehr gut widerspiegeln. Wir können also davon ausgehen, **dass die Feurungsraten der Neuronen des 8. Nerves die mechanische Bewegung der BM recht getreu widerspiegeln.**

In Abb. 4.6 sind neuronale tuning Kurven bei von mehreren Neuronen mit verschiedener CF dargestellt.

Die Neuronen des 8. Nerves benutzen also das *line coding* zur Übermittlung der Information, wo die BM erregt wurde, und die Feurungsrate zur Übermittlung der Information über die Auslenkung oder Geschwindigkeit der BM an der entsprechenden Stelle. Aber auch die zeitliche

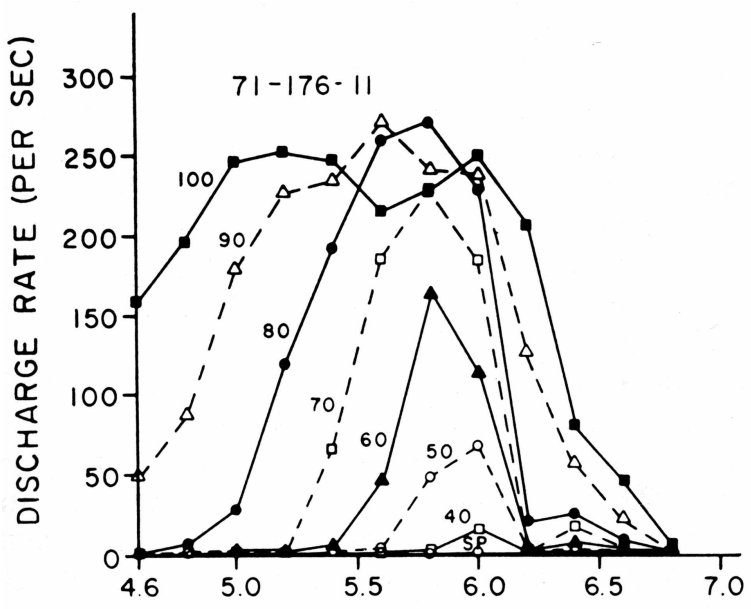


Abbildung 4.4: Erhöhung der Feurungsrate in einem Neuron in Abhängigkeit von der Frequenz für verschiedene Schallpegel, gemessen an einem Pinseläffchen, nach [?]

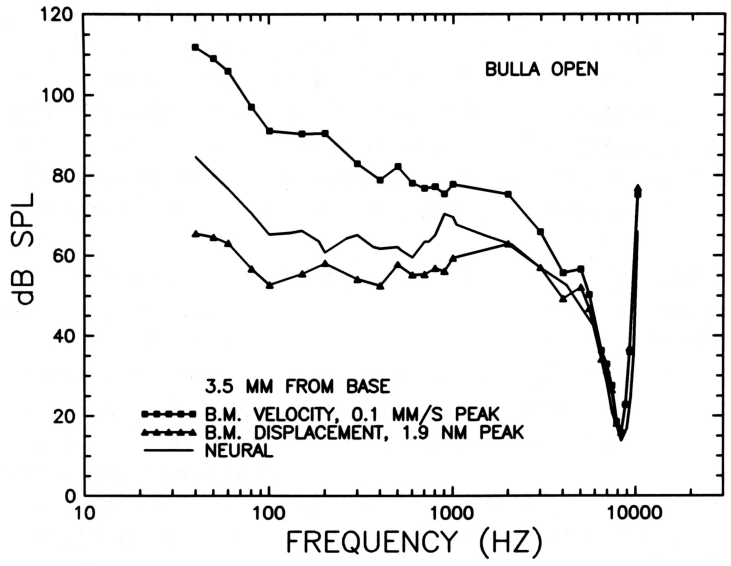


Abbildung 4.5: Neuronale (ohne Symbole) und mechanische Tuning Kurve an einem Chinchilla, 3.5 mm von der Basis, Quadrate: gleiche geschwindigkeit der BM (0.1 mm/s), Dreiecke: gleiche Auslenkung (1.9 nm). nach [**?**]

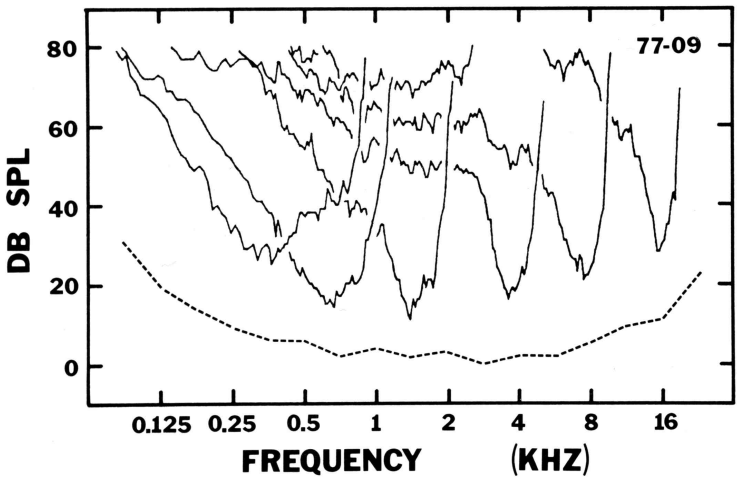


Abbildung 4.6: Neuronale tuning Kurven an einem Chinchilla. nach [**?**]

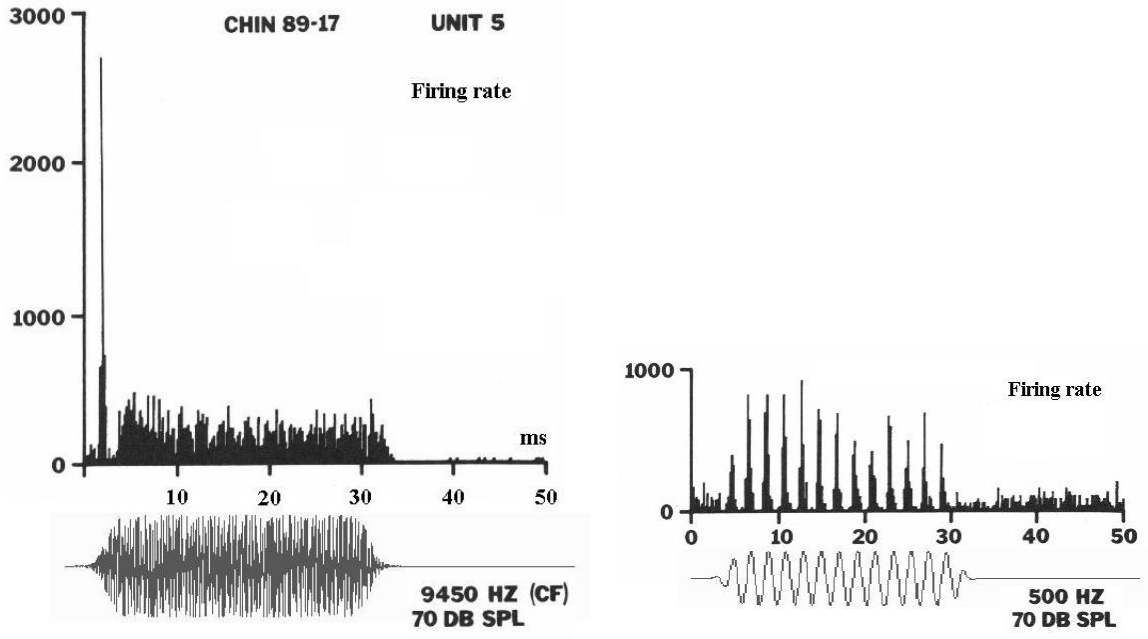


Abbildung 4.7: Zeithistogramm der Feurungsrate eines Neurons mit der CF 9450 Hz und einer spontanen Feurungsrate von 59 Spices/s. Der Stimulus hatte eine An- und Abfallzeit von 4 ms und dauerte 25ms. **Links** Frequenz 9450 Hz,**Rechts** 500 Hz; nach [**?**]

Kodierung wird verwandt, und zwar in der Form einer festen Phase zwischen Signal und Aktionspotential, wie aus Abb. 4.7 ersichtlich. Hier ist die zeitliche Abfolge der Aktionspotentiale für ein Neuron mit einer CF von 9450 Hz dargestellt. Bei dieser hohen Frequenz ist zwar die Feurungsrate zu Beginn des Signals besonders hoch, aber eine feste Phasenbeziehung zwischen dem Signal und der Feurungsrate ist nicht beobachtbar. Ist die eingestrahlte Frequenz aber 500 Hz, die Periode also 20 ms, dann erfolgt auch die Feurungsrate in Salven mit 20 ms Abstand.

Die Frequenz eines Tones wird also 2 mal kodiert: einmal über das line coding (Ortstheorie) und einmal über die zeitliche Kodierung (Zeittheorie). Höchstwahrscheinlich gehen beide Kodierungen in die Empfindung der **Tonhöhe** ein. Allerdings ist die Zeitauflösung im Millisekundebereich nur für die frühen (subkortikalen) Stufen der Hörbahn gegeben, im Kortex ist die Zeitauflösung schlechter als etwa 10 ms. Die temporale Information muss also, wenn sie kortikal verwendet wird, auf der Zwischenstufe zwischen auditorischen Nucleus und Kortex in eine Ortsinformation umgewandelt werden.

Vgl. Abb. 3.4 und Beispiele `praats/pitch-sensation`

Wir betrachten zunächst die neuronalen Feurungsraten asl antwort auf einen Klick, d.h. ein Signal mit der SP-Kurve $p(t) = \delta(t)$. Wir hatten gesehen dass die bewegung der BM in diesem Falle gegeben ist durch $V_y^w(x,t)$, s. 3.22 und Abb. 3.21. Wir betrachten zunächst die neuronalen Feurungsraten asl antwort auf einen Klick, d.h. ein Signal mit der SP-Kurve $p(t) = \delta(t)$.

Interessant ist auch die Antwort auf einen Click, d.h. ein signal mit der SP Kurve $p(t) = c\delta(t)$ die eins, d.h konstant ist, wird dir ganze BM angeregt. Wir hatten gesehen dass die Bewegung der BM an einer feten Stelle x in diesem Falle gegeben ist durch das "wavelet" $V_y^w(x,t)$, s. 3.22 und Abb. 3.21. In Abb. 4.8 ist links die Feurungsrate für einen Depressionsklick und Kompressionsklick eingezeichnet, wobei die Feurungsrate für den Kompressionsklick mit -1 multipliziert wurde. Hierbei wird zweierlei sichtbar: Die BM schwinkt nach oben (s. vestibuli) und nach unten (s. tympani)doch die Aktionspotentiale werden nur durch die Schwingung in eine Richtung ausgelöst (Gleichrichterwirkung der Haarzellen), die eingezeichnete Kombination von Kompressions und Depressionsklick kompensiert diese Gleichrichterwirkung. Zum andern sehen wir deutlich das ,,Nachklingeln der BM, mit einer Dämpfung, die umso grösser ist, je höher die Frequenz ist. Die zumindest qualitative Übereinstimmung zwischen der Rechnung mit dem einfachen BM-Modell (Abb. 3.21) und den Feurungsraten Abb. 4.8 ist deutlich. Ich will nicht näher darauf eingehen, wie die Gleichrichterwirkung zustande kommt. Früher wurde allgemein angenommen, dass nur die Bewegung nach unten die Feurungsrate erhöht, doch ergaben neuere Messungen, dass die Situation komplizierter ist, und sowohl von der CF als auch von der Lautstärke abhängt.

In Abb. 4.8 ist rechts die Korrelation $c(\tau)$ zwischen Schalldruck eines Rauschsignals und der Feurungsrate (REVCOR) dargestellt: $c(\tau) = \int dt\, p(t)\, f(t+\tau)$, wobei $p(t)$ der Schalldruck und $f(t)$ die Feurungsrate ist. Man kann sich diese Korrelationsfunktion als die Überlagerung ganz vieler Klick-Antworten vorstellen. Dies kann man quantitativ ausführen.

Die Antwort auf einen Klick an einer festen Stelle x sei gegeben durch die Feurungsrate $f^w(t)$. Die Antwort auf ein kontinuierliches Gräusch $p(t)$ ist dann:

$$f^p(\tau) = \int dt\, p(t) f^w(\tau - t) \tag{4.1}$$

Bilden wir nun die Faltung

$$c(\tau) = \int dt'\, p(t') f^p(\tau - t') = \int\int dt\, dt'\, p(t) p(t') f^w(\tau + t' - t) \tag{4.2}$$

und führen die neuen Variablen

$$t_1 = t' - t, \quad t_2 = (t' + t)/2 \tag{4.3}$$

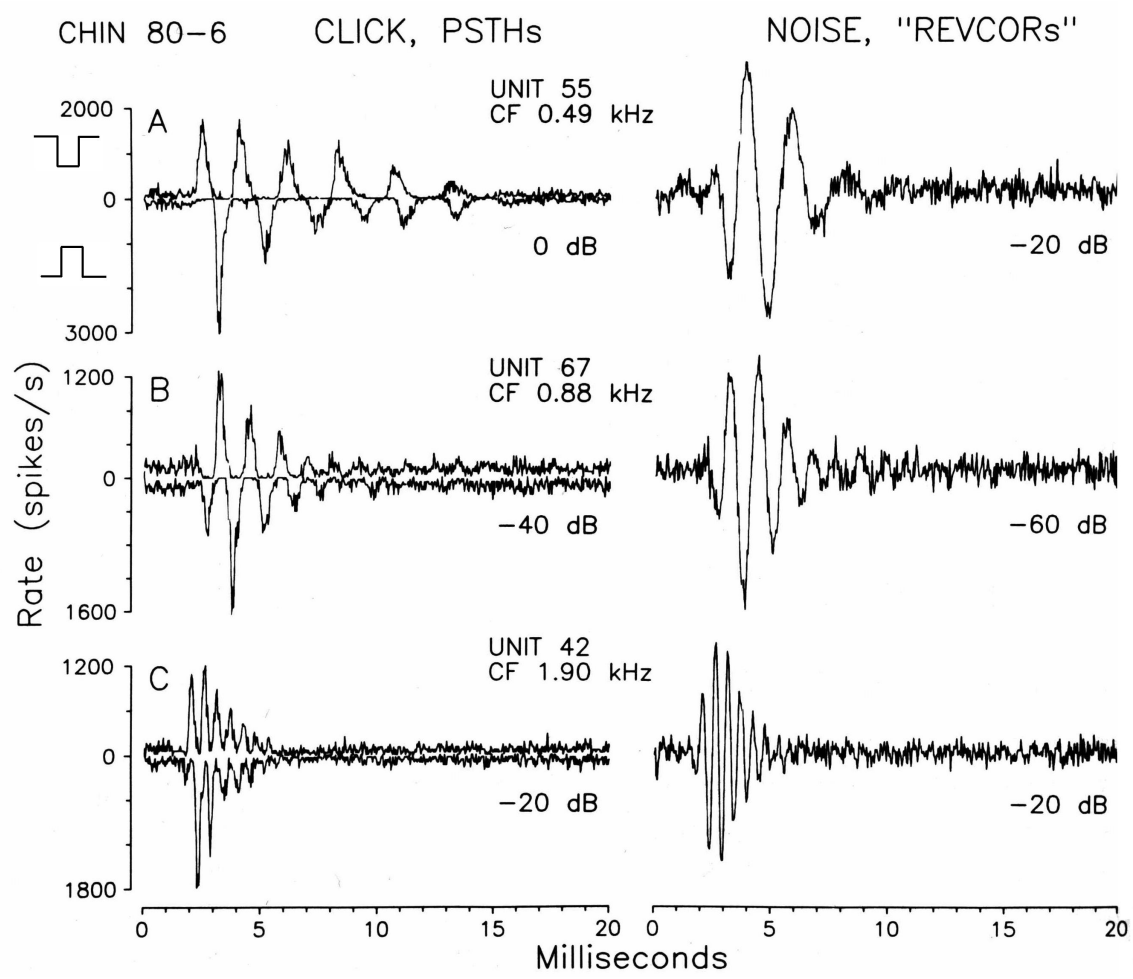


Abbildung 4.8: ,,Nachklingeln" der Feurungsraten an verschiedenen Stellen der BM, s. Text**Links** Antwort auf Depressions und Kompressionsklick,**Rechts** Korrelationsfunktion von Rauschen und Feurungsrate; nach [**?**]

dann ergibt sich:

$$c(\tau) == \int\int dt_1\, dt_2\, p(t_2 + t_1/2) p(t_2 - t_1/2) f^w(\tau + t_1) \tag{4.4}$$

Ist $p(t)$ ein Rauschen, dann gilt

$$\int dt\, p(t)\, p(t + \tau) = \delta(\tau) \tag{4.5}$$

und wir bekommen

$$c(\tau) = f^w(\tau) \tag{4.6}$$

Da im Rauschen sich Kompressionen und Dekompressionen zufällig abwechseln, wird $c(tau)$ die Überlagerung von den beiden wavelets darstellen.

