4. Basics of binaural Signal Processing

This chapter shall specify the environment for a model of binaural signal processing. It contains, for instance, the description of ear signals, formula symbols and the used nomenclature.

Based on these premises binaural models shall be described, and conditions shall be derived, to evaluate the directions and amplitudes of multiple sound sources. By means of a cross correlation function based model in the frequency range possibilities for the construction of a Cocktail-Party-Processor are presented.

In the following transfer functions and ear signals within critical bands are treated. In this context mainly "head oriented", "binaural" transfer functions are described, whereby natural heads or dummy heads can rather be understood as one (relatively complicated) example of it. Other recording systems like stereo microphones or microphone equipped objects (for example telephone chassis with microphones) can be described in the same way.

Subsequently the interaural time difference τ is considered as the basic term for describing directions. All other binaural terms will be related to τ.

It is assumed, that the receiver (ear/microphone) is always located in the far field of the sound sources and that the distance between the receivers is small compared to the distance to the sound sources. As a consequence, a point source's wave fronts are plane at the receiver's location and are perpendicular to the line source-receiver.

4.1. The Transmission Path from the Sound Source to the Ear

Head related Signals and Outer Ear Transfer Functions

When sound from a sound source arrives at the head in the free field, the transmission path of the sound from the source to the left or the right ear can be described with the help of the outer ear transfer functions $\underline{H}_{ql}(f,\tau)$, $\underline{H}_{qr}(f,\tau)$. These functions are composed of the transfer function sound source - center of the head $\underline{H}_{qk}(f)$ and the free field outer ear transfer functions $\underline{H}_{lf}(f,\tau)$,$\underline{H}_{rf}(f,\tau)$. If only the ear signals and the free field outer ear transfer functions are known and no information about the transmission path sound source - head location is given, then only the free field sound signal at the head location $\underline{A}(f)$ can be determined instead of the sound signal at the location of the sound source $\underline{A}_q(f)$.

In the following considerations the free field outer ear transfer functions (transmission paths head position - ear) $\underline{H}_{lf}(f,\tau)$, $\underline{H}_{rf}(f,\tau)$ are split into transmission paths to a reference point "center of the head" with the transfer functions $\underline{H}_m(f,\tau)$, $\underline{H}_l(f,\tau)$ and $\underline{H}_r(f,\tau)$ (see Fig. 4.1). "Center of the head" refers to a constructed point, where the interaural transfer function $\underline{H}_{rl}(f,\tau)$ can be split into two transfer functions $\underline{H}_l(f,\tau)$ and $\underline{H}_r(f,\tau)$, which are reciprocal to each other and describe the transmission path s from the "center of the head" to the ears.

$$\underline{H}_r(f,\tau) = 1/\underline{H}_l(f,\tau) = \sqrt{\underline{H}_{rl}(f,\tau)}$$

$\underline{H}_m(f,\tau)$ describes the transmission path free field - "center of the head". By choosing transfer functions, which are related to the "center of the head" considerable algorithmic simplifications can be achieved.

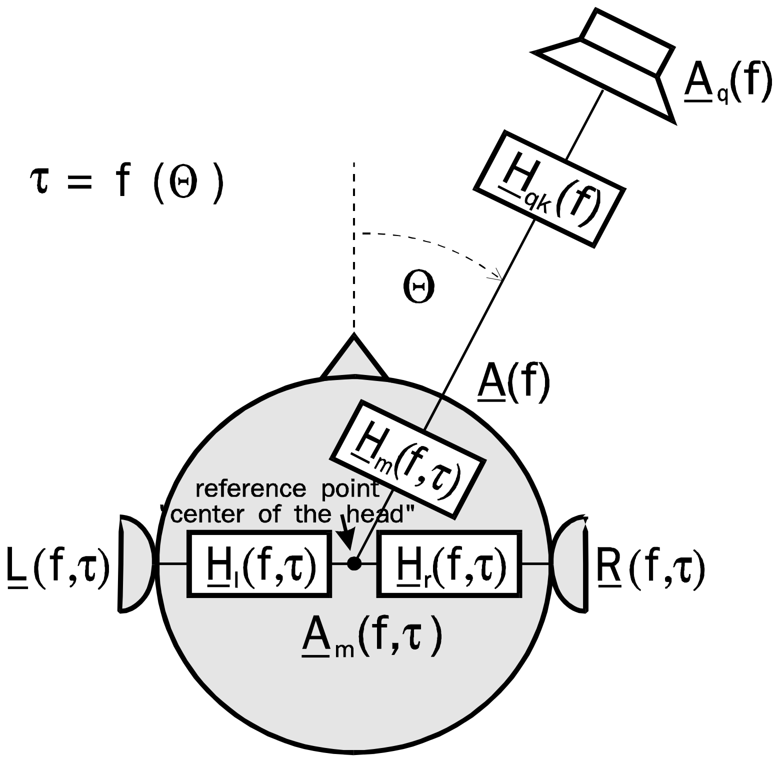

Fig. 4.1: Description of the transmission path sound source-head

All transfer functions are a function of the frequency and of the incidence angle as well. Subsequently dependencies on the incidence angle are represented by dependencies on the interaural time difference τ. The transmission paths result as follows:

Head position-"center of the head": $\underline{H}_m(f,\tau) = \sqrt{\underline{H}_{lf}(f,\tau)\,\underline{H}_{rf}(f,\tau)}$

"Center of the head"-ears: $\underline{H}_l(f,\tau) = \sqrt{\underline{H}_{lf}(f,\tau) / \underline{H}_{rf}(f,\tau)}$

$$\underline{H}_r(f,\tau) = \sqrt{\underline{H}_{rf}(f,\tau) / \underline{H}_{lf}(f,\tau)} \qquad (4.1/1)$$