4.1.3 Nichtlineare Effekte

Maskierung

Auch die Effekte der Maskierung lassen sich an den Aktionspotentialen im 8. Nerv nachweisen. In Abb. 4.9 sind die Zacken der Aktionspotentiale für drei verschiedene Fälle an einem Neuron mit der CF = 360 Hz dargestellt. Bild A ist die Antwort auf einen kontinuierlichen Ton von 360 Hz, bei 40 dB SPL, Bild B auf einen Pieps mit 720 Hz. In Bild C schliesslich ist die Reaktion auf die beiden gleichzeitig angebotenen Töne dargestellt. Man sieht deutlich, dass der Pieps die Feurungsrate stark unterdrückt. Dass dies praktisch instantan geschieht spricht dafür, dass diese Unterdrückung auf die BM Mechanik zurückzuführen ist, und nicht auf einen nachträglichen Rückkopplungseffekt.

4.1.4 Kombinationstöne

Kombinationstöne sind nichtlineare Effekte. Haben wir die Überlagerung von zwei Sinusrtönen, z. B. $p(t) = \sin(2\pi\nu_1 t) + \sin(2\pi\nu_2 t)$ so hören wir, bei genügendhohem Schallpegel Kombinationstöne, dies sind spektrale Komponenten, die in höheren Potenzen von $p(t)$ enthalten sind. So enthält z.B. $p(t)^2$ die Komponente mit der Frequenz $\nu_1 - \nu_2$, die bei genügend hohen Schallpegeln sehr deutlich zu hören ist. Auch Kombinationstöne sind in den Aktionspotentialen nachzuweisen, was allerdings wenig verwunderlich ist. In Abb. 4.10 ist dies deutlich zu sehen.

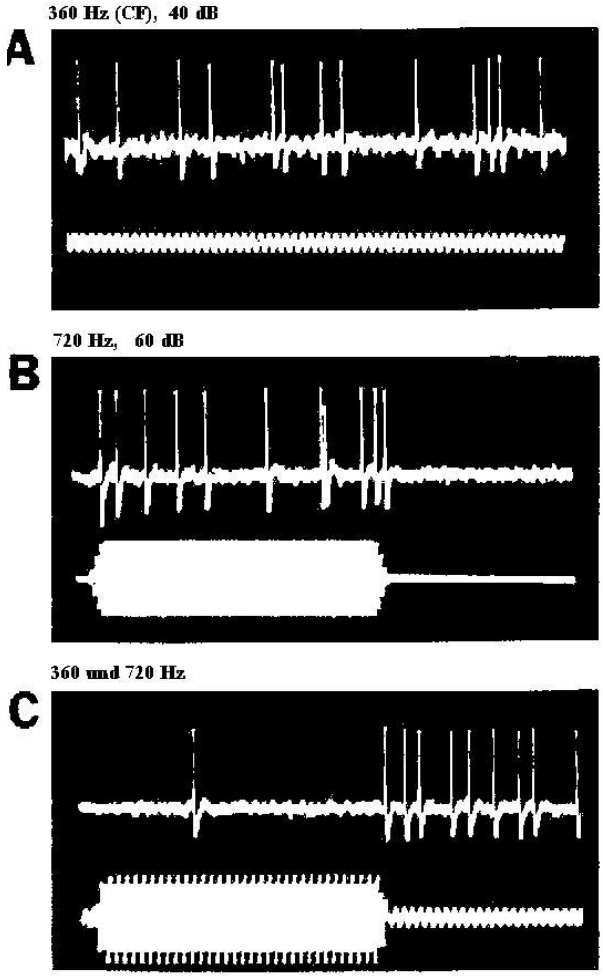


Abbildung 4.9: Maskierung bei den Aktionspotentialen, s. Text; nach [**?**]

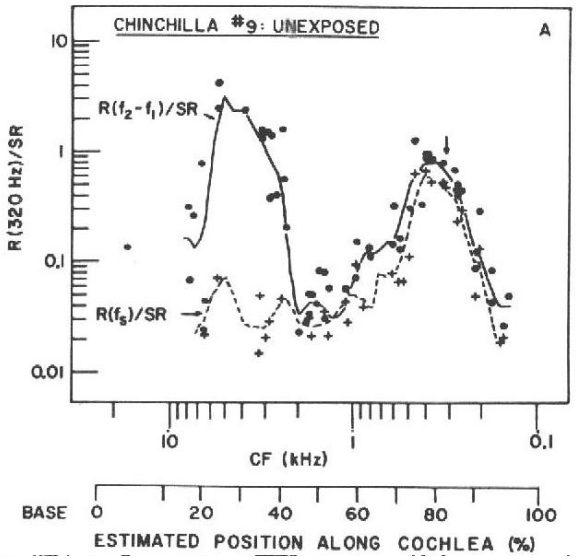


Abbildung 4.10: Relative Feurungsrate von Nerven entlang der BM eines Chinchilla einmal für einen Zweitonkomplex mit 3680 und 4000 Hz (Kreise, volle Linie) und für einen Sinuston von 320 Hz (Kreuze, gestrichelte Linie); nach [?]; Tonbeispiel `praats/diference-tone`

4.2 Einfluss der Haarzellen

Da die afferenten Nertven von den inneren Haarzellen ausgehen, ist es nicht verwunderlich, dass wenn innere Haarzellen abgetötet werden, die spontane Feurung der entsprechenden afferenten Nerven aufhört und akustische Reize keine Aktionspotentiale auslösen. Bei einer Schädigung der inneren Haarzellen wird die Schwelle der tuning Kurven erhöht (Schwerhörigkeit).

Weniger evident und interessanter ist der Effekt der äusseren Haarzellen, die ja gar keine Aktionspotentiale aussenden. Werden äussere Haarzellen partiell zerstört, so steigt die Schwelle der tuning Kurve für die afferenten Nerven an, die an inneren Haarzellen nahe den geschädigten äusseren Haarzellen enden, in Gebieten weit davon entfernt ist die Schwelle wenig geändert. Ausserdem geht die Frequaenzselektivität weitgehend verloren. Dies ist in Abb. 4.11 deutlich zu sehen. Die äusseren Haarzellen eines Chinchillas sind nahe dem Apex zerstört und dementsprechend tritt eine starke Reduktion der Hörfähigkeit bei tiefen Frequenzen ein. Dies wird sowohl aus den neuronalen tuning Kurven als auch aus dem durch Verhalten gemessenem Audiogramm (dicke Linie) sichtbar. Bei hohen Frequenzen sind die tuning Kurven normal, doch bei niederen geht die Frequaenzselektivität verloren. Da die Feurungsraten des 8. Nerves die Mechanik der Cochlea recht genau wiedergeben, ist dies ein sehr wichtiges Indiz dafür, dass die äusseren Haarzellen einen grossen Einfluss auf die Mechanik der Cochlea (Entdämpfung) haben.

Die Einflüsse des Haarzellenverlustes sind in Abb. 4.12 schematisch dargestellt, ebenfalls dargestellt ist die berechnete tuning-Kurve der BM, wobei der Verlust der äusseren Haarzellen (rote Kurve) durch eine Vegrösserung der Dämpfung durch einen Faktor 667 simuliert wird.

In die gleiche Richtung weist auch, dass die Maskierung in Zweitonkomplexen durch Verlust der äusseren Haarzellen eliminiert wird. Auch dies ist ein wichtiges Indiz dafür, dass die Nichtlinearität (z.B. amplitudenabhängige Entdämpfung) durch die äusseren Haarzellen verursacht wird.

Ein weiterer Hinweis dafür, dass die äusseren Haarzellen die Mechanik der Cochlea beeinflussen besteht darin, dass elektrische Reize die auf die efferenten Nerven des SOC ausgeübt werden, die Antworten der afferenten Nerven nahe der CF beeinflussen.

4.2.1 Otoakustische Emission

Das Ohr empfängt nicht nur Töne, sondern sendet auch welche aus, und zwar sowohl spontane als auch evozierte. Während die spontanen direkt nachgewiesen werden können, s. Abb. 4.13,links muss man zum Nachweis der evozierten rafinierte Methoden anwenden. Man kann

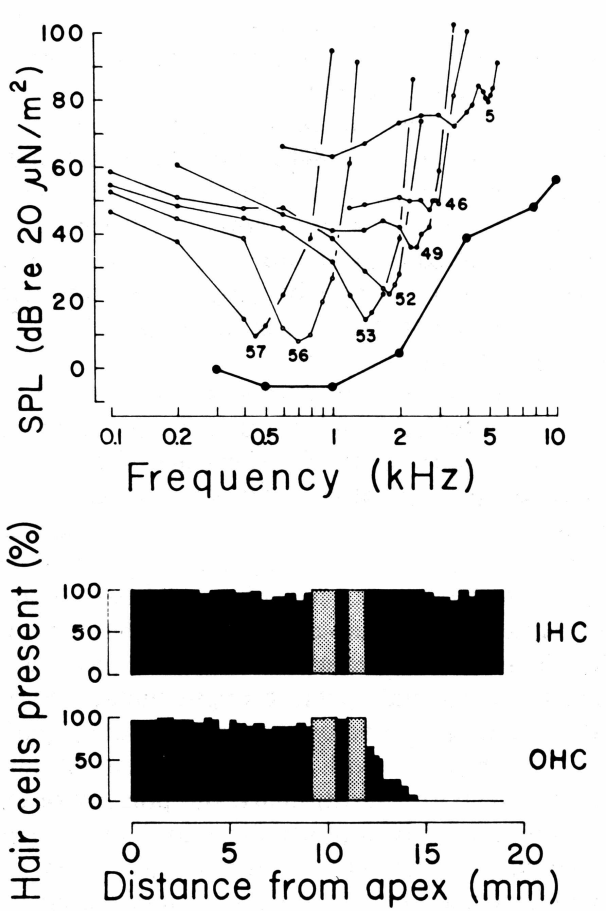


Abbildung 4.11: Einfluss des Verlustes eines Teils der äusseren Haarzellen, s. Text;nach [**?**]

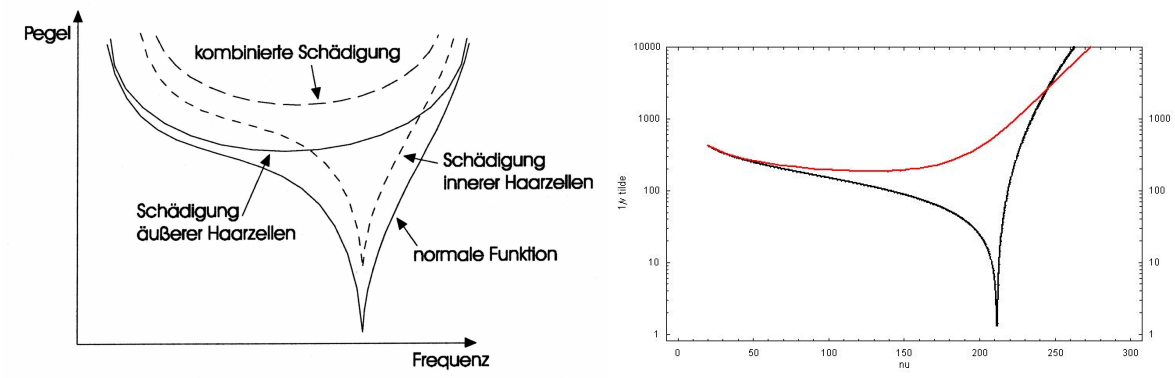


Abbildung 4.12: Links: Schematische Darstellung des Einflusses von Schädigungen der Haarzellen auf eine neuronale tuning Kurve, Rechts: Tuning Kurve auf der BM, schwarz mit Dämpfungsfaktor $\delta = 0.0015$, rot mit $\delta = 0.4$

z.B. ein kurzes Tonsignal eingeben, und danach Schallemmissionen des Ohres nachweisen, siehe Abb. 4.13, rechts.

Eine andere Möglichkeit besteht in der Differenzverstärkung, die darauf beruht, dass otoakustische Emissionen mit zunehmendender Schallstärke saturieren, während der direkt nachgewiesene Schall dies nicht tut.

Auch die otoakustischen Emissionen sind ein Hinweis auf die aktive Rolle der äusseren Haarzellen. Bewirken sie tatsächlich eine Entdämpfung, dann heisst dies, dass sie tatsächlich aktiv zur Bewegung der BM beitragen, wie man aus der folgenden Gleichung sieht:

$$m\partial_t^2 Y = -SY - (R - f(Y))\partial_t Y + F$$

Hier ist Y die Auslenkung der BM, S die Steifigkeit, F die Kraft durch den Druck in der Lymphe, R ist die passive Dämpfung, die durch die aktive Kraft $+f(Y)\partial_t Y$ veringert (entdämpft) wird. Diese aktive Kraft, die auf die BM wirkt, führt ebenfalls zu einer Wanderwelle in der Cochlea, die zum runden Fenster läuft und dann am Trommelfell Schall erzeugt. Da diese aktive Kraft nichtlinear ist ($f(Y)$ geht wahrscheinlich gegen Null mit $Y \to \infty$) enstehen auch Kombinationstöne in der otoaksutischen Emission, insbesondere der kubische mit $\nu = 2\nu_1 - \nu_2$.

Ich moechte nun ganz kurz die Hörbahn, d.h. den Weg von der BM zum Zentrum der höheren Verarbeitung akustischer Signale, dem Hörkortext schildern. Bevor wir zu einzelnen Punkten kommen, möchte ich noch einmal eine funktionale Karte des menschlichen Gehirns zeigen, Abb. 4.14. Es fällt auf, das der Hörkortex sehr nahe am Ohr liegt und inmitten des hinteren Sprachzentrums.

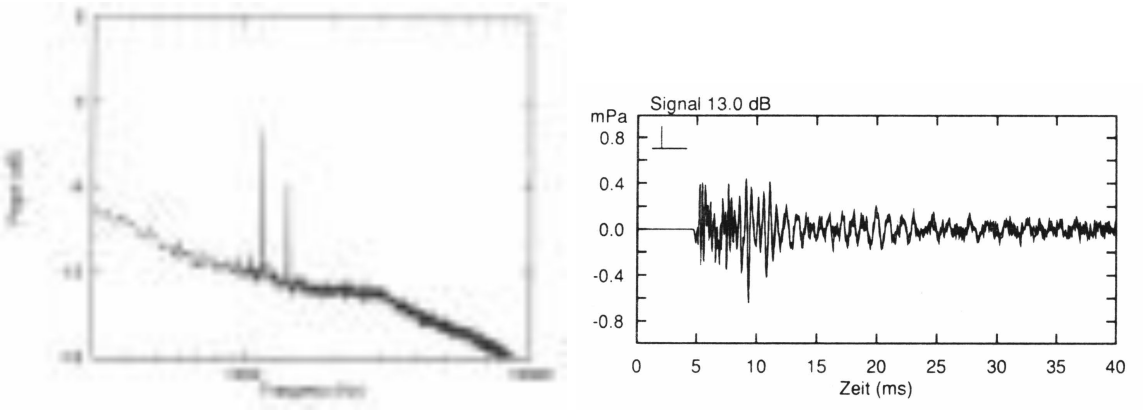


Abbildung 4.13: **Links:** Spektrum einer spontanen otoakustischen Emission. **Rechts:**Zeitlicher Verlauf einer durch einen Klick evozierten otoakustischen Emission.

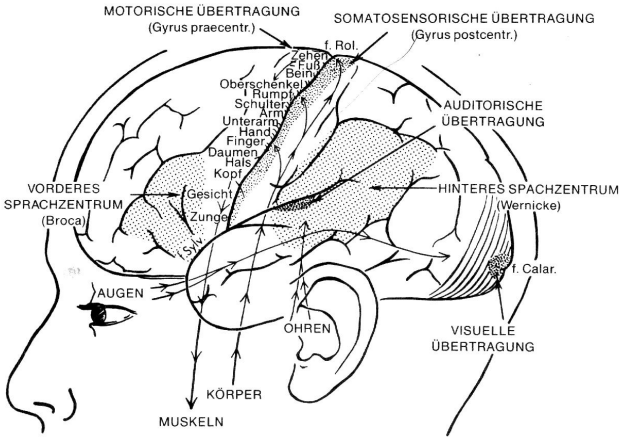


Abbildung 4.14: Funktionale Karte des menschlichen Gehirns, aus [**?**]

Kapitel 5

Pitch perception

The physiological and first neurophysiological stage of sound perception is at least qualitatively understood: The sound waves are transformed in travelling waves in the BM and the motion of the BM is transformed into firing of the auditory nerve. The position of the maximal amplitude of BM movement depends on the frequency, low frequencies generate a wave which travels through nearly the whole cochlea and excite the BM near the helicotrma, high frequencies travel only a short distance and excite the BM bear the oval window (finestra vestibuli). The firing of the nerves is for frequencies below ca 2000 Hz in phase with the BM movement, the nerves fire if the MB moves from the scala tympani to the scala vestibuli. In that case the information on the frequency is doubly encoded: In the position of the BM (spectral information, tonotopic) and in the frequency of the firing burst. There is still a vivid discussion, which of the both information is most important for pitch perception.

The following example shows, that the situation is indeed intricate and that the question of "higher" and "lower" pitch is – at least in some cases – not uniquely determined. This is shown in the following example, going back to Smorenburg and developed by Peter Schneider here in HD. `186-155-teil.wav`. In average one half of the subjects say that the pitch rises, the other half that is falls.