Ear-ear (interaural transfer function) $\underline{H}_{rl}(f,\tau) = \underline{H}_{rf}(f,\tau) / \underline{H}_{lf}(f,\tau)$

Head position-ears (free field outer ear transfer function)

$$\underline{H}_{lf}(f,\tau) \text{ or } \underline{H}_{rf}(f,\tau)$$

Sound source - Head position (free field transfer function)

$$\underline{H}_{qk}(f,\tau)$$

The subsequent considerations can be simplified by describing the sound situation with the help of center-of-the-head transformed sound signals (for example $\underline{A}_m(f,\tau)$). Hereby the free field signal $\underline{A}(f)$ can be evaluated easily with the help of the outer ear transfer functions.

$$\underline{A}(f) = \underline{H}_m(f,\tau)^{-1}\,\underline{A}_m(f,\tau)$$

The transfer function head position - "center of the head" can be split into absolute value and phase, whereby the absolute value represents the mean damping α_o of the free field outer ear transfer function and the phase represents the mean interaural runtime τ_o. of this function.

$$\underline{H}_m(f,\tau) = e^{\alpha_o(f,\tau) + j2\pi f\,\tau_o(f,\tau)}$$

If a *microphone pair* in the free field (punctual microphones) is used as receiver instead of a (dummy) head and if the distance source - receiver is large compared to the receiver distance (plain wave fronts), then the absolute values of the transfer functions $\underline{H}_m(f,\tau)$, $\underline{H}_l(f,\tau)$, $\underline{H}_r(f,\tau)$ will be one in this special case. Then the phases will be proportional to the path difference of the incoming waves.

$\underline{H}_{lo}(f,\tau)= e^{-j\pi f\tau}$ $\quad$ $\underline{H}_{ro}(f,\tau)= e^{+j\pi f\tau}$ $\quad$ $\underline{H}_{mo}(f,\tau)= 1$

$\tau = d/c_{schall} \sin\theta$ $\quad$ d= microphone distance, c_{schall}= sound velocity $\quad$ θ= incidence angle

Interaural Phases and interaural Damping

The absolute values of the functions $\underline{H}_l(f,\tau)$ and $\underline{H}_r(f,\tau)$ are dependent on the interaural level difference ΔL, the phases are dependent on the interaural time difference τ.

$$\underline{H}_r(f,\tau) = 10^{+\Delta L(f,\tau)/40dB}\ e^{+j\pi f\tau(f)}$$

$$\underline{H}_l(f,\tau) = 10^{-\Delta L(f,\tau)/40dB}\ e^{-j\pi f\tau(f)}$$

Herewith an *interaural damping* $\alpha(f,\tau)$ and an *interaural phase* $\beta(f,\tau)$ can be defined, which allow, to describe these two transfer functions much easier.

$$\alpha(f,\tau) = \Delta L(f,\tau)\ \ln(10) / 20dB$$

$$\beta(f,\tau) = 2\pi\ f\ \tau(f) \qquad (4.1/2)$$

$$\underline{H}_r(f,\tau) = e^{+½\ \alpha(f,\tau)\ +j½\ \beta(f,\tau)}$$

$$\underline{H}_l(f,\tau) = e^{-½\ \alpha(f,\tau)\ -j½\ \beta(f,\tau)} \qquad (4.1/3)$$

Ear Signals

If the signal of a sound source is given at the head position $\underline{A}(f)$ under free field conditions, then the interaural parameters of the sound source a are $\alpha_a(f)=\alpha(f,\tau_a)$ and $\beta_a(f)=\beta(f,\tau_a)$ and the ear signals in the *frequency domain* result to:

$$\underline{R}(f,t) = \underline{A}_m(f,\tau)\ e^{+½\alpha_a(f)\ +j½\beta_a(f)} \qquad ;\ \underline{A}_m(f,t)=\underline{A}(f)\ \underline{H}_m(f,\tau_a)$$

$$\underline{L}(f,t) = \underline{A}_m(f,\tau)\ e^{-½\alpha_a(f)\ -\ j½\beta_a(f)} \qquad (4.1/4)$$

The ear signals in the *time domain* r(t) and l(t) can be evaluated by convoluting the free field sound signal a(t) with the impulse responses of the corresponding transfer functions:

$$r(t) = a_m(t) * h_r(t)$$

$$l(t) = a_m(t) * h_l(t) \qquad ;\ a_m(t)=a(t) * h_m(t) \qquad (4.1/5)$$

If for the considered input direction the free field outer ear transfer functions (which mean interaural and mean delay τ, τ_o, interaural and mean damping α, α_o) are frequency independent inside a critical band, then the convolution with the ear impulse responses can be replaced by a multiplication with a constant factor and a delay of the signals.

$$r(t) = a(t-\tau_o(\tau)+½\tau_a)\ e^{\alpha_o(\tau)+½\alpha_a} \qquad \text{for } \alpha,\ \alpha_o,\ \tau,\ \tau_o \neq f(f)$$

$$l(t) = a(t-\tau_o(\tau)-½\tau_a)\ e^{\alpha_o(\tau)-½\alpha_a} \qquad (4.1/6)$$

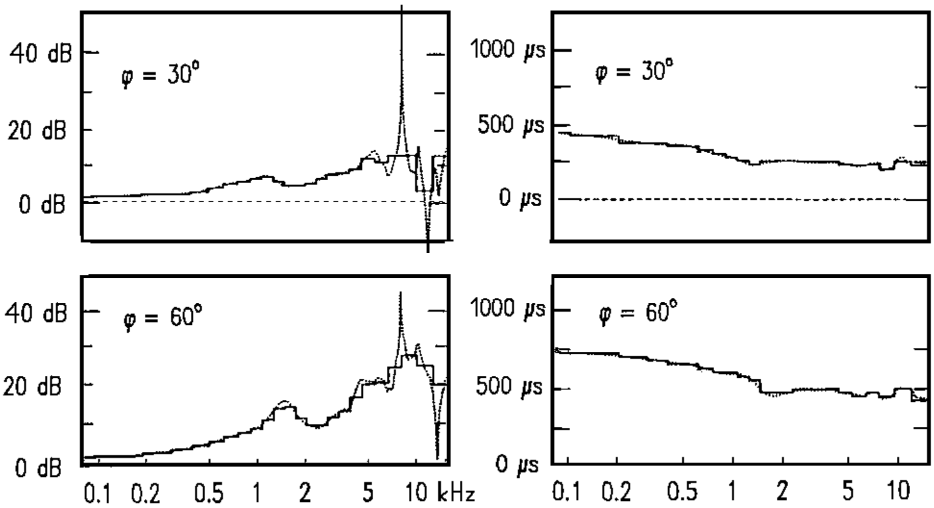

Fig. 4.2: Measured interaural level differences and phase delays for azimuth angels of 30° and 60° compared values, which have been averaged o ver a critical band.
For frequencies up to 7 kHz interaural parameters in critical bands can be described sufficiently accurate by constant values (according to GAIK [20])
•••••••••• *measured values*
------- *critical band averaged values*

For low frequencies these conditions are fulfilled in a first approximation (see Fig. 4.2). Gaik [20] has established a constant relationship between interaural parameters for a wide frequency range. After averaging over different elevation angles, the interaural level difference ΔL could be described as a frequency independent function of the interaural time difference for each critical band, whereby the deviations of this averaged function against measured values remained relatively small.

Analytic Time Signals

Subsequently mainly the description via *analytic time signals* is used. For positive frequencies the Fourier Transformation of analytic time signal $\underline{A}_a(f)$ correspond to the real time signal $\underline{A}(f)$. For negative frequencies the Fourier Transformation of the analytic time signal is zero.

$$\underline{A}_a(f) = \underline{A}(f)$$
$$\underline{A}_a(-f) = 0 \qquad ; f>0$$

Because of the symmetrical properties of the Fourier-Transformation for real and pure imaginary functions

$$\underline{X}(f) = \underline{X}(-f) \qquad \text{for } \underline{X}(f) = \mathcal{F}\{ x(t) \};\ x(t) \text{ real}$$
$$\underline{X}'(f) = -\underline{X}'(-f) \qquad \text{for } \underline{X}'(f) = \mathcal{F}\{ j\, x(t) \};\ x(t) \text{ real}$$

the Fourier-Transformation of the analytic time signal $\underline{A}_a(f)$ results to

$$\underline{A}_a(f) = \mathcal{F}\{ \mathrm{Re}\{\underline{a}(t)\} \}_{(f>0)} + \mathcal{F}\{ \mathrm{Im}\{\underline{a}(t)\} \}_{(f>0)}$$
$$\underline{A}_a(-f) = \mathcal{F}\{ \mathrm{Re}\{\underline{a}(t)\} \}_{(f>0)} - \mathcal{F}\{ \mathrm{Im}\{\underline{a}(t)\} \}_{(f>0)} \qquad ; f>0$$

Therefore the real part of the analytic time signal $\underline{a}(t)$ corresponds to the half real time signal $\frac{1}{2}a(t)$, and the imaginary part results from the Fourier-Transformation of the real time signal $\mathcal{F}\{a(t)\}$ as follows:

$$\mathrm{Im}\{\underline{a}(t)\} = \mathcal{F}^{-1}\{\mathcal{F}\{a(t)\}\,\underline{H}_a(f)\,\} \qquad \text{with} \quad \begin{array}{ll} \underline{H}_a(f)= j\frac{1}{2} & \text{for } f>0 \\ \underline{H}_a(f)= 0 & \text{for } f=0 \\ \underline{H}_a(f)=-j\frac{1}{2} & \text{for } f<0 \end{array}$$

Describing the transmission characteristics of the head results into similar formula context when describing it by analytic time signals of the ear signals $\underline{r}(t)$,$\underline{l}(t)$ than describing it by real time signals. If $\underline{h}(t)$ denotes the analytic time signal, which belongs to the impulse response $h(t)$ of the transfer function $\underline{H}(f)$, then the analytic time signals of the ear signals result to:

$$\begin{aligned} \underline{r}(t) &= \underline{a}_m(t) * \underline{h}_r(t) \\ \underline{l}(t) &= \underline{a}_m(t) * \underline{h}_l(t) \qquad ;\ \underline{a}_m(t)=\underline{a}(t) * \underline{h}_m(t) \end{aligned} \qquad (4.1/7)$$

Analogous to formula 4.1/6 the terms for the analytic time signals of the ear signals can be simplified, if the transfer functions $\underline{H}_r(f)$, $\underline{H}_l(f)$, $\underline{H}_m(f)$ remain constant within a critical band:

$$\begin{aligned} \underline{r}(t) &= \underline{a}_m(t+\tfrac{1}{2}\tau)\, e^{+\frac{1}{2}\alpha_a} \\ \underline{l}(t) &= \underline{a}_m(t-\tfrac{1}{2}\tau)\, e^{-\frac{1}{2}\alpha_a} \qquad \text{for } \alpha_a,\tau \neq \text{function of f} \end{aligned} \qquad (4.1/8)$$

The phase of the analytic time signal can be described with the help of the time dependent instantaneous frequency Ω and of the instantaneous phase Φ.