5.1 Spectral theory

In the 40s of the 19th century was a intensive discussion between Ohm and Seebeck about the "essence of the sound". Seebeck had worked with the newly developed siren and came to the conclusion that periodicity determines the pitch, whereas Ohm, applying the shortly before

developed theory of Fourier, claimed that the spectral properties are essential. At the time, Seebeck seemed to be the winner, but then Helmholtz published his classical book "Die Lehre von den Tonempfindungen", where he defended and "proved" Ohms Theory. Let me first state the essence of the Ohm-Helmholtz theory:

A "Klang" consists of a sound with a harmonic spectrum. The lowest harmonic determines the pitch, the higher ones the timbre of the sound. Few strong harmonics for instance give a soft sound (flute e.g.), many strong harmonics yield a sharp, a metallic sound, see and hear.(`praats-bsp/flute-spinett)`

Sounds with distinct non-harmonic spectra play in western a minor role, though they may set very important accents (cymbals, e.g.). They are typically produced by two-dimensional generating bodies (Drums, cymbals) in contrast to one-dimensional generators (strings, air-columns, vibrating tongues).

Noise finally is sound with no discrete spectrum. In white noise the spectrum is approximately constant over the full range of audible frequencies, see fig. **??**. Noise occurs frequently in nature, for instant in the rustling of the wind. It it can be generated electronically by random numbers.

A click, which is very short in time, has a also a broad spectrum, in the limiting case of a click with a $\delta(t)$-like pressure, the spectrum is constant. But both noise and (single) clicks have in common, that they have no pitch. The spectra of continuous noise and a click are however different if one considers not only thw absolute value of the FT (power spectrum) but also the phases. The FT of a δ click has a constant phase,

$$\int_{-\infty}^{\infty} dt\, \delta(t - t_0)\, e^{i\omega t} = e^{i\omega t_0} \tag{5.1}$$

whereas in continuous noise the pase varies stochastically.

Harmonic sounds have a periodic SP curve, see figure 5.1 flute and virginal. The sound pressure of the cymbal and the white noise look both erratic, but the spectrum reveals the qualitative difference, which is very audible, indeed.

In speech sounds with harmonic and with continuous spectrum play an important role. In figure **??** the spectrogram of the vowel "e" is shown, the harmonics are clearly visible. In the sung tone, the spectra are, in spite of the vibrato, sharper than in the spoken tone. There are certain regions of the spectrum of a vowel which are particularly enhanced. These regions are called formants. They are (approximately) independent of the pitch of the vowel. These formants make the difference between the different vowels. In figure 5.3 the spectrograms of the vowel 'u' and 'i' are displayed. For a speaker untrained in singing the vowel u has practically no formants, but for a singer, who is trained to enhance the formants, the formant spectrum of the "u" is as rich as as that of the "i". Nevertheless the vowels are clearly recognized in both

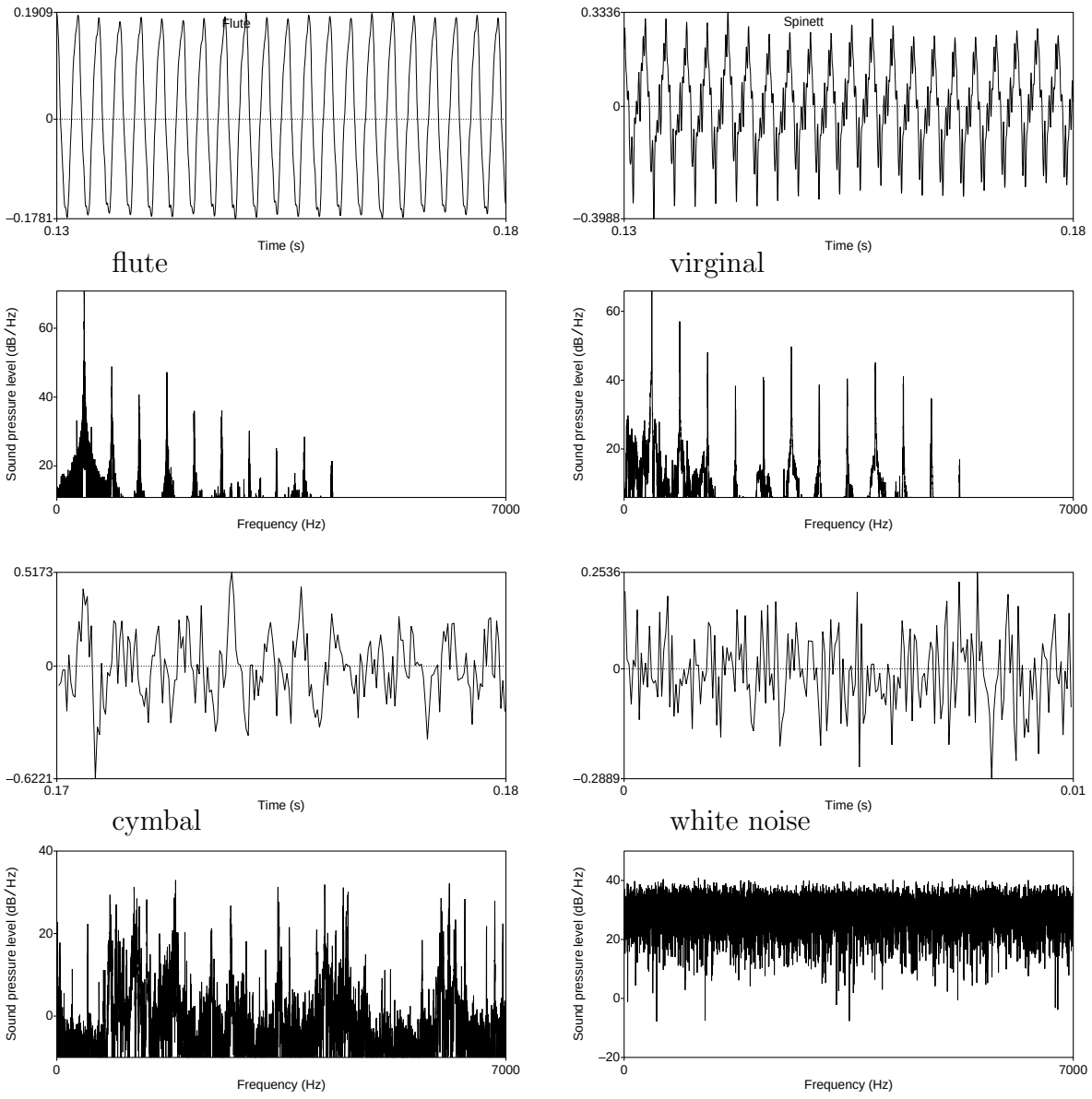


Abbildung 5.1: SP and spectrum of a a) flute, b) virginal, c) cymbal and d) white noise

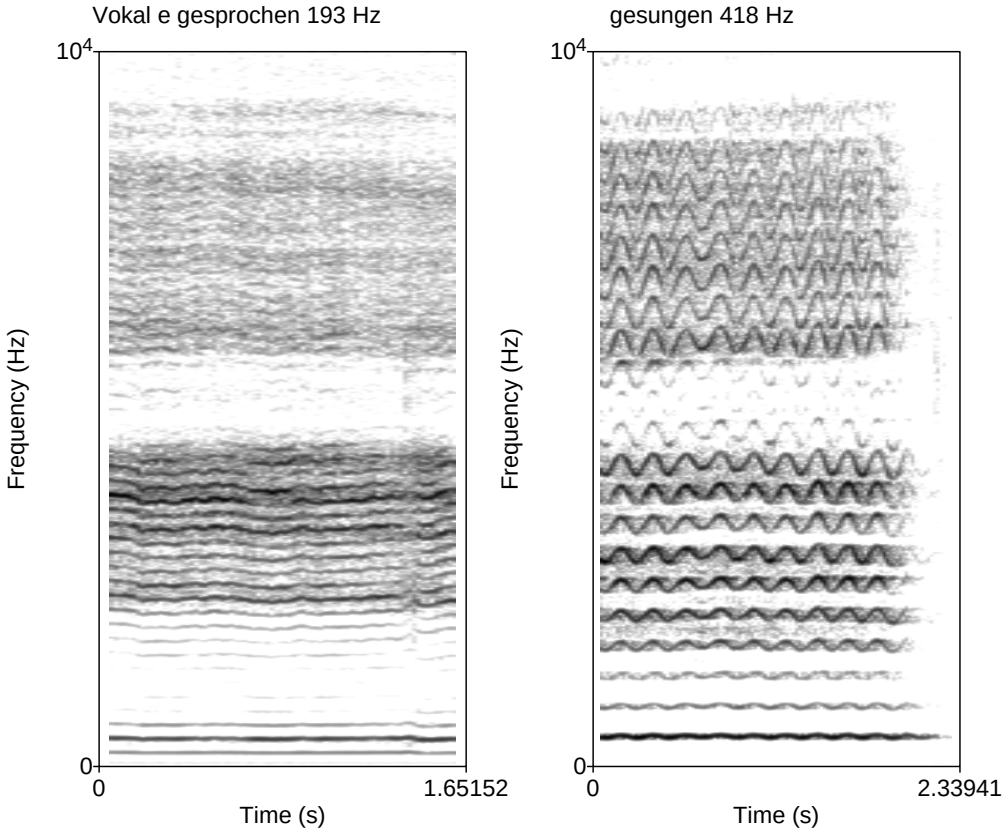


Abbildung 5.2: Spectrogram of the vowel ‘e’ with fundamental frequency of 193 Hz (spoken) and 418 Hz (sung) from an amateur singer (lady).

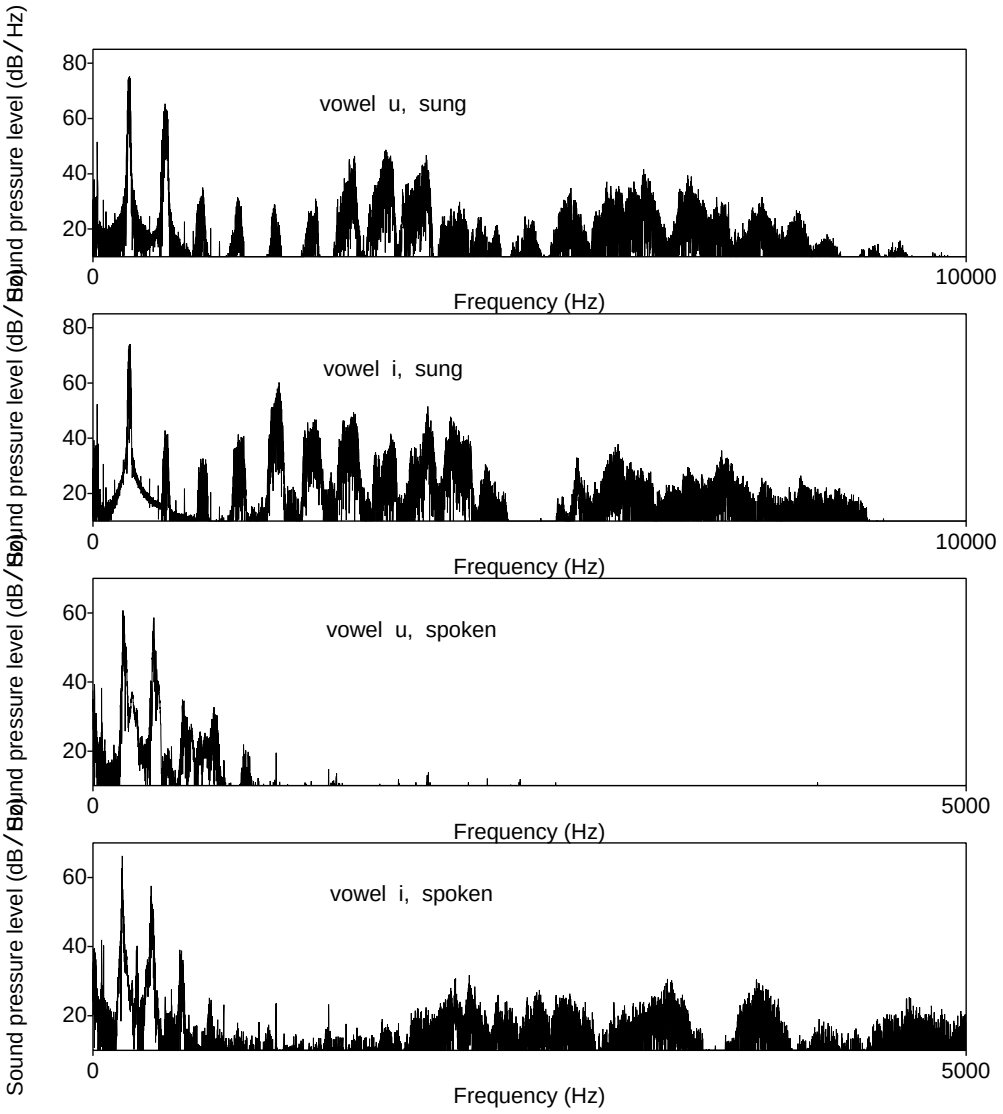


Abbildung 5.3: Spectra of the vowels u and i sung by a amateur singer (lady) and spoken by a non-singer (male)

cases, `praat-bsp/vokale` As in human voice, the formants of a musical instrument are decisive for it's timbre and especially it's quality.

In figure 5.5 the spectrogram and cuts of the SP of the word "Schnee" are displayed (very

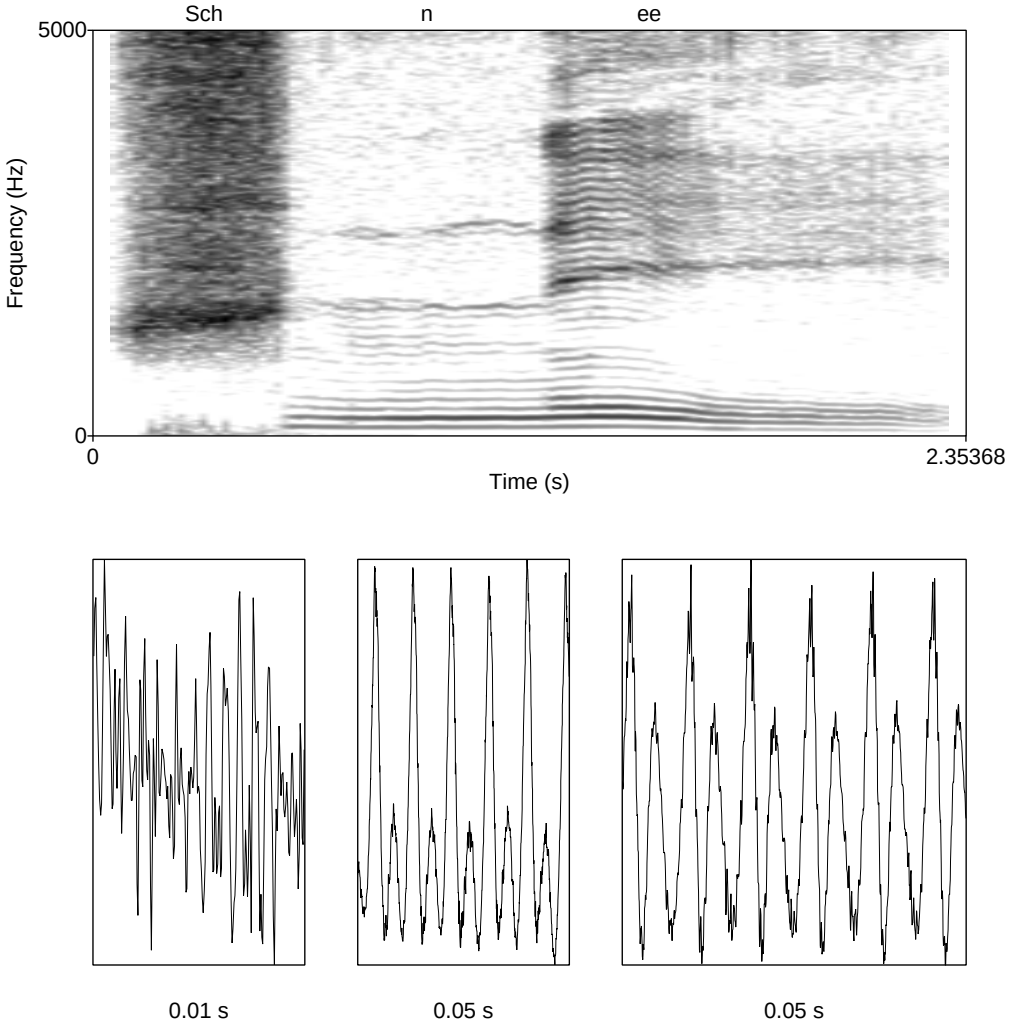


Abbildung 5.4: Spectrogram and cuts of the SP of the word “Schnee”

unnaturally spoken Sch...sch n...n e...e). Whereas the “Sch” has a continuous spectrum, the “n” is as periodic as the vowel e, but has a much poorer formant structure.

The consonant t has also a continuous spectrum, but it is in nature much closer to a click than a longer lasting noise.