$$\arg\{\,\underline{a}_m(t)\,\} = \Omega_a(t)\,t + \Phi_a(t)$$

With $\beta_a=\Omega_a\tau$ and $|\underline{a}_m(t\pm\frac{1}{2}\tau)|\approx|\underline{a}_m(t)|$ formula 4.1/8 will be simplified to:

$$\begin{aligned} \underline{r}(t) &= |\underline{a}_m(t)|\ e^{j\Omega_a(t)t + j\Phi_a(t)}\ e^{+\frac{1}{2}\alpha_a + j\frac{1}{2}\beta_a} \\ \underline{l}(t) &= |\underline{a}_m(t)|\ e^{j\Omega_a(t)t + j\Phi_a(t)}\ e^{-\frac{1}{2}\alpha_a - j\frac{1}{2}\beta_a} \end{aligned} \qquad (4.1/9)$$

The approximations for formula 4.1/8 and formula 4.1/9 are not valid for high frequencies, where information about input angles are coded via minimums and maximums of the free field outer ear transfer function and where interaural parameters change inside critical bands (see Fig. 4.2). The frequency averaged interaural parameters get dependent on the signal spectrum.

4.2. Binaural Information at the Presence of multiple Sound Sources

Ear signals at two Sound Sources

If multiple sound sources are present, the sound signals of all source interfere. For two sound sources a and b, analogous to formula 4.1/9, the following analytic time signals of the ear signals appear at the reference point "center of the head" in the case of frequency independent free field outer ear transfer functions inside critical bands.

$$\begin{aligned} \underline{r}(t) &= a_m(t)\, e^{+\frac{1}{2}\alpha_a + j\frac{1}{2}\beta_a}\ e^{j\Omega_a t + j\Phi_a} + b_m(t)\, e^{+\frac{1}{2}\alpha_b b + j\frac{1}{2}\beta_b}\ e^{j\Omega_b t + j\Phi_b} \\ \underline{l}(t) &= a_m(t)\, e^{-\frac{1}{2}\alpha_a - j\frac{1}{2}\beta_a}\ e^{j\Omega_a t + j\Phi_a} + b_m(t)\, e^{-\frac{1}{2}\alpha_b b - j\frac{1}{2}\beta_b}\ e^{j\Omega_b t + j\Phi_b} \end{aligned} \qquad (4.2/1)$$

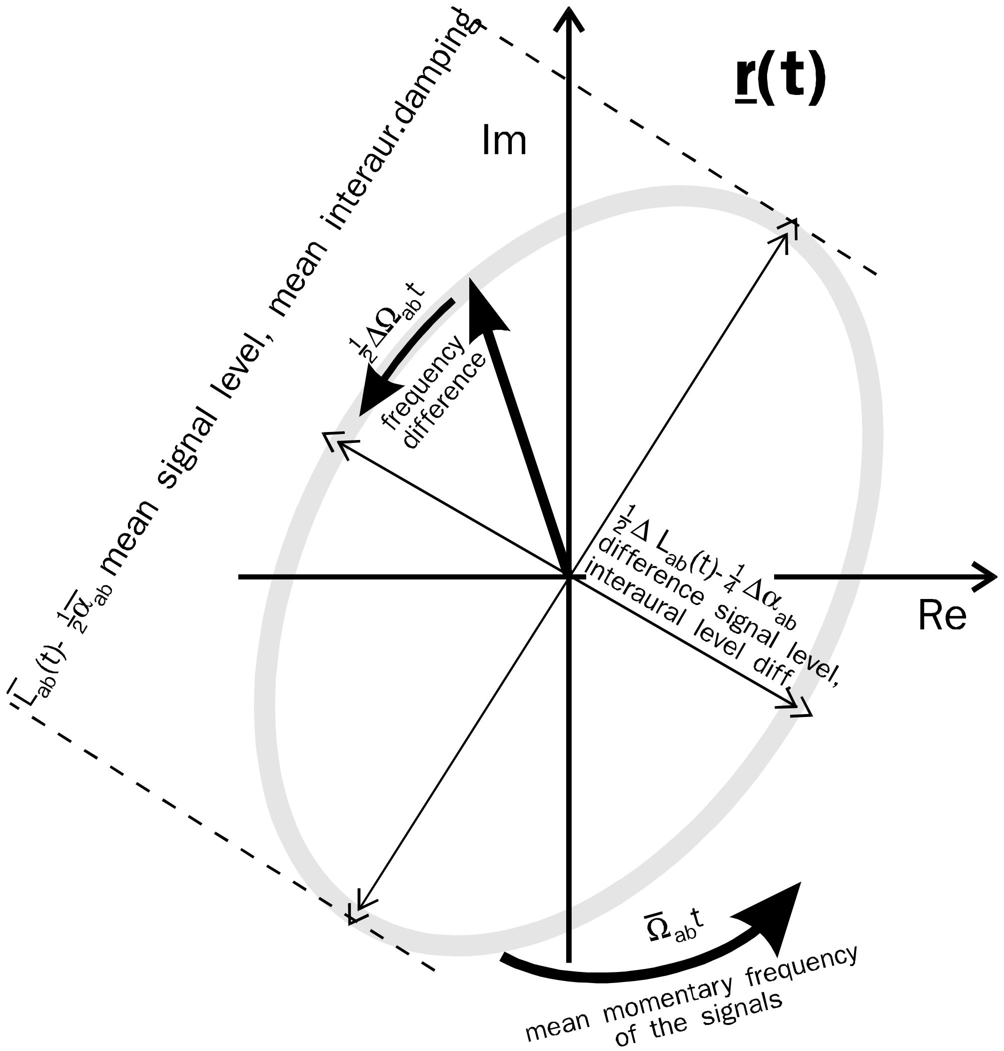

Fig. 4.3: Time locus curve of the analytic time signals of the ear signals. Here: right ear signal $\underline{r}(t)$

From this follows:

$$\underline{r}(t) = 2\sqrt{a_m(t)\,b_m(t)}\; e^{+\frac{1}{4}(\alpha_a+\alpha_b)}\; e^{j\frac{1}{2}[(\Omega_a+\Omega_b)t+\Phi_a+\Phi_b+\frac{1}{2}(\beta_a+\beta_b)]}$$
$$\cosh\Big(\ln\sqrt{a_m(t)/b_m(t)} + \tfrac{1}{4}(\alpha_a-\alpha_b) + j\tfrac{1}{2}[(\Omega_a-\Omega_b)t+\Phi_a-\Phi_b+\tfrac{1}{2}(\beta_a-\beta_b)]\Big)$$

$$\underline{l}(t) = 2\sqrt{a_m(t)\,b_m(t)}\; e^{-\frac{1}{4}(\alpha_a-\alpha_b)}\; e^{j\frac{1}{2}[(\Omega_a+\Omega_b)t+\Phi_a+\Phi_b-\frac{1}{2}(\beta_a-\beta_b)]}$$
$$\cosh\Big(\ln\sqrt{a_m(t)/b_m(t)} - \tfrac{1}{4}(\alpha_a-\alpha_b) + j\tfrac{1}{2}[(\Omega_a-\Omega_b)t+\Phi_a-\Phi_b-\tfrac{1}{2}(\beta_a-\beta_b)]\Big) \qquad (4.2/2)$$

This representation of the ear signals consists of a term, which is formed from the mean source signals and the mean interaural parameters and of a "cosh" term, which is controlled from the signal differences and the differences of the interaural parameters.

The time locus curve of a cosh-function with a complex argument and a time dependent imaginary part results into an ellipse in the complex plane. The real part of the argument determines the main axis ratio of the ellipse, the imaginary part determines the angle, under which a point of the ellipse appears to the center of the ellipse. The cosh-term gets multiplied with a function of the form $e^{j\Omega t+\Phi}$, which describes a function, which causes a time dependent rotation in the complex plane. That implies, when displaying the ear signals in the complex plane as a function of time, they result into ellipses, which rotates around the origin of the complex plane (Fig. 4.3).

Size and position of the ellipse depend on the mean signal levels and on the mean interaural damping (constant factor $2(a_m(t)b_m(t)e^{\pm½\alpha_a \pm½\alpha_b})^{½}$)
The gradient for the time t=0 depends on the mean signal phase and on the mean interaural time difference (factor $e^{j(\Phi_a+\Phi_b \pm½\beta_a \pm½\beta_b)}$)
The main axis ratio depends on the differences between the signal levels and on the interaural phase difference (real part of the cosh-term)
The (zero) phase angle depends again on the differences between the interaural phases and signal phases (time independent part of the imaginary part of the cosh-term).
The rotation frequency of the ellipse corresponds to the mean instantaneous frequency of the signals (factor $e^{j½(\Omega_a+\Omega_b)t}$).
The angular frequency by which the ellipse is passed, depends on the difference between the instantaneous signal frequencies (time dependent part of the imaginary part of the cosh-term)

The Ellipse is rotating around the origin with the mean signal frequency, but it is at the same time passed through with the difference between the signal frequencies. Since the rotation frequency is bigger than the circulation frequency of the ellipse, the time locus curve will turn to be a spiral.

The interaural Quotient

The following considerations shall give information, whether the results of the auditory experiments with two active sound sources can be explained by a simple analysis of interaural parameters or whether additional signal processing steps have to be considered.

The following interaural parameters inside critical bands will be investigated below:

- *Interaural group delays* correspond to the runtime between corresponding points in time of the signal envelopes. The course of the signal envelope corresponds to the course of the absolute value of the analytic time signal.
- *Interaural phase delays* correspond to the runtime between corresponding points in time of identical phase. This corresponds to the phase of the analytic time signal.
- *Interaural level differences* correspond to the level differences between the ear signals. This corresponds to the quotient of the absolute values of the analytic time signal.

Interaural level differences ΔL and phase runtime differences τ_Φ can be described by the quotient of the analytic time signals of the ear signals. This quotient is named *interaural quotient* $\underline{d}(t)$

$$\underline{d}(t) = \frac{\underline{r}(t)}{\underline{l}(t)} = \frac{|\underline{r}(t)|}{|\underline{l}(t)|} e^{j\arg\{\underline{r}(t)\} - j\arg\{\underline{l}(t)\}}$$

$$\Delta L(t) = 20\text{ dB } \lg |\underline{d}(t)| \qquad \tau_\theta(t) = \arg\{ \underline{d}(t)\} / \Omega \qquad (4.2/3)$$

For one sound source from the input direction θ the ear signals $\underline{r}(t)$ and $\underline{l}(t)$ inside one critical band can be described according to formula (4.1/9). Then the interaural quotient $\underline{d}(t)$ corresponds to the interaural transfer function of this input direction.

Interaural Group Delays for two Sound Sources

The interaural group delay corresponds to the runtime between two corresponding maxima of the envelope, i.e. between two corresponding maxima of the analytic time signals of the ear signals. Such maxima appear at the times t_r(right) and t_l(left), when the phase angle of the ellipse above becomes to an odd or even multiple of $\pi/2$, which means, that the imaginary part or the real part of the "cosh"-term is zero. "Odd" or "even" depends on the main axis ratio, whether it is bigger or smaller than one. With this condition $\mathrm{Im}\{ \mathrm{Arg}\{\cosh(...)\} \}=j\pi/2$, formula 4.2/2 becomes to:

$$t_r = \frac{i_r\,\pi/2 - (\Phi_a - \Phi_b) - ½(\beta_a - \beta_b)}{\Omega_a - \Omega_b} \qquad i_r \in \mathbf{G} \text{ or } i_r \in \mathbf{U}$$

$$t_l = \frac{i_l\,\pi/2 - (\Phi_a - \Phi_b) + ½(\beta_a - \beta_b)}{\Omega_a - \Omega_b} \qquad i_l \in \mathbf{G} \text{ or } i_l \in \mathbf{U}$$