The fact that a musical sound contains several partial sounds was noticed presumably first by

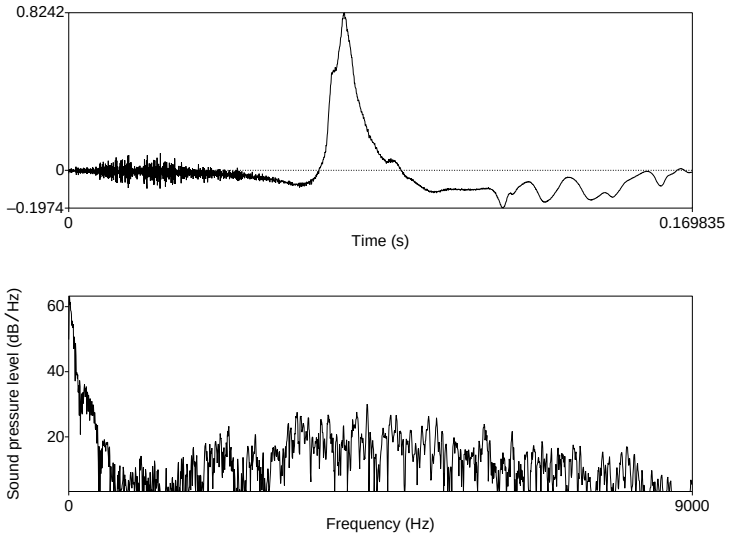


Abbildung 5.5: Spectrogram and SP of the consonant t

Descartes. It is very clearly stated in D'Alembert's "Elemens de musique" and is the basis of the theory of harmony as established by Rameau, which was prevailing until the 20th century. One can sharpen its attention, by first listening to an isolated harmonic and the try to find it in the complex tone. If one adds to an incomplete complex sound the missing partial, ome first hears this partial quite distinctly, but after a second or so, it melts with the others to a homogeneous sound.

5.1.1 The missing fundamental

If one cuts one or several harmonics from a complex harmonic sound, only the timbre changes. See `praats-bsp/missing-harmonic`, completely in accordance with spectral theory. Change of timbre by changing the spectral composition is the basis of instrumental virtuosity (flageolet, e.g.), of the registers of the organon and of course of the synthesizers.

But what happens, if we cut not an upper harmonic, which is anyhow only responsible for the timbre, but the fundamental, which is responsible for the pitch? If you cut a few, only the timbre changes, if you cut more, than also you may jump by an octave. `praats-bsp/missing-fundamentals`

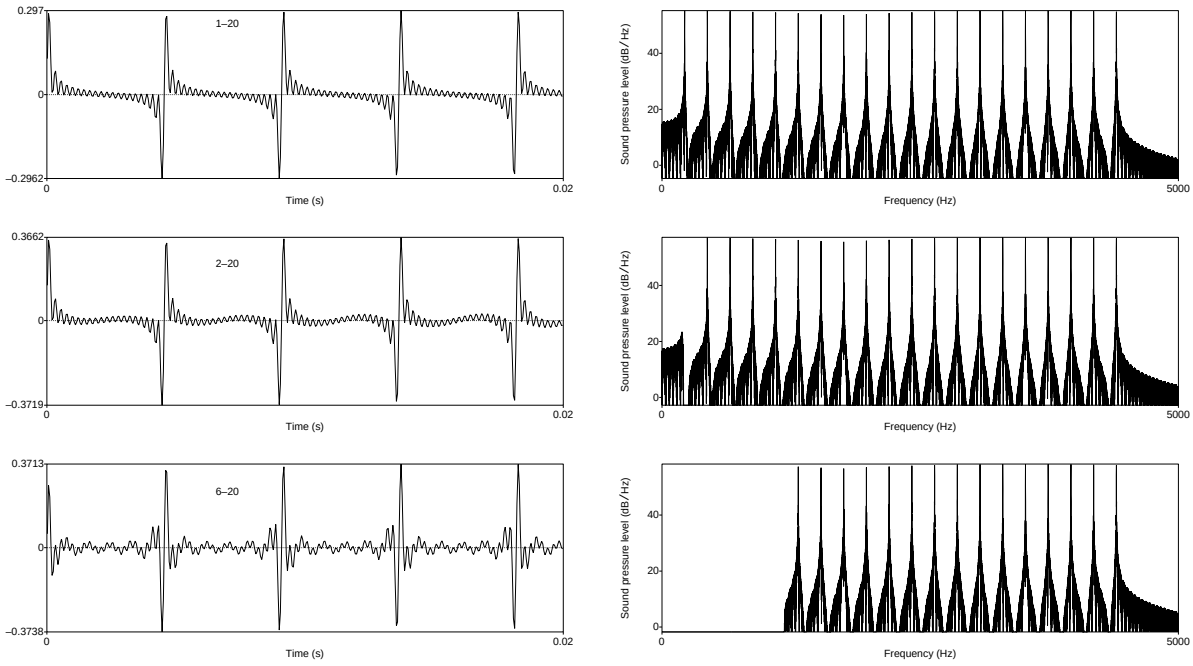


Abbildung 5.6: SPL and spectrum missing fundamentals

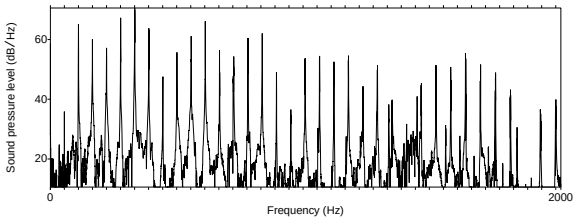


Abbildung 5.7: Spectrum of the A(55) of a Steinway D

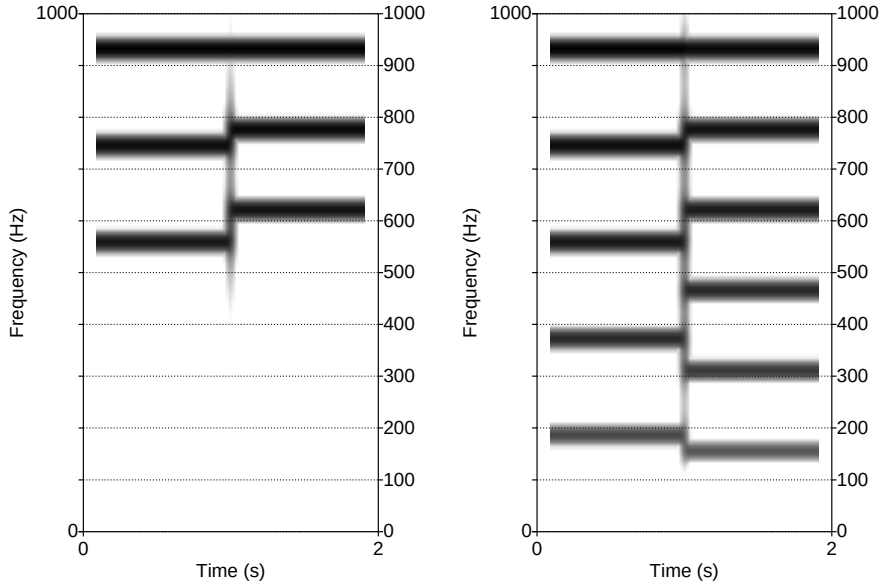


Abbildung 5.8: Spectrogram of the Smorenburg test.

This effect even plays a role with musical instruments. In a grand piano, even a Boesendorfer Imperial, the fundamental tone is missing in the lowest tones. Even at the second A(55) the fundamental tone in a Steinway D is practically not present (30 dB suppressed), see fig. **??**.

We can now better understand the example, which I played at the beginning. It goes back to Smorenburg and was essentially refined by Peter Schneider. In that example the fundamental of the second tone (155 Hz) is lower than that of the first (186 Hz), but all partials of the first have frequencies which are lower or are equal to those of the second.

The periodicity does however not change, if partial tones, even the lowest ones, are cut, and this gave much support to the temporal pitch theories.

5.1.2 Spectral vs temporal theories

As can be seen from figure 5.6, the character of the tones changes substantially, though the difference in the SP is not so drastic. On the other hand, rather drastic changes in the temporal behaviour do hardly influence the timbre, as can be seen e.g. from the Schroeder phase complex.

A Schroeder phase complex is given by

$$p(t) = \sum_{m=n_1}^{n_2} \sin(\left(2\pi m\nu_0 t + \pi \frac{m(m-1)}{n_2 - n_1 + 1}\right)$$

In fig. **??** the SPL of a Schroeder phase complex is displayed, with $\nu_0 = 22$ Hz, $n1 = 1;\ n2 = 10$. Below is shown the same complex with no phase, though the fine structure of the time behaviour looks very different, one does hear no difference of the two tones. This phase independece of complex sounds, which Helmholtz had shown in really ingenious experiments, was one of his main arguments for the spectral theory. The phase independence (or near phase independence) was for him one of the strongest arguments for the spectral pitch theory.

Another argument brought frequently forward in favour of the spectral theory is the chopped noise. If noise is presented in chops, it nevertheless stays noise and the frequency of the bursts leads to no pitch. In figure 5.11 the chopped noise:

$$p(t) = \sin^8(2\pi\nu_0 t)\phi(t) \tag{5.2}$$

with $\phi(t)$ white noise and $nu_0 = 200$ Hz. We see the flat spectrum and hear the noise, as can be heard in `praats-bsp/chopped-noise`.

Also presented is the SP curve

$$p(t) = \sin^8(2\pi\nu_0 t)\sin(2\pi\nu_1 t) \tag{5.3}$$

with $\nu_0 = 200$ and $\nu_1 = 5000$ Hz. One hears distinctly a pitch, perhaps some octave of 200 Hz. The spectrum is indeed that of a incomplete harmonic sound with fundamental frequency of 200 Hz.

The comparison is however not fair. The special form of the SP curve in fig. 5.11 is very sensitive to the phase, if there is some distortion of the phases, the curve might look quite differently. In fig. **??** the curve, as given electronically to the loud speaker and the SP curve as measured by a microphone in the distance of ca 1 m from the loudspeaker is shown. It can be seen that from the chopping little is left in the recorded SP curve.

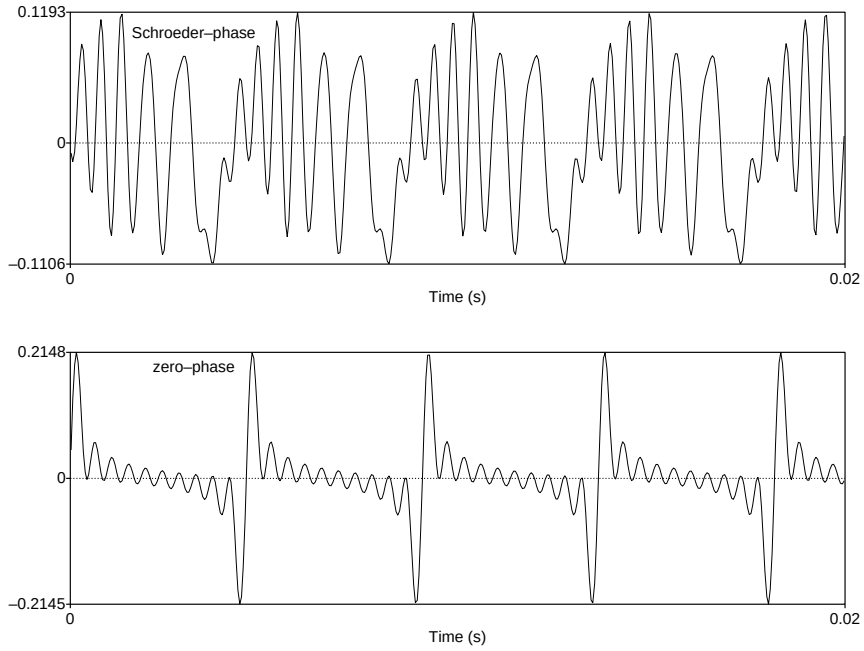


Abbildung 5.9: Complex sound with 10 partials with Schroeder phase abd zero phase.

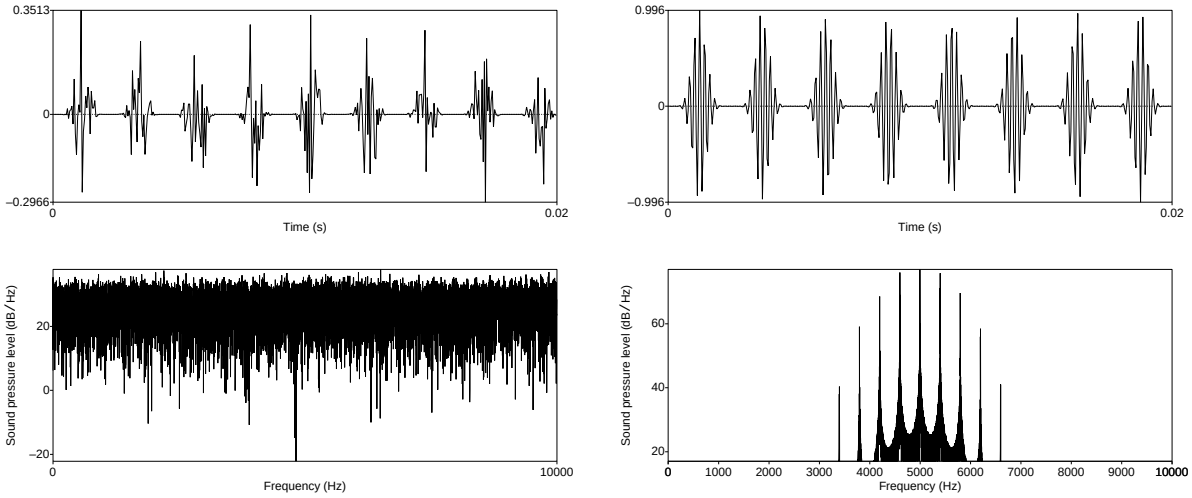


Abbildung 5.10: Chopped noise and a chopped sinusoidal tone of 5000 Hz. The chopping is in both cases made by the envelope $\sin^8(2\pi 200t)$

.

5.1.3 Pitch extraction in spectral theory

The "missing" fundamental tone as difference tone

. The fact that a missing fundamental does not change the pitch was not unknown. The general assumption was, that the missing fundamental is created by nonlinear distortions. In fact, if one plays for instance the tone created from the 2 to 20th partial very loud and records it, the linear distortions create physically the fundamental tone which is visible in the spectrum. The reason for that can be shown very easily.

Be $p(t) = \sin(\omega_1 t) + \sin(\omega_2 t)$ a complex tone. The distorted signal can be expanded in a power series:

$$D[p(t)] = \alpha p(t) + \beta(p(t))^2 + \gamma(p(t))^3$$

The quadratic distortion term is

$$\begin{aligned} p^2(t) &= \Big(\sin(\omega_1 t) + \sin(\omega t_2)\Big)^2 \\ &= -\frac{1}{4}\Big(e^{\omega_1 t} + e^{-i\omega_1 t} + e^{i\omega_2 t} - e^{-i\omega_2 t}\Big)^2 \\ &= e^{2i\omega_1 t} + e^{-2i\omega_1 t} + e^{2i\omega_2 t} + e^{-2i\omega_2 t} + 2e^{i(\omega_1+\omega_2)t} + 2e^{-i(\omega_1+\omega_2)t} \end{aligned} \tag{5.4}$$

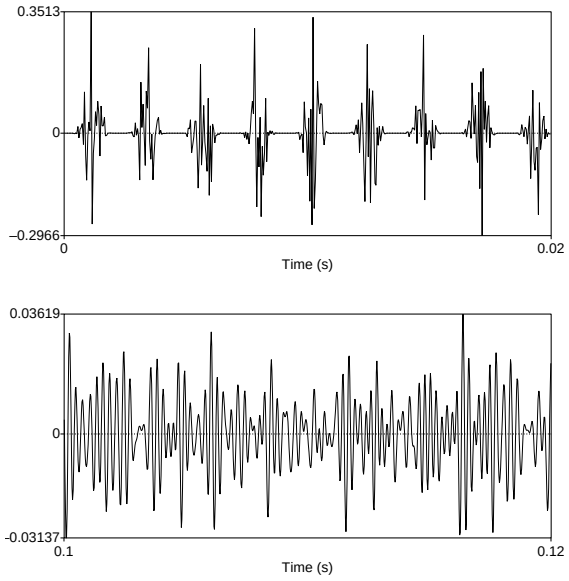


Abbildung 5.11: Chopped noise as given to the loudspeaker (top) and as recorded by a microphone in a distance of 1 m from the loudspeaker(bottom).)
.

$$-2e^{i(\omega_1-\omega_2)t} + 2e^{-i(\omega_1-\omega_2)t} - 4 \tag{5.5}$$

If $\omega_1 = n\omega_0$ and $\omega_2 = (n+1)\omega_0$ then $p(t)$ is a harmonic complex tone with missing fundamental ω_0. But in the distorted signal the fundamental is physically present through $2\ e^{i(\omega_1-\omega_2)t} + 2e^{-i(\omega_1-\omega_2)t} = 2\cos(\omega_0 t)$

The shifted harmonic complex

Schouten, who was the first in modern times to challenge the spectral theory, could however show by ingenious experiments, that this is not the case. One of his decisive experiments is particularly simple. Each partial tone of harmonic complex with missing fundamentals is shifted by a fixed frequency δ, that is the partials have now the frequencies

$$\{n_1\nu_0 + \delta, (n_1+1)\nu_0 + \delta, \ldots (n_1+k_1)\nu_0 + \delta\} \tag{5.6}$$

A nonlinear distortion of this tone leads to an interference tone with frequency ν_0. So, if the interference tone replaces the missing fundamental, then the pitch of the tone complex should be independent of the shift δ. But this is not the case, as can be heard clearly `praats-bsp/pitchshift`. In the following we shall consider as reference the harmonic complex the signal:

$$p_0(t) = \sum_{k=n_1}^{n_1+k_1} \sin(2\pi\nu_0\, t) \tag{5.7}$$

with $n_1 = 4$, $k_1 = 11$, $nu_0 = 200$. In fig. 5.12 this signal $p_0(t)$ (red) is displayed as well as the signal $p_\delta(t)$ (black), where the frequency of each partial is shifted by $\delta = 40$ Hz. As can be seen, the shifted signal is nearly harmonic and Schouten argued that the approximate perodicity determines the pitch.

Spectral pitch extractor

This is however a bit vague, as it was already in the time of Seebeck and Ohm, and the spectral theory was fighting back with a more sophisticated pitch extractor.We e have seen that the simplest model for a pitch determination, namely the difference between the partial frequencies, works only for harmonic tones. Therefore Goldstein proposed a more refined processor, based an an optimal adaption of the presented tone complex by an harmonic tone. More precisely, if the frequencies of the presented tone are $\nu_{n_1}, \nu_{n_1+1}, \ldots \nu_{n_1+k_1}$, then look for an harmonic tone

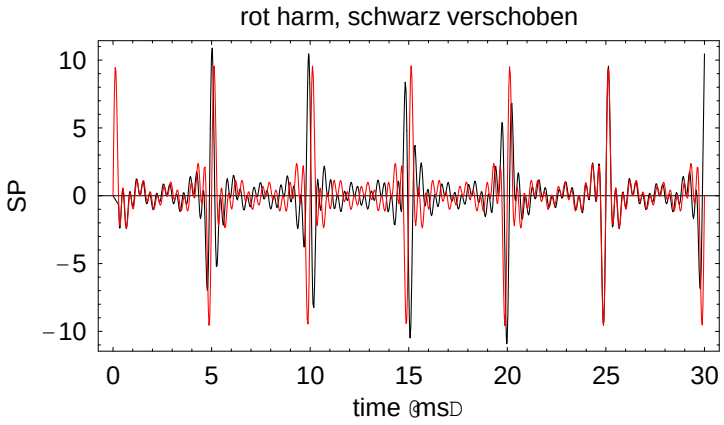


Abbildung 5.12: The SP of the harmonic signal $p_0(t)$ (red) and the shifted signal $p_\delta(t)$ (black) with $\delta = 40$ Hz

.

with fundamental frequency ν_g such that the difference

$$\Delta_g = \sum_{k=n_1}^{n_1+k_1} \Big(\nu_{n_1} - (m_1 + k_1 - 1)\nu_g\Big)^2 \tag{5.8}$$

is minimal. In figure 5.13 the SP and the spectral lines of the shifted complex (black) as well as of the optimized harmonic tone are displayed as a function of the frequency ν, blue for $m_1 = n_1 = 4$ and magenta for $m_1 = n_1 + 1 = 5$. In fig. 5.14 the difference Δ_g, see (5.8) is displayed as function of ν_0. As expected, the minimum is lower for the choice $m_1 = n_1$, in which case $\nu_g > \nu_0$. If however the shift δ is bigger than $\nu_0/2$, the choice $m_1 = n_1 + 1$, with $\nu_g < \nu_0$ yields a better approximation. This implies that near the shift $\delta = 100$ Hz the resulting harmonic approximation is ambiguous.

In the simple case of a shifted harmonic tone the minimum of Δ can be determined analytically:

$$\Delta_g(\nu_0) = -\frac{3\,\delta\,k_1 - 6\,\delta\,m_1 - k - 1\,\nu_0 - 2\,k_1^2\,\nu_0 - 3\,k_1\,m_1\,\nu_0 - 3\,k_1\,n_1\,\nu_0 - 6\,m_1\,n_1\,\nu_0}{k_1 + 2\,k_1^2 + 6\,k_1\,m_1 + 6\,m_1^2} \tag{5.9}$$

In fig. 5.15 the optimal frequencies ν_g are displayed as function of the shift δ, blue for $m_1 = 4$, magenta for $m_1 = 5$.

Apart from some subtleties this processor works quite well, an apparent improvement would

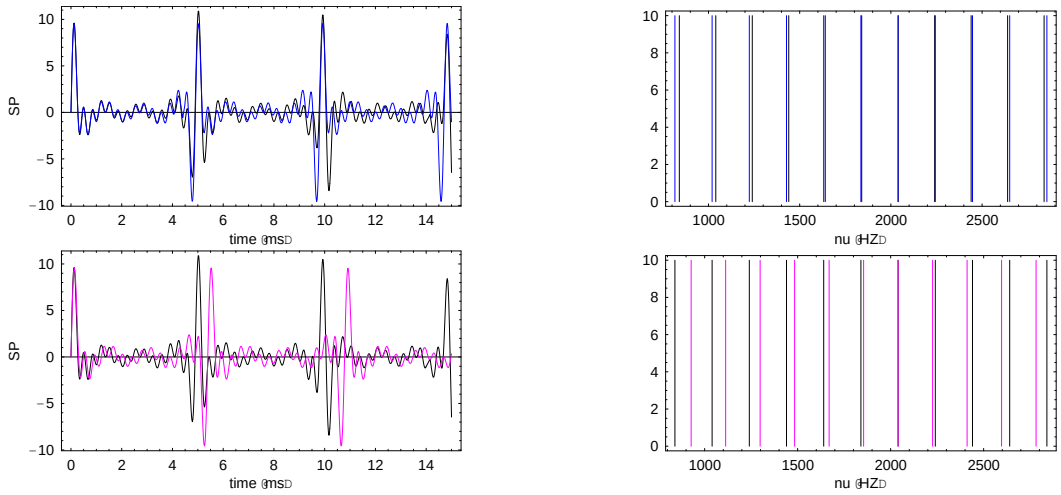


Abbildung 5.13: SP and the spectral lines of the shifted complex (black) as well as of the optimized harmonic tone, blue for $m_1 = n_1 = 4$ and magenta for $m_1 = n_1 + 1 = 5$.

.

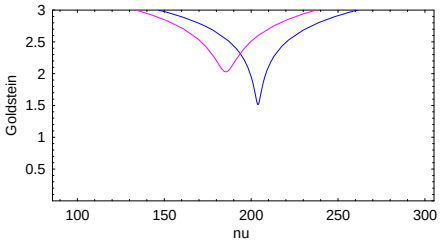


Abbildung 5.14: The difference Δ_g, (5.8).

.

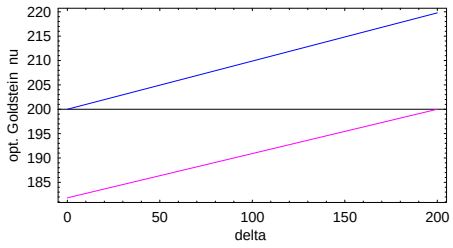


Abbildung 5.15: The optimal frequencies ν_g as function of the shift δ, blue for $m_1 = 4$, magenta for $m_1 = 5$.

.

be to minimize not the algebraic difference between the spectral lines, but the distance on the BM.

The spectral theory relies on the resolution of the partial tones on the BM. There is some discussion about this resolution, but for very high partials a resolution is very improbable. Nevertheless one can evidently recognize the fundamental tone, even if one cuts all partials between the tenth one. Even if one plays at a low key and cuts all below the 20th one, one sill can discern clearly the beginning of the theme of the *Kunst der Fuge*.

As the spectral theory with its pitch extractors, also the temporal theory can be refined and come to results which are in mant cases very similar to the spectral theory. To see this, we need a new mathematical tool, which is very important in data analysis and also “data mining”, namely the correlation and the autocorrelation.

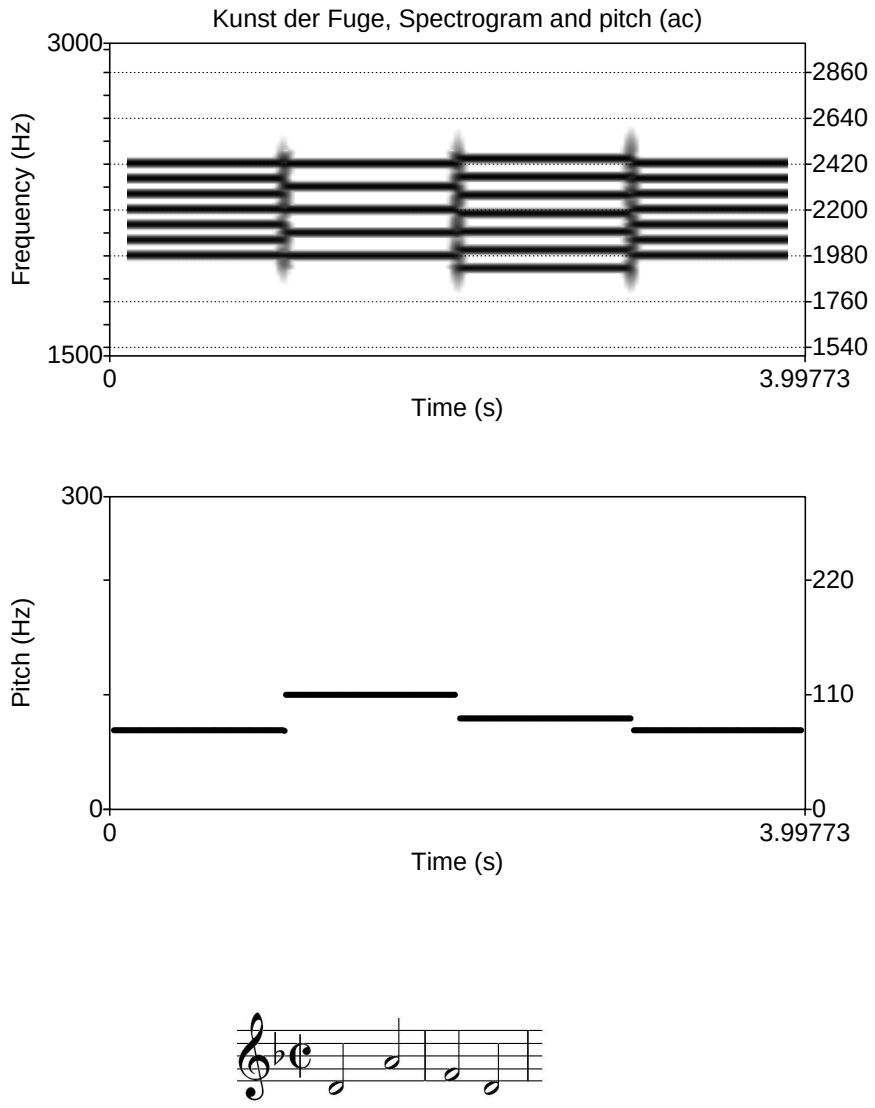


Abbildung 5.16: Spectrogram and pitch (PRAAT ac) the high passed beginning of the theme of the *Kunst der Fuge*.

.

5.1.4 Correlation and autocorrelation

Correlation and deconvolution

The correlation function between to functions $f(t)$ and $g(t)$ is defined as:

$$c(t) \equiv \int_{-\infty}^{\infty} dt'\, f(t')g(t+t') = \int_{-\infty}^{\infty} dt'\, f(t'-t)g(t') \tag{5.10}$$

It is important to note that in the convolution

$$(f * g)(t) = \int_{-\infty}^{\infty} dt'\, f(t')g(t-t') = \int_{-\infty}^{\infty} dt'\, f(t-t')g(t') \tag{5.11}$$

the integration variable t' in the two functions has the opposite sign, whereas in the correlation it has the same sign.

In order to define the correlation as a convolution, on has to define a time-reversed function, $g^T(t) \equiv g(-t)$, then (5.11) can be written as

$$c(t) = (f * g^T)(t) \tag{5.12}$$

The correlation can work true miracles in data analysis, especially if the data are very "noisy". For instance if we want to obtain the influence of the sunspots on the climate, it is best to investigate the correlation function between the sunspot number and the average temperature, the latter being a notoriously noisy quantity. I have mentioned in sect. **??** that by by correlating the firing rate of a nerve at a certain position on the MB with the noise which evokes this movement, one obtains the response of this part of the BM to a momentary excitation. I shall show step by step how this works and illustrate it by an example. This realistic example performed at a computer, that is all functions are represented as discrete sets in a list and the integrals are approximated by finite sums. So for instance the signal $p(t)$ in the interval $[-T, T]$ is represented by the list

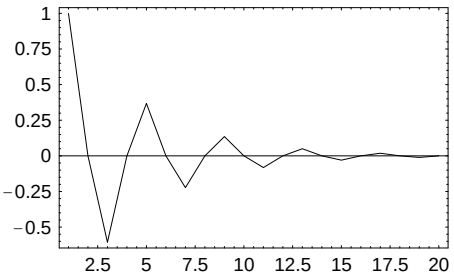
$$\{p_1, p_2, \ldots p_N\}, \quad \text{with } p_k = p(-T + (k-1)/\Delta),\ (N-1)/\Delta = 2T. \tag{5.13}$$

$1/\Delta$ is the sampling rate. The correlation 5.10 of two functions f and g is then expressed as:

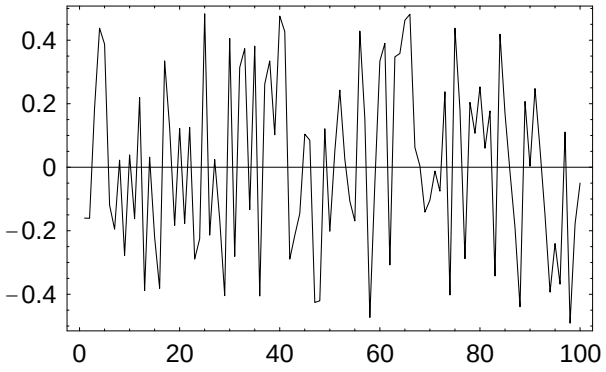
$$c_\ell = \sum f_k\, g_{k+\ell} \tag{5.14}$$

It is a new list $\{c_\ell\}$ with $c_\ell \approx c(\ell/\delta)$. Care must be taken that the element $g_{k+\ell}$ is defined, that is the integration in the correlation extends over an interval which is smaller than $[-T, T]$.

Be $f^c(t)$ the response (firing rate) of an auditory nerve to to a click, it has e.g. in a discretized version the following form



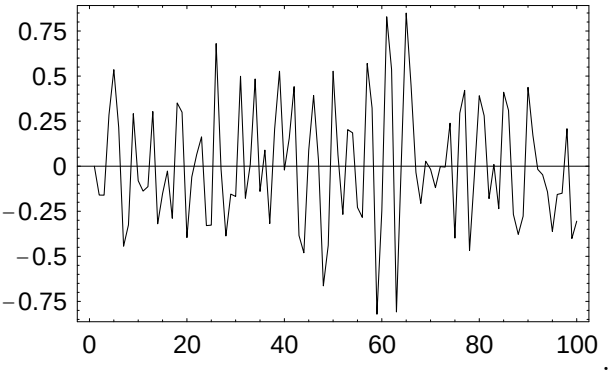
The auditory nerve is stimulated by a noise $p^g(t)$:



The resultin response of the auditory nerve is the result of the **convolution**

$$f^g(t) = \int_{-\infty}^{\infty} dt'\, p^g(t')\, f^c(t - t'). \tag{5.15}$$

It looks as chaotic as the original noise:

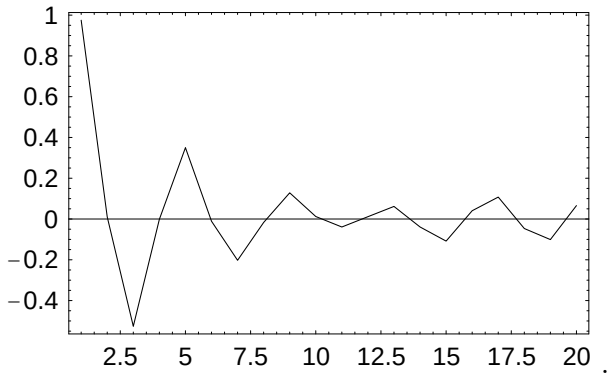


.

But if we now form the **correlation**integral of the response with the evoking noise:

$$f^{cc}(t) = \int_{-\infty}^{\infty} dt''\, p^g(t'')\, f^g(t + t'') \tag{5.16}$$

we obtain back the response to a click : $f^{cc} = f^c$:



In nature, no numerical integral can be performed over an infinite time, but only over a finite interval, therefore the response is still a bit noisy and only approximative, $f^c \approx f^w$. The correlation 5.16 thus leads to th "deconvolution" of the convolution 5.15.