With $\beta = \Omega\tau$ the interaural time difference between the envelope-maxima results to:

$$t_r - t_l = \frac{(i_r - i_l)\pi/2 - \Omega_a\tau_a + \Omega_b\tau_b}{\Omega_a - \Omega_b}$$

If a main axis ratio is given, which leads to $i_r = i_l$ (combination with minimal phase difference), the interaural group delay corresponds to the differences between the interaural phases of the signals, normalized to difference of the instantaneous frequencies. If band-pass filters are used with a bandwidth Δf_{FG} and a center frequency f_m (for example critical band filter), the minimal group delays can be estimated as:

$$t_r - t_l = -½\,\frac{\Omega_a\tau_a + \Omega_b\tau_a - \Omega_a\tau_b - \Omega_b\tau_b + \Omega_a\tau_a - \Omega_b\tau_a + \Omega_a\tau_b - \Omega_b\tau_b}{\Omega_a - \Omega_b}$$

$$t_l - t_r = ½\,\frac{\Omega_a + \Omega_b}{\Omega_a - \Omega_b}(\tau_a - \tau_b) + ½(\tau_a + \tau_b)$$

$$t_l - t_r \geq f_m/\Delta f_{FG}\,(\tau_a - \tau_b) + ½(\tau_a + \tau_b) \qquad (4.2/4)$$

The interaural group delay differs from the mean interaural time difference $½(\tau_a + \tau_b)$ only by the term $f_m/\Delta f_{FG}(\tau_a - \tau_b)$. For narrow band signals the term $f_m/\Delta f_{FG}$ becomes bigger than one. This implies that even small incidence angle differences between the sound sources can lead to interaural group delays, which are outside the natural interaural time difference range.

For third octave band filters with 25% relative band width applies $f_m/\Delta f_{FG} = 4$. For two sound sources with incident angles of +20° and -20° and corresponding normalized interaural time differences (formula 3.1/1) of ± 210 µs, interaural group delays of ± 840 µs would appear, according to formula 4.2/4. This is outside the naturally occurring range (of ± 625 µs).

Interaural Phases and Level Differences of two Sound Sources

When presenting two sound sources, then the interaural quotient, as a measure for interaural parameters, results, according to formula 4.2/1, into:

$$\underline{d}(t) = e^{½(\alpha_a + \alpha_b)}\, e^{j½(\beta_a + \beta_b)}\, \frac{\cosh\{½\ln(a_m(t)/b_m(t)) + j½(\Omega_a t - \Omega_b t + \Phi_a - \Phi_b) + ¼(\alpha_a - \alpha_b) + j¼(\beta_a - \beta_b)\}}{\cosh\{½\ln(a_m(t)/b_m(t)) + j½(\Omega_a t - \Omega_b t + \Phi_a - \Phi_b) - ¼(\alpha_a - \alpha_b) - j¼(\beta_a - \beta_b)\}}$$

(4.2/5)

Averaged over the time, the interaural quotient corresponds to the average of the interaural phases and of the interaural level differences of the involved sound sources,

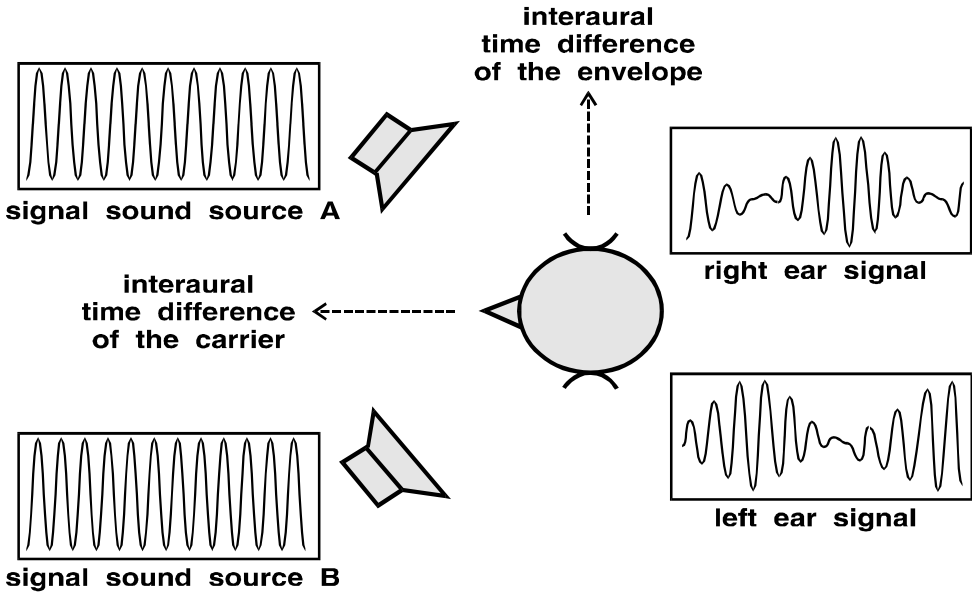

Fig. 4.4: Ear signals, interaural parameter and expected auditory events at 2 sinus signals from different directions and with different frequency.

Interaural Parameters and Results of the Auditory Experiments

Assuming that perceived directions are only determined from interaural phases, group delays and level differences, then a direct analysis of the interaural parameters would expect 2 auditory events in the sound situation of Fig. 4.4: one interaural phase and interaural level difference based auditory event from the middle between both sound sources , and one interaural group delay based auditory event with big interaural time differences, from the side or as a monaural auditory event. Similar conclusions would result from binaural models, which are based on the analysis of interaural differences.

Fig.3.11: shows the result of a model simulation of the binaural model according to Lindemann [25]. This model analyses interaural differences (phase, group delay and level differences) and shows the behavior, as expected above. Maxima of the cross correlation function appear at the middle between both sound sources and at the border of the analyzed directional range.

For untrained test persons or for very small spectral differences between the test signals the pure analysis of the interaural parameters did indeed correspond to the perceived auditory events. There had been auditory events at the extreme right and at the extreme left (perceived incidence angle >70° at loudspeaker positions ≤35°). This auditory events did not depend on the loudspeaker positions (Slatky [39]), the pitch of these auditory events was described as very low, corresponding to the period of the envelope. In addition also auditory events appeared from the middle between the loudspeakers. But, in contrast, trained test persons were able to localize both sound sources correctly, even at rather low frequency differences between the sound source signals.

As a consequence, a pure evaluation if interaural parameters is not sufficient, to explain the appearing auditory events. From the pure evaluation of the interaural parameters only auditory events from the direction of the middle between both sound sources or from the extreme right or the extreme left would be expected. But auditory events from the directions of both sound sources, as observed in the auditory experiments, cannot be explained by this approach.

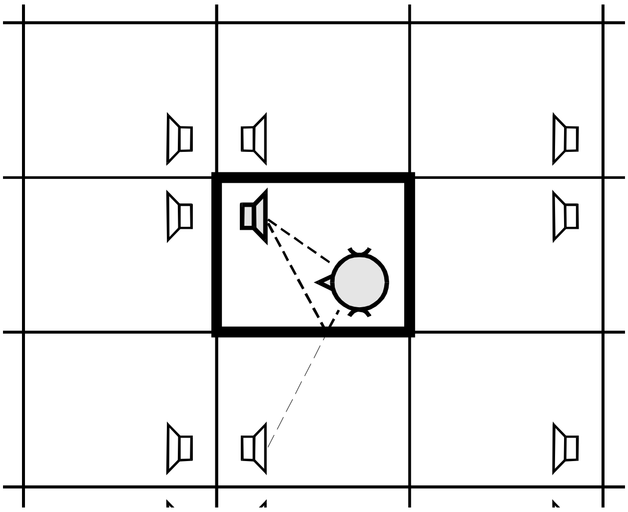

Fig. 4.5:
Mirror sound source model:
Describing reflected sound portions by mirror sound sources
- - - *reflected sound beam*
- - - *sound beam of the corresponding mirror sound source*

Sum Localization

If two identical signals $\underline{A}(f)=\underline{B}(f)$ from different directions are presented, the arguments of the cosh-terms in formula 4.2/1 become identical for both ear signals. The interaural parameters of the ear signals are then only determined from the mean interaural differences of the sound sources:

$$\underline{d}(t) = e^{½(\alpha_a+\alpha_b)}\ e^{j½(\beta_a+\beta_b)} \qquad (4.2/6)$$

According to these interaural parameters only one auditory event from the middle between both signal directions would be expected. This matches to the psychoacoustical observations.

Intensity Stereophony

Two sound sources with identical signals are arranged symmetrically ($\tau_a=-\tau_b=\tau$; $\alpha_a=-\alpha_b=\alpha$; $\beta_a=-\beta_b=\beta$) and the sound source levels are changed anti-symmetrically ($\ln|A(f)|=-\ln|B(f)|$). When introducing a term L_S for the level difference between both sound sources with $L_S=\ln\{|\underline{A}(f)/\underline{B}(f)|\}$, the interaural quotient will result to:

$$\underline{d}(t) = \frac{\cosh(\ ½L_S + ½\alpha_a + j\,½\beta)}{\cosh(\ -½L_S + ½\alpha_a + j\,½\beta)} \qquad (4.2/7)$$

Level differences between the sound sources change the real parts of the corresponding cosh-terms. This changes the main axis of the ellipses, which represents these terms, differently in both ear signals. Changes of the level differences between the sound sources (i.e. changing L_S) result into changed interaural parameters. As long as the ratio between interaural phases and interaural levels corresponds to naturally combinations a fused auditory event appears with an input angle, depending on L_S.

Reflections and Reverberation

Under the presence of reflecting surfaces the ear signals result from direct sound interfering with reflected sound portions. Early reflections have a sound field characteristics like a limited number of additional discrete sound sources. The late reverberation corresponds more likely to a diffuse sound field.

The ear signals result from an interfering of all mirror sound sources (Fig. 4.5):

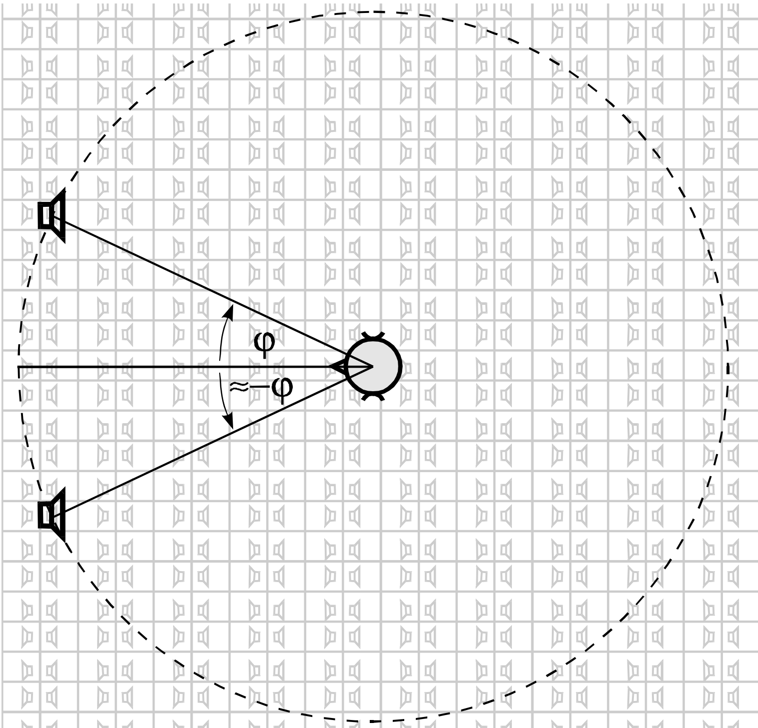

Fig. 4.6:
Mirror sound source model: within the late reverberation (meaning at big distances between the mirror sound sources and the head) the mirror sound sources can be combined to pairs with equal runtime to the receiver, but reciprocal interaural parameters.