The proof is easy: We insert (5.15) in (5.16) and obtain:

$$f^c(t) = \int_{-\infty}^{\infty} dt''\, p^w(t'') \int_{-\infty}^{\infty} dt'\, p^g(t')\, f^w(t + t'' - t') \tag{5.17}$$

Substituting the integration variable $t'' = \tau + t'$ we obtain

$$\int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} dt'\, p^n(\tau + t')\, p^g(t')\, f^w(t + t'' - t') \tag{5.18}$$

For a noise signal, the integral $\int_{-\infty}^{\infty} dt'\, p^n(\tau + t')\, p^n(t')$ is the δ distribution:

$$\int_{-\infty}^{\infty} dt'\, p^n(\tau + t')\, p^n(t') = \delta(\tau) \tag{5.19}$$

from which follows

$$f^c(t) = \int_{-\infty}^{\infty} d\tau\, \delta(\tau)\, f^w(t + \tau) = f^w(t) \tag{5.20}$$

One can use (5.19) as definition for a noisy tone, but one can make it also easily plausible: We discretize the noise to a list $p_1^g, p_2^g, \ldots p_N^g$ as explained above (5.13. Because of the stochastic nature of the noise, one has for $\ell \neq k$:

$$\sum_{k=1}^{N} p_k \, p_\ell \sim \sqrt{N}, \tag{5.21}$$

whereas one has for $\ell = k$:

$$\sum_{k=1}^{N} p_\ell^2 \sim N. \tag{5.22}$$

Therefore if we form the ratio

$$\lim_{N\to\infty} \frac{1}{N\,\Delta t} \sum_{k=1}^{N} p_k \, p_\ell = \delta_{k\ell} \frac{1}{\Delta t}, \tag{5.23}$$

we obtain a discretized version of the δ distribution, in our example **??** one has: $\Delta t = 2T/\Delta$.

Autocorrelation

If in a convolution 5.11 the two functions are identical, one speaks of an autocorrelation:

$$p_{ac}(t) = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} dt' p(t') p(t + t') \tag{5.24}$$

One can visualize the autocorrelation of a signal $p(t)$ in the following picture: A copy of the signal moves with velocity 1 to the left and the autocorrelation at a given time t is the momentaneous integral over the product of the two curves. One can show that the autocorrelation of a signal has its maximum at $t = 0$ (triangular inequality):

$$p_{ac}(t) \leq p_{ac}(0) \tag{5.25}$$

For a periodic function with period T ome has $p(t) = p(t + T)$ and therefore the $p_{ac}(t + T) = p_{ac}(t)$. This means that the autocorrelation reaches this maximal value again for $t = T$, $2T \ldots$. Therefore the autocorrelation is a very precise way, to determine the period, and the corresponding frequency is given by $\nu = 1/T$. This procedure is independent of phases. We show this in a simple example of the autocorrelation of a sinusoidal tone:

$$p(t) = \sin(\omega t + \delta)$$

.

$$p_{ac}(t) = -\lim_{T\to\infty} \frac{1}{4T}\int_{-T}^{T} dt'(e^{i\omega t'+i\delta} - e^{-i\omega t'-i\delta})(e^{i\omega(t+t)'+i\delta} - e^{-i\omega(t'+t)-i\delta}) \tag{5.26}$$

Among the products in the integrand there are terms oscillating in t', e.g. $e^{i\omega t'+i\delta+i\omega(t+t)'+i\delta}$ and those which do not depend on t', e.g. $e^{i\omega t'+i\delta-i\omega(t'+t)-i\delta}$ The terms oscillating in t' do not increase linearly with T and in the limit $T \to \infty$ only the t'-independent term

$$p_{ac}(t) = \tfrac{1}{2}\cos\omega t \tag{5.27}$$

survives.

We can normalize the autocorrelation by dividing through $\hat{p}_{ac}(t) \equiv p_{ac}(t)/p_{ac}(0)$.

The temporal pitch models are all based on the autocorrelation. Therefore we apply the autocorrelation to the shifted harmonic complex.

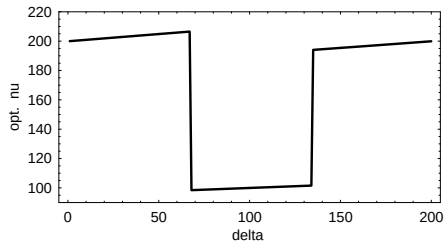
In figure **??** the normalized autocorrelation function of the shifted harmonic complex given by **??** is displayed as a function of the **inverse** correlation time. We see in the left figure, where the shift $\delta = 40$ Hz, besides the absolute (nontrivial) maximum of the autocorrelation function at $\nu = 1/t = 203.9$ Hz a smaller local maximum near 100 Hz. If the shift δ increases this local maximum increases whereas the one near 200 Hz decreases, see the right figure, where $\delta = 80$ Hz.

In fig. 5.17, left, the frequencies yielding the absolute maximum of the correlation function are displayed as a function of δ. Below $\delta \approx 67$ Hz the autocorrelation has a maximum somewhat below 1/200 s, between 67 and 133 Hz near 1/100 s, and above 133 Hz the maximum is slightly abobe 1/200 s. We also show in the same figure, right, the value corresponding to the absolute minimum of the Goldstein pitch extractor, for that case there is no region where the period is predicted to be near 1/100 s.

So both methods yield similar results, except for the region $\delta = \nu_0/2$ where the pitch is not very well defined. It is plausible to assume that for the low partial tones, which lead on the BM to well separated excitations the spectral mechanism is dominant, whereas for high partial tones the temporal extractor is relevant. We shall discuss this again later in connection with the iterated rippled noise (INR).

The Wiener theorem

It is not an accident that in large regions the result of the spectral and the temporal analysis agree. This is due to an intimate connection between the autocorrelation function and the (power) spectrum, expressed by the theorem of Wiener:



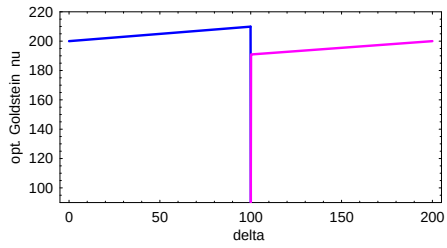


Abbildung 5.17: Left: The autocorrelation function of the shifted harmonic complex for the shift $\delta = 40$ Hz. Right: The optimal frequency, as obtained by the autocorrelation (black) and from the spectral Goldstein pitch extractor (blue and magenta), for the shifted harmonic complex, as a function of the shift δ

The Fourier transform of the autocorrelation function of a signal is the power spectrum of that signal.

Proof: One can show this directly by performing the Fourier analysis of the ac function of the real signals $p(t)$, and making use of the miracle formula (2.17). But more elegant is the following proof. It has been pointed out, (**??**), that the autocorrelation can be expressed by the convolution of $p(t)$ and $p^T(t) = p(-t)$:

$$p_{ac} = \int_{-\infty}^{\infty} dt'\, p(t')\, p(t+t') = (p * p^T)(t) \tag{5.28}$$

We make use of the result that the Fourier transform of a convolution is the product of the Fourier transforms.

$$\tilde{p}_{ac}(\omega) = \tilde{p}(\omega)\, \tilde{p}^T(\omega) \tag{5.29}$$

$$\begin{aligned} \tilde{p}^T(\omega) &= \int_{-\infty}^{\infty} dt e^{i\omega t} p(-t) \\ &= \int_{-\infty}^{\infty} dt e^{-i\omega t} p(t) = (\tilde{p}(\omega)^*. \end{aligned} \tag{5.30}$$

qed.

Iterated rippled noise (INR)

An important example where the autocorrelation reveals the periodicity hidden in a noisy signal is the iterated rippled noise. It is defined in the following way: Be $p(t)$ the SP of a noise, the

INR with ℓ iterations and the gain factor g is defined as

$$p^{H\,\ell}(t) = \sum_{k=0}^{\ell} g^k\, p(t + k\,\Delta) \tag{5.31}$$

The autocorrelation is given by

$$\begin{aligned} p_{ac}^{H\,\ell}(t) &= \int_{-\infty}^{\infty} dt' \sum_{k=0}^{\ell} g^k\, p(t' + k\,\Delta) \sum_{k'=0}^{\ell} g^{k'}\, p(t + t' + k'\,\Delta) \\ &= \sum_{k,k'=0}^{\ell} g^{k+k'} \delta(t + (k' - k)\Delta) \end{aligned} \tag{5.32}$$

There is a finite set of correlation times $T_{k',k} = (k'-k)\Delta$, that is times, where the autocorrelation function has a maximum:

$$p_{ac}^{H\,\ell}(t) = (\ell + 1)\delta(t) + g\,\ell\delta(t - \delta) + g^2\,(\ell - 2)\delta(t - 2\Delta) + \ldots \tag{5.33}$$

Allthough the SP of INR looks as noisy as simple noise, see fig. 5.18, the ear is evidently well aware of the periodicity – or of the spectral structure. The normalized autocorrelation function for a simulated noise generated by random numbers is displayed in figure 5.19. One hears indeed for the rippled noise in the noise a tone with the frequency $1/\Delta$. `praats-bsp/huygens`.

But the same conclusion which can be drawn from the ac function can also be drawn from the spectra. In fig. **??** the spectra and the autocorrelation function for INR with 1 and 5 iterations and a delay time of 1/440 s and positive gain, $g = 1$, is shown. The autocorrelation function has been determined here by a (backwards) Fourier transformation of the power spectrum using Praat. Since this is based on a finite integration and discretization the resulting qutocorrelation function is a bit "noisy". For 5 iterations the autocorrelation function has a higher maximum at $t = 1/440$ s than the INR with 1 iteration, but also the spectral peaks at $n \cdot 440$ Hz are sharper, both could explain the greater salience of the pitch in that case.

Interesting is the case for negative gain, $g = -1$ in 5.33. Here we hear a less distinct pitch in the noise. The spectral representation, **??**, upper ropw, shows now peaks at $(2n - 1) \cdot 220$ Hz, $n = 1, 2, \ldots$. For 5 iterations the spectral peaks are distinctly sharper again.

A complex tone with partial tones at these frequencies has a distinct pitch of the a(220). The autocorrelation function shows for negative gain and 1 iteration no nontrivial maximum at all. For 5 iterations it shows a clear maximum at 1/220 s, in accordance with spectral theory. Since the INR with negative gain is definitely different from the continuous noise, the INR with $g = -1$ seems to be only in agreement with the spectral theory. However the adherer of the

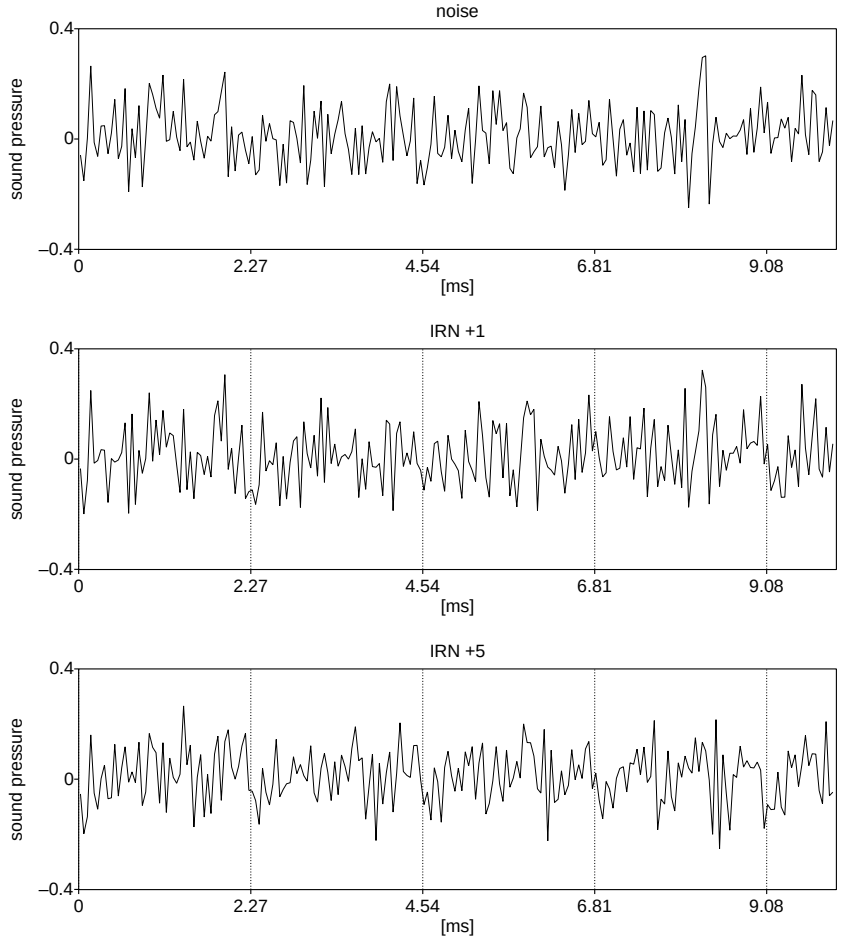


Abbildung 5.18: AP of noise and INR with 1 and 5 iterations and gain factor $g = 1$.

.

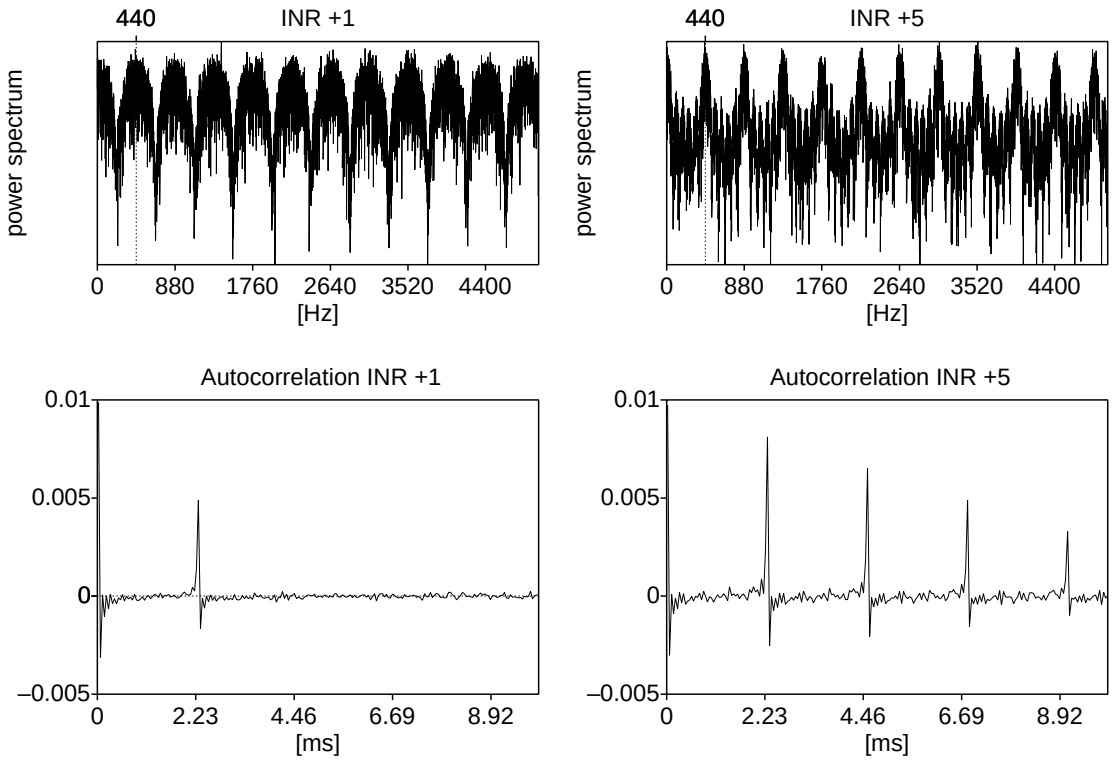


Abbildung 5.19: Spectra and autocorrelation functions of huygens noise with $\ell = 1$ and with $\ell = 5$ repetitions

.

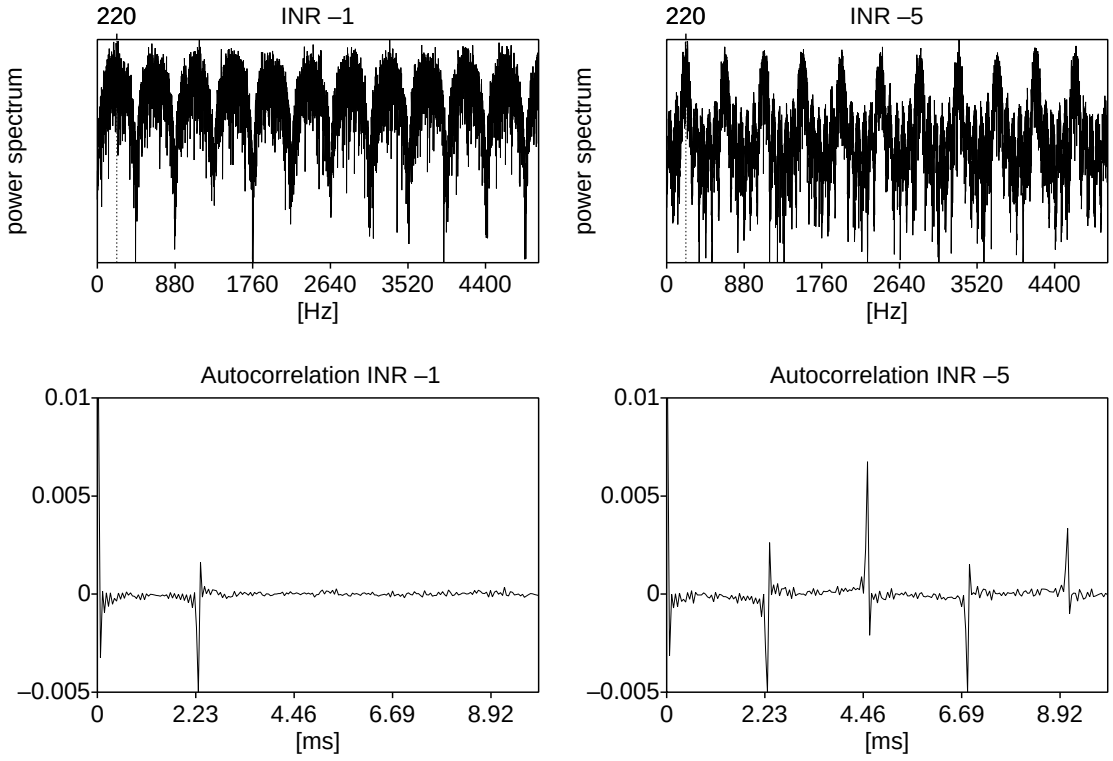


Abbildung 5.20: Spectra and autocorrelation function of huygens noise with $g = -1$ and with $\ell = 1$ and 5 repetitions

.

temporal theory argue, that the autocorrelation is performed only for a certain band in the spectrum, corresponding to the region of the BM which is simultaneously excited, this band has the width of a few critical band widths. In figure **??** the spectrum and the autocorrelation function for anti-Huygens noise, band pass filtered between 2200 and 4400 Hz is displayed. Now the ac function exhibits two nearly equally high local maxima near $t = 1/440Hz$.

Kapitel 6

Nichtinvasive Methoden in der Neurophysiologie

6.1 Kernspintomographie (NMRI und FNMRI)

Die hauptsächlichen Bestandteile der Materie, Elektronen, Protonen und Neutronen haben einen Spin und mit ihm ein magnetisches Moment. Die Werte sind:

$$\begin{aligned} \text{electron} \quad & \mu_B = e\hbar/(2m_e) = 5.8 \times 10^{-11} MeVT^{-1} \\ \text{proton} \quad & \mu_p = 1.4e\hbar/(2m_p) = 1.6 \times 10^{-14} MeVT^{-1} \end{aligned}$$

(T steht für Tesla)

Ein Spin kann sich in einem Magnetfeld $\vec{B}$ entweder parallel (Energie $W = -\mu B$) oder antiprallel ($W = +\mu B$) ausrichten. Im thermischen Gleichgewicht ist die Zahl der parallel ausge-

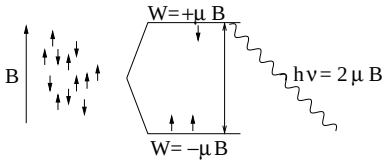


Abbildung 6.1: Spinverteilung und Energieniveaux im Magnetfeld

richteten (N_-) zur Zahl der antiparallel ausgerichteten (N_+)gegeben durch

$$\frac{N_+}{N_-} = \frac{e^{-W_+/(kT)}}{e^{-W_/(kT)}} = e^{-2\mu B/kT} \tag{6.1}$$

Bei Zimmertemperatur ist $kT \approx 1/40eV$, d.h. im Magnetfeld der Erde (1 gauss = 1/10000 tesla) ist das Verhälltnis für Protonen $e^{-6.410^{-11}} \approx 1 - 6.410^{-11}$, In einem Magnetfeld von 1 Tesla immerhin $e^{-6.410^{-7}}$

Durch ein eingestrahltes HF Signal der Frequenz $\hbar\nu = 2\mu B$ können die Spins umgeklappt werden und aus dem thermischen Gleichgewicht gebracht werden. Dabei senden sie wieder HF Signale aus, die nachgewiesem werden, und aus denen das Bild konstruiert wird.

Da die Protonen nicht isoliert, sondern in einer chemischen Umgebung sind, ist das Magnetfeld an ihrer Stelle nicht nur durch das äussere B-Feld sondern auch durch die Elektronenkonfiguration der Umgebung bestimmt. Dies macht die Frequenz des ausgesandten Signals empfindlich für die chemische Bindung. So können z.B. die Protonen im Wasser von denen in einer anderern Umgebung unterschieden werden. Eine einfachste NMR Analyse misst also den Wassergehalt und kann damit z.B. Liquor (prakt. reines Wasser) von Neuronen (reich an Wasser) und Myelin (arm an Wasser) durch die Intensität des für H_2O typischen Signals unterscheiden.

Die Umgebung hat auch noch einen grossen Einfluss auf die Relaxaktion, d.h. die Zeitkonstante in der die angeregten Spins ein Signal aussenden (umklappen) oder auch ausser Phase geraten, die entsprechenden Zeitkonstanten werden T_1, T_2, T_2^* genannt. Durch selektive Auswahl der einzelnen Relaxaktionszeiten lässt sich also auch die chemische Umgebung feststellen.

Dies ist besonders wichtig für die Funktionelle Kernmoment Bildgebung (FNMRI). Das Hämoglobin im Blut hat sehr verschiedene magnetische eigenschaften, je nachdem ob es oxidiert ist (HbO_2) oder reduziert ($HB,\ Hbr$).Im HBr treten im Gegensatz zum HBO_2 ungepaarte Elektronen auf, deshalb hat Hbr ein starkes Magnetisches Moment und beeinflusst das lokale Magnetfeld stark, HbO_2 hat dagegen nur gepaarte Elektronen, es beinflusst das lokale Magnetfeld also nur schwach. Bei verstärkter Durchblutung einer Gebietes nimmt das Hbr im venösen Blut ab, da dann ein Überangebot an Sauerstoff vorliegt. Damit nimmt bei erhöhter Durchblutung die Unregelmässigkeit im lokalen Magnetfeld ab und dies beeinflusst die relaxationszeit.

Bei der FNMRI konzentiert man sich auf die HF Signale, die auf die Änderungen im lokalen Magnetfeld empfindlich reagieren ($T2^*$). Misst man diese HF Signale mit Reiz und zieht davon die Signale ohne Reiz ab so spiegelt die Differenz der Signale die Differenz zwischen Hbr und HbO_2 wieder, also letzlich die Durchblutung der untersuchten Region (BOLD-Effekt). Damit zeigt die Signaldifferenz gerade die Gebiete an die bei durch Reiz besonders aktiviert werden.

Leider ist offenbar bei den $T2^*$-Signalen die lokale Genauigkeit besonders schlecht. Natürlich

ist auch die zeitliche Auflösung im Sekundenbereich nicht im entferntesten mit der des MEG vergleichbar. Ein grosser Vorteil ist allerdings, dass man nicht den Beschränkungen der inversen Methode unterworfen ist, da die Signale wirklich direkt aus dem Innern des Hirns kommen.

6.2 Positronenemissionstomographie (PET)

Bei dieser Methode wird ein β^+-Strahler ins Blut injiziert. Die beim Zerfall auftretenden Positronen vernichten sich mit Elektronen, die beiden bei der Annihilation auftretenden γ-Quanten werden nachgewiesen. In besonders stark durchblutete Gebiete wird auch besonders viel β^+-Strahler transportiert und deswegen ist dort die Vernichtungsstrahlung besonders intensiv. Auf diese Weise lässt sich feststellen, welche Gebiete des Hirns bei gewissen Tätigkeiten oder reizen besonders stark durchblutet sind, die PET ist also ähnlich der FMRI, vielleicht etwas stabiler, aber dafür doch etwas invasiver (Injektion eines radioaktiven Präparates)

6.3 Elektro- und Magneto-Enkephalographie

6.3.1 Some electrophysiology for MEG

[?]

The gray matter in the cortex consists mainly of neurons, there are about 10^{10} neurons in the cortex, connected among each other by about 10^{14} synapses. In figure 4.3 we show a very schematic picture of a neuron. The dendrites receive signals from other neurons through synapses. An active synapsis gives rise to a Postsynaptic Potential (PSP) which in turn leads to a current dipole of ≈ 20 fAm.

A typical current dipole moment measured in MEG is several tens of a nAm, that is there must be 10^6 active synapses for a signal. The resulting measured magnetic fields are around 100 fTesla, that is around 10^{-8} of the magnetic field of the earth.

6.3.2 SQUID magnetometers

The measurement of such weak fields became only possible through the development of very sensitive magnetometers, the SQUIDS. These magnetometers consist of closed supraconductors interrupted by thin normal conducting layers (Josephson junctions) , see figure 6.2. In the

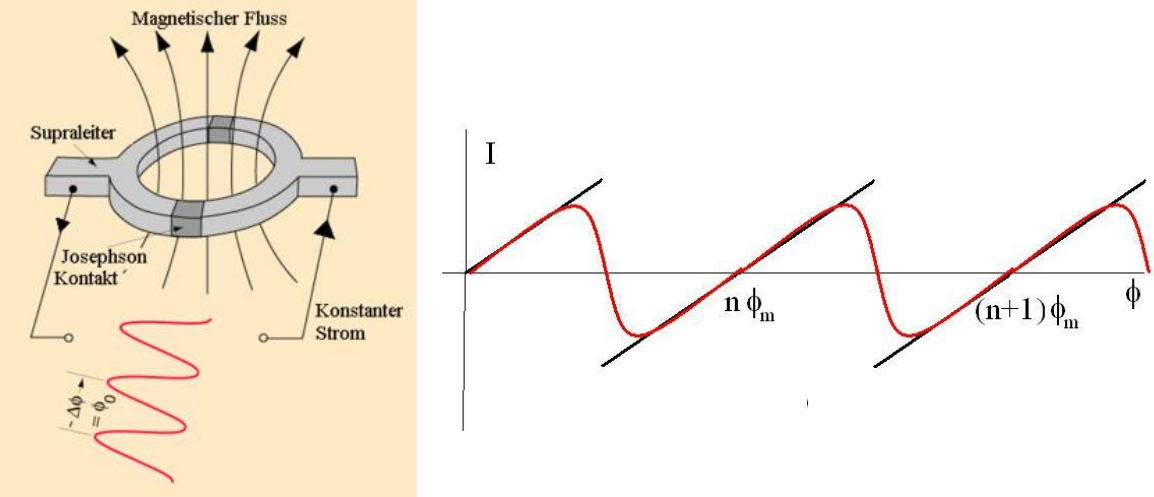


Abbildung 6.2: Construction principle of a SQUID
and the compensating currents, black theoretical, without Josephson junction, red with Josepson junction.

supraconducter electrons form Cooper pairs which condense and are the carriers of the current in the supraconductor. According to QM the magnetic flux, $\Phi = \int_F \vec{B}.d\vec{f}$, through the squid is quantized in flux quanta:

$$\Phi_m = \frac{\hbar c 2\pi}{2e} \approx 4 \cdot 10^{-7} \text{ gaus cm}^2 = 4 \cdot 10^{-15} \text{ Wb} = 0.0001 \; fT\, cm^2 \tag{6.2}$$

(1 WB (Weber) is 1 T m^2)

If the magnetic flux is larger than an integer multiple of Φ_m a compensating current must flow in the squid which reduces or increases the flux to an integer multiple. This current leads to a potential difference across the Josephson junctions and can be measured. The Squid measures therefore not the absolute value of the magnetic field, but only the difference to the next multiple of the flux quant, it is therefore sensitive to changes in the magnetic field to the fraction of a flux quant and thus lies indeed in the femto-Tesla region, as can be seen from 6.2.

6.3.3 Electrodynamics for MEG

Construction of magnetic fields and potentials from given primary currents

In EEG one studies the potentials induced on the surface of the head induced by currents flowing in cells of the brain (mostly synaptic currents). In MEG on studies the magnetic fields outside (but near) the head induced by those currents. The basis are off course the Maxwell equations. The frequencies involved are low enough to justify the quasi static approximation, that is we start with the equations:

$$\vec{\partial}\vec{D} = 4\pi\rho \tag{6.3}$$
$$\vec{\partial}\vec{B} = 0 \tag{6.4}$$
$$\vec{\partial} \times \vec{E} = 0 \tag{6.5}$$
$$\vec{\partial} \times \vec{H} = \vec{j} \tag{6.6}$$
$$\vec{j} = \sigma \vec{E} \tag{6.7}$$

In the vacuum and the head we can put $\vec{B} = \vec{H}$.

Because of 6.5 we can express $\vec{E}$ by the quasi static potential ϕ

$$\vec{E}(\vec{x}\,) = -\vec{\partial}\phi(\vec{x}\,) \tag{6.8}$$

We are interested in this potential ϕ on the surface of the head and on the magnetic field $\vec{B}$ outside the head. For further reference we denote by
• G the interior of the head and by
• ∂G the surface of the head.

In one cell or "a few" neighboring ones (mostly synaptic cells) flows a well localized current with current density $\vec{j}_p$, (black box in figure 6.3). It is called the primary current. Since the brain is conducting, the potential difference created by the primary current leads to a current distributed over the whole brain, the so called volume current with current density $\vec{j}_v$ (dashed lines in the same figure).

The primary current $\vec{j}_p$ together with the volume current create a magnetic field $\vec{B}$. It can be calculated with the help of Maxwell's equations. 6.4 allows to express $\vec{B}$ as the rotation of a vector potential

$$\vec{B} = \vec{\partial} \times \vec{A} \tag{6.9}$$

which can be chosen to fulfill the Coulomb gauge

$$\vec{\partial}\vec{A} = 0 \tag{6.10}$$

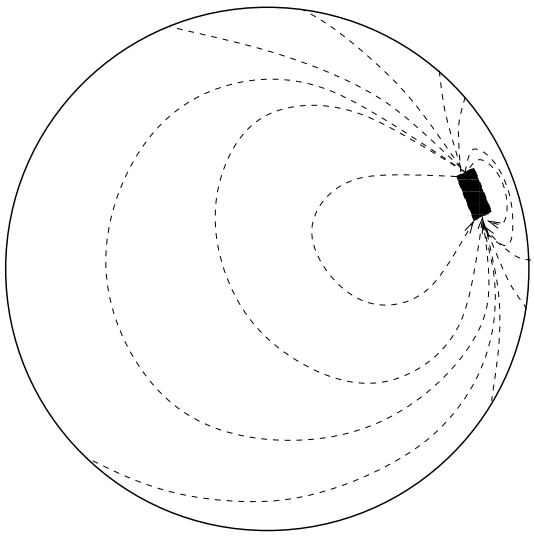


Abbildung 6.3: Current distribution in the brain initiated by a primary current in the black box; the dashed lines are the stream lines of the volume current.

6.6 becomes then

$$\vec{\partial}^2 \vec{A} = -\vec{j} \tag{6.11}$$

With a solution vanishing at $\vec{|}x\,| \to \infty$ 6.11 leads to the standard solution::

$$\vec{A}(\vec{x}\,) = \frac{1}{4\pi} \int_G d^3x' \frac{\vec{j}(\vec{x}\,')}{|\vec{x} - \vec{x}\,'|} \tag{6.12}$$

If the current distribution vanishes sufficiently fast the magnetic field is obtained from 6.6 as:

$$\vec{B}(\vec{x}\,) = \frac{1}{4\pi} \big[\vec{\partial} \times \int_G d^3x' \frac{\vec{j}(\vec{x}\,')}{|\vec{x} - \vec{x}\,'|} = \frac{1}{4\pi} \int_G d^3x' \frac{[\vec{j}(\vec{x}\,') \times (\vec{x} - \vec{x}\,')]}{|\vec{x} - \vec{x}\,'|^3} \tag{6.13}$$

Putting the rotation in $\vec{B} = \vec{\partial} \times \vec{A}$ inside the integral, switching from $\vec{\partial}$ to $\vec{\partial}'$ and performing a partial integration one arrives at the following form:

$$\vec{B}(\vec{x}) = \frac{1}{4\pi} \int d^3x' \frac{1}{|\vec{x} - \vec{x}\,'|} \vec{\partial}' \times \vec{j}(\vec{x}\,') \tag{6.14}$$

Because of 6.7 and 6.8 the total current is

$$\vec{j}(\vec{x}) = \vec{j}_p(\vec{x}) + \vec{j}_v(\vec{x}) = \vec{j}_p(\vec{x}) - \sigma(\vec{x}) \vec{\partial} \phi(\vec{x}). \tag{6.15}$$

Simple manipulations (most easily performed using the ϵ-tensor) lead to

$$\vec{\partial} \times \vec{j} = \vec{\partial} \times (\vec{j}_p + \phi \vec{\partial} \sigma) \tag{6.16}$$

that is besides the primary current the volume current contributes to the magnetic field as $\vec{j}_{eff} = \phi(\vec{x}\,')\vec{\partial}\,'\sigma(\vec{x}\,')$. Inserting this into 6.14 and partial integration yields finally to the result:

$$\vec{B}(\vec{x}) = \frac{1}{4\pi} \int d^3x' \frac{\left(\vec{j}_p(\vec{x}\,') + \phi(\vec{x}\,')\vec{\partial}\,'\sigma(\vec{x}\,')\right) \times \left(\vec{x} - \vec{x}\,'\right)}{|\vec{x} - \vec{x}\,'|^3} \tag{6.17}$$

If it where not for the space dependence of the conductivity σ the contribution of the volume current would vanish. Even if we assume a homogeneous conductivity inside the head, there is an x dependence, since outside the had the conductivity is zero.

The potential ϕ is related to the primary current $\vec{j}_p$ through 6.15 and $\vec{\partial}.\vec{j} = 0$, a consequence of 6.6. it yields

$$\vec{\partial}.(\sigma \vec{\partial}\,\phi) = (\vec{\partial}\sigma).(\vec{\partial}\phi) + si\vec{\partial}^2\phi = \vec{\partial}.\vec{j}_p \tag{6.18}$$

Outside the head the conductivity is zero. Because of the continuity of the current components the current normal to ∂G inside G is zero and therfore

$$\hat{n}_{\partial G}.\vec{\partial}\phi = 0 \tag{6.19}$$

on ∂G. where $\hat{n}_{\partial G}$ is the normal on ∂G, the surface of the head.

Thus we can calculate the potential from 6.18 and the boundary condition 6.19. From the potential and with 6.3.3 we can calculate the magnetic field.

If the volume element, where $\vec{j}_p \neq 0$ is small we can replace it by

$$\vec{j}_p(\vec{x}) \approx I\ \vec{l}\ \delta(\vec{x} - \vec{x}_Q) = \vec{Q}\ \delta(\vec{x} - \vec{x}_Q). \tag{6.20}$$

Here $\vec{x}_Q$ is the position where the current density is different from zero, I is the current strength and $\vec{l}$ is the direction vector of the current density. $\vec{Q}$ is called the current dipole,

$$\vec{Q} = \int d^r x'\ \vec{j}_p(\vec{x}\,') \tag{6.21}$$

The contribution of a primary current described by the current dipole $\vec{Q}$ at position $\vec{x}_Q$ is thus ,see :

$$\vec{B}(\vec{x}) = \frac{1}{4\pi} \frac{\vec{Q} \times (\vec{x} - \vec{x}_Q)}{|\vec{x} - \vec{x}_Q|^3} \tag{6.22}$$

The results obtained above can be summarized as follows:

The potential and the Magnetic field at any position outside and on the surface of the head can be calculated from $\vec{j}_p$ and depend linearly on it:

$$\phi(\vec{x}) = \int d^3x' \; \vec{\mathcal{L}}(\vec{x},\vec{x}\,').\vec{j}_p(\vec{x}\,') \tag{6.23}$$

$$\vec{B}_\alpha(\vec{x}) = \int d^3x' \; \vec{\mathcal{L}}^m_\alpha(\vec{x},\vec{x}\,').\vec{j}_p(\vec{x}\,') \tag{6.24}$$

where $\vec{\mathcal{L}}$ depends only on the geometry and conductivity of the head.

For a current dipole we obtain from 6.24 and 6.20:

$$\vec{B}_\alpha(\vec{x}) = \vec{\mathcal{L}}^m_\alpha(\vec{x},\vec{x}_Q).\vec{Q} \tag{6.25}$$

We have the important theorem:

Theorem: The magnetic field outside the head is uniquely determined by the component $B_n(\vec{x}) = \hat{n}_{\partial G}.\vec{B}(\vec{x})$ with $\vec{x} \in \partial G$, that is the normal component of $\vec{B}$ on the surface of the head.

Proof: Be G' the complement of G, that is the space outside the head. In G' we have $\vec{j} = 0$ and hence $\vec{\partial} \times \vec{B} = 0$. Therefore we can in G' express $\vec{B}$ through a scalar potential ϕ_M

$$\vec{B} = \partial\phi_M \tag{6.26}$$

ϕ_M satisfies in G', because of 6.4

$$\vec{\partial}^2\phi_M = 0 \tag{6.27}$$

The solution is determined through the Neumann boundary condition on the surface of G', that is the surface of the head and infinity. At infinity everything is assumed to be zero (shielding !) and therefore the normal derivative on the surface of the head determines ϕ_m and thus $\vec{B}$ in G'.

The solution is given by

$$\phi_m(\vec{x}\,) = -\frac{1}{4\pi}\int_{\partial G} d^2a' \; G_N(\vec{x},\vec{x}\,').\hat{n}_{\partial G}.\vec{\partial}\phi_m(\vec{x}\,') \tag{6.28}$$

G_N is the Greens function satisfying the Neumann condition

$$\hat{n}_{\partial G}.\vec{\partial}\,' G_N(\vec{x},\vec{x}\,') = 0 \quad \vec{x}\,' \in \partial G \tag{6.29}$$

(see e.g. Jackson, 1.10).

The normal derivative $-\hat{n}_{\partial G}.\vec{\partial}\phi_m(\vec{x}\,')$ is just B_n, the component of $\vec{B}$ normal to the surface of the head.

Spherically symmetric case

Theorem If the head is a spherically symmetric system, that is $\sigma(\vec{x}\,) = \sigma(|\vec{x}\,|)$ and G is a sphere, the contribution of the volume current to the normal component of $\vec{B}$ on the surface of the head is zero.

Proof For a spherically symmetric system we have $\sigma = \sigma(|\vec{x}\,|)$ and $\hat{n}_{\partial G} = \vec{\partial}|\vec{x}\,|\vec{x}/|\vec{x}\,| \equiv \hat{x}$.

We have $\partial\sigma(|\vec{x}\,|) = \sigma'(|\vec{x}\,|)\hat{x}$, σ' is the derivative of $\sigma|\vec{x}\,|)$ with respect to $|\vec{x}\,|$.

We therfore deduce from 6.3.3 the contribution of the volume current:

$$\vec{B}^{\,V}(\vec{x}) = \frac{1}{4\pi}\int d^3x'\phi(|\vec{x}\,|)\sigma(|\vec{x}\,|)(\vec{x}' \times \vec{x})\frac{1}{|\vec{x} - \vec{x}\,'|^3} \tag{6.30}$$

which is evidently orthogonal to $\vec{x}$ and thus to $\hat{x}$. Thus $B_n^V = \hat{x}.\vec{B}^{\,V} = 0$ and therefore only the primary current contributes to the normal component.

For a current dipole we have:

$$B_r(\vec{x}) \equiv \vec{B}(\vec{x}).\hat{x} = -\frac{1}{4\pi}\frac{\vec{Q}.[\vec{x} \times \vec{x}_Q]}{|\vec{x} - \vec{x}_Q|^3} \tag{6.31}$$

This is summarized in Figure 6.4

According to the theorem derived above the radial component of the magnetic field is thus exclusively determined by the primary current and no detailed knowledge of $\sigma|\vec{x}\,|)$ is necessary. This is a big advantage of the MEG. Even if the spherically symmetric approximation might not be very trustworthy, it is quite good if the source is near the surface of the head (as the auditory cortex) and if the local curvature is taken into account, see figure **??**

This is pleasant, but it has also has an important (and unpleasant) consequence:

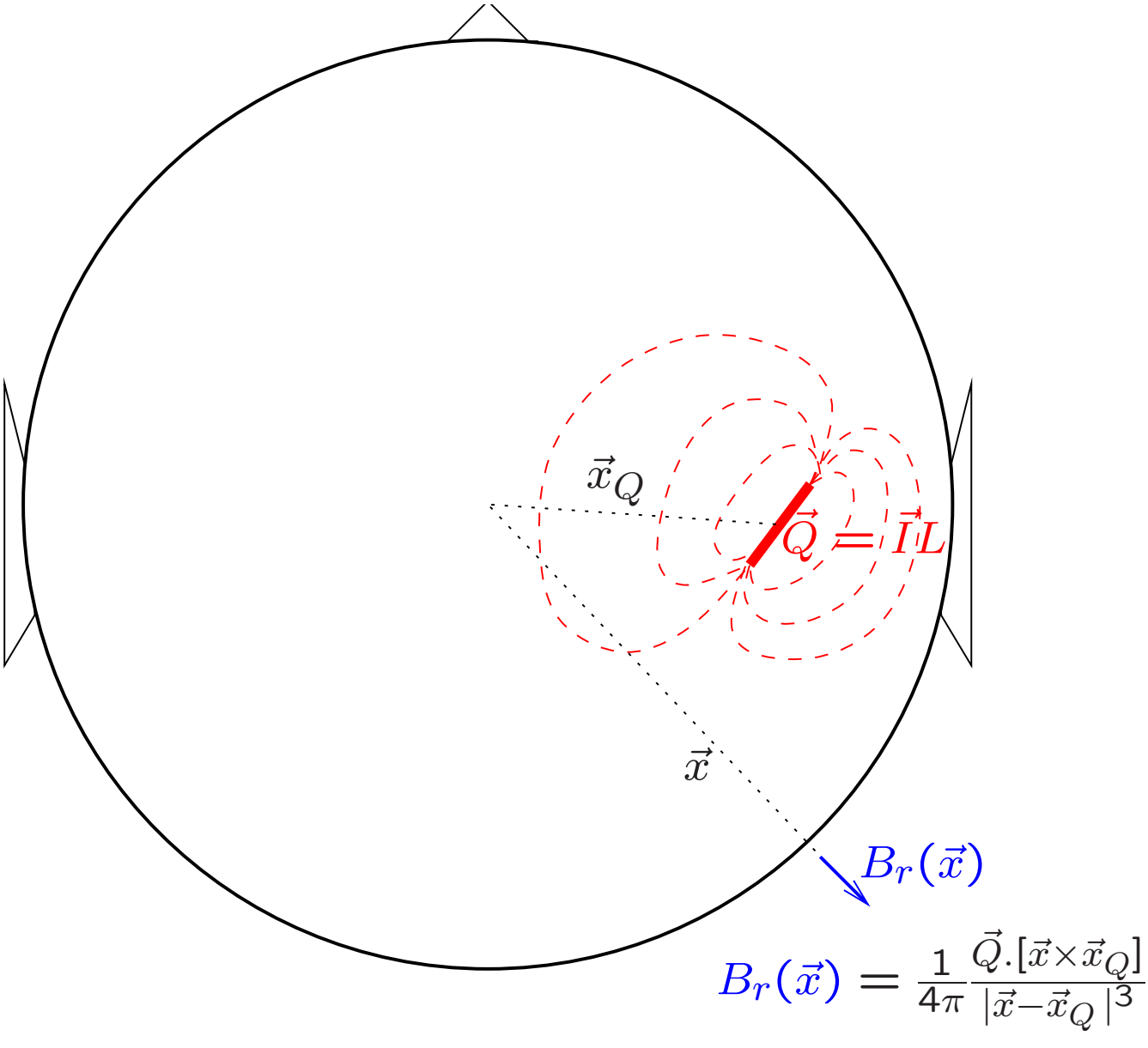
If a current dipole is in radial direction, that is if $\vec{Q} = |\vec{Q}|\hat{x}_Q$ then $\vec{B}(\vec{x}) = 0$ outside the head.

The proof is simple: One sees easily from 6.31 that $B_r = 0$ on the surface of the head in that case and therefore the Neumann boundary volume problem has only the trivial solution $\vec{B} = 0$.

Not that the primary current leads to tangential magnetic field components but that these are compensated by the volume current.

6.3.4 The inverse Problem

The problem to obtain potentials or fields from given sources is straightforward. Mathematically and physiologically much more interesting is the so called inverse problem, namely to obtain



Q für Neuron: ≈ 20 fAm
Q für Signal ≈ 20 nAm $\rightarrow$ 100 fTesla

Aus $B_r(\vec{x})$ kann $\vec{B}(\vec{x})$ konstruiert werden (Green).

Abbildung 6.4: Dipole solution in the spherical head

information on the primary from measured potentials or fields on the surface of the head. It has been shown already in 1853 by Helmholtz that this problem has no unique solution.

That this is the case follows from the fact that there are non vanishing primary currents which leave no signal on the surface of the head. We have seen already in the previous section that in the spherically symmetric case a radial current dipole $\vec{Q} = |\vec{Q}\,|\hat{x}$ cannot create a magnetic field outside the head. There are analog cases for the potential and even those which leave neither an electric nor a magnetic signal. Therefore the solutions found are never unique. Fortunately in the auditory cortex must primary currents seem to be parallel to the surface of the head and thus accessible to MEG measurements.

6.3.5 Dipole models

One makes the model that the (measurable components of) the magnetic field are initiated by one or several current dipoles and the external field is therfore given by 6.25 or by a superposition of these terms in addition to the corresponding volume currents which are also determined by strength an position of the current dipoles. One tries to choose these parameters in such a way as to minimize the discrepancy between the observed and the calculated field. Off course given the background noise and the experimental errors, only a limited amount of dipoles makes sense, since mor dipoles lead to more free parameters and therefore always lead to a solution with an excellent fit. This solution might however be very far from the true solution. This is known to everybody who has made multipara meter fits with noisy signals and a somewhat shaky theoretical background.

6.3.6 Minimal norm solution

Given Head model: Form and conductibility $\sigma(x)$.

Wanted: Postsynaptic current distribution $\vec{j}_P(x)$.

ED:
$\vec{B}(x) = \int dx' \mathcal{L}_m(x, x') j_P(x')$;
$\Phi(x) = \int dx' \mathcal{L}_e(x, x') j_P(x')$;

Be V_i, $i = 1, \ldots, N$ the signals registered at position x_i, i.e. $V_i = \Phi(x_i)$ or say $B_n(x_i)$. Then

that is

$$V_i = \int dx' \mathcal{L}(x_i, x') j_P(x')$$

where $\mathcal{L}(x_i, x')$ depends on the head model, for EEG crucially on $\sigma(x)$

We want now to reconstruct from the measured signals V_i the distribution of the primary currents j_P. There are two Problems:
1) Fundamental, inverse Problem: There are primary currents,
which give no signals, that is:

$$0 = \int dx' \mathcal{L}(x_i, x') j_P^0(x')$$

2) Technical: Only finite Number of signals V_i.

For the inverse problem, there is no cure, only remedy: Ignore those currents!
Set: $j_P = \hat{j}_P + j_P^0$ and call $\hat{j}_P$ your estimate for j_P.

Then we make for $\hat{j}_P$ the ansatz:

$$\hat{j}(x) = \sum_{k=1}^{N} w_k \, \mathcal{L}(x_k, x) \tag{6.32}$$

This is the best we can do, since solutions, which cannot be expressed in that way, e.g. j_P^0, cannot be determined anyhow.

We insert the ansatz and obtain now the system of linear equations:

$V_i = \int dx' \mathcal{L}(x_i, x') \, \hat{j}(x') = \int dx' \mathcal{L}(x_i, x') \, \sum_{k=1}^{N} w_k \, \mathcal{L}(x_k, x')$

$V_i = w_k K_{ik}, \;\; K_{i,k} = \int dx' \mathcal{L}(x_i, x') \, \mathcal{L}(x_k, x')$

Theoretically clear: Linear equations, easy to solve.

$w_k = \sum K_{ki}^{-1} V_i$

If we have the w_k we can obtain $\hat{j}_P$ from **??**, since $\mathcal{L}(x_k, x)$ are known functions.

But there is a serious problem: Small errors in V_i or $\mathcal{L}(x_i, x')$ my lead to huge errors in w_k. Errors in V_i are measuring errors, errors in $\mathcal{L}(x_i, x')$ are model errors, for instance entering through the simplification of the asumptions for $\sigma(x)$.

The following is not a realistic example, but rather a carricature, whichaccentuates essential problems of the method.

Be $x_1 \approx x_2$, then $\mathcal{L}(x_1, x') \approx \mathcal{L}(x_2, x')$
then $K_{11} \approx K_{12} = K_{21} \approx K_{22}$

$$3 \;\; = \;\; 1\, w_1 + 0.998\, w_2$$

$$3.1 \quad = \quad 0.998\, w_1 + 1\, w_2$$

Certain: $w_1 + w_2 \approx 3$,

But any solution

$w_1 = 1.5 - x,\ w_2 = 1.5 + x$

is approximately equally good.

$K.\{w1, w2\} = \{2.9976, 2.9964\}$ solution: $\{w1, w2\} = \{1.8, 1.2\}$

But:

$K.\{w1, w2\} = \{3.0, 3.0\}$ solution: $\{w1, w2\} = \{1.5015, 1.5015\}$

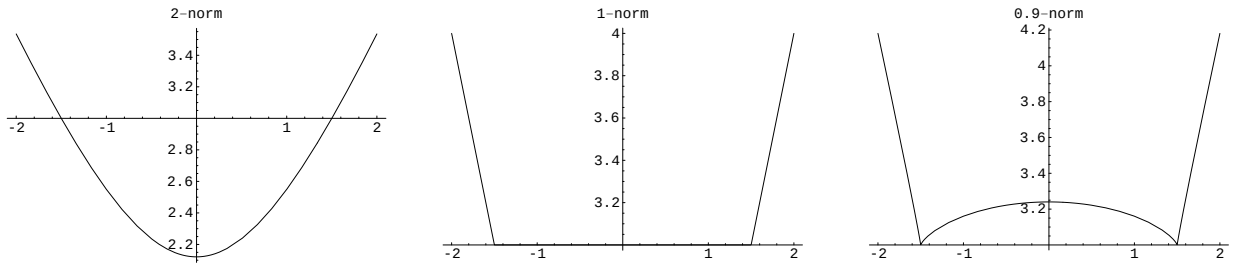
$K.\{w1, w2\} = \{2.99, 3.01\}$ solution: $\{w1, w2\} = \{-3.5, 6.5\}$

Create a "unique solution":

p-Norm of vector $\vec{w} = (w_1, w_2)$: $|\vec{w}\,|_p \equiv \left(|w_1|^p + |w_2|^p\right)^{1/p}$

Minimal Norm estimate: Choose x such, that $|\vec{w}\,|_p$ minimal.

Practical: set $\{w1, w2\} = \{1.5 + x, 1.5 - x\}$ and find minimum of p-norm.



As can be seen from fig. **??**, the results differ considerably:

2-Norm (the usual) $\vec{w} \approx (1.5, 1.5)$

1-Norm (also popular) $\vec{w} \approx (3, 0) \ldots (0, 3)$

0.9-Norm (creativ) $\vec{w} \approx (3, 0)$ or $(0, 3)$