$$\underline{L}(f) = \sum_i \underline{A}_q(f)\, \underline{H}_{qki}(f)\, \underline{H}_{lf}(f,\tau_i)$$

$$\underline{R}(f) = \sum_i \underline{A}_q(f)\, \underline{H}_{qki}(f)\, \underline{H}_{rf}(f,\tau_i)$$

In case of a symmetrical head, meaning $\underline{H}_{rf}(f,\tau)=\underline{H}_{lf}(f,-\tau)$, mirror sound sources with opposed interaural time difference i_1, i_2 can in each case be evaluated in conjunction.

$$\underline{L}(f) = \underline{A}_q(f) \sum_i \underline{H}_{qki1}(f)\, \underline{H}_{lf}(f,\tau_i) + \underline{H}_{qki2}(f)\, \underline{H}_{rf}(f,\tau_i)$$

$$\underline{R}(f) = \underline{A}_q(f) \sum_i \underline{H}_{qki1}(f)\, \underline{H}_{rf}(f,\tau_i) + \underline{H}_{qki2}(f)\, \underline{H}_{lf}(f,\tau_i)$$

If the head axis corresponds to a symmetry axis of the room, then $\underline{H}_{qki1}(f)=\underline{H}_{qki2}(f)$, and the ear signals become identical. Therefore only a localization in the median plane remains possible.

In an ideal diffuse sound field the energy is on average distributed equally over all input directions and over the time. As a consequence, each axis is a symmetry axis of the sound field. According to the considerations above, on average no interaural differences would be expected in this case. This would lead to a resulting localization in the median plane.

Inside real rooms such a sound field can approximately be expected during late reverberation, when -due to the high density of reflections- all mirror sound sources with equal runtime between sound source and head can be combined to pairs with opponent interaural differences (Fig. 4.6).

Translated to other cases this means:

- If reverberations can be described by an ideal diffuse sound field, directional information of the median plane is supported.
- Directional information, which deviates from the median plane, must arise from direct sound and early reflections.
- For binaural signal processing models this behavior could pretend additional signals from the front direction and lead to failures in processing the front direction.

4.3. Algorithms for binaural Processing of Ear Signals

Subsequently an attempt shall be made to relate the results of the auditory experiments and the considerations concerning occurring ear signals to existing binaural models and to derive from this possibilities for modeling the Cocktail-Party-Processor phenomenon.

The most important category of binaural models are cross correlation models or models of related functions. These models transform the input signals in such a way, that binaural relevant information can easily be derived from it, like input directions, interaural parameters, power of the sound signals. Signal phases are in this context of minor interest.

The cross correlation function $S_{rl}(\tau)$ between the real time functions of the right and left ear signal $r(t)$ and $l(t)$ is defined as:

$$S_{rl}(\tau) = \lim_{T\to\infty} 1/2T \int_{-T}^{T} r(t)\, l(t+\tau)\, dt$$

If $r(t)$ and $l(t)$ are time delayed representations of the same signal $s(t)$, it results to:

$$r(t)=s(t-\tau_r) \qquad l(t)=s(t-\tau_l)$$

$$S_{rl}(\tau_r-\tau_l) = \lim_{T\to\infty} 1/2T \int_{-T}^{T} s^2(t)\, dt \qquad \text{(mean power of the signal)} \qquad (4.3/1)$$

$$S_{rl}(\tau_r-\tau_l) \geq S_{rl}(\tau) \qquad \text{for all } \tau$$

A maximum of the cross correlation function appears, if the displacement parameter τ corresponds to the time delay between the signals. The absolute value at this point corresponds to the mean power of the signal. From displacement and amplitude of the maximum of the cross correlation function the interaural time difference and mean power of the signal can be determined.

The same considerations are valid for the sliding cross correlation function, but some constraints arise by using a window function $w(t)$ ($w(t)=0$ for $|t|>T$):

$$S_{rl}(\tau) = 1/2T \int_{t-T}^{t+T} r(t')\, l(t'+\tau)\, w(t-t')\, dt' \qquad (4.3/2)$$

Modifications of this algorithm have been made in order to adjust the cross correlation function better to the behavior of the human auditory system. By introducing inhibitory elements (Lindemann [25] and. Gaik [20]) or by post-processing of the correlation patterns (Stern/Colburn [43]) a reaction of cross correlation functions on interaural level differences can be achieved, similar to the human auditory system. This kind of adapted functions is particularly suitable for modeling localization phenomena. But the advantage of a auditory system adapted representation is on the other hand attended by losses in signal processing (longer computation times, no power proportional output data)

Besides that there are a couple of other models for describing binaural phenomena, which are based on other types of ear signal transformations, like the "central-spectrum-model" of Raatgever/Bilson [33] for modeling binaural pitch problems, like the EC-model of Durlach [14] for modeling binaural masking experiments. In a special position are physiological motivated models, which attempt to model binaural interactions by neuronal spike series. Wolf [49] showed, that the results of psychoacoustical models like those of Lindemann[25]/Gaik[20] can also be reproduced on the level of neuronal interactions and that these models are therefore physiologically possible.

Gaik [17] and Bodden [10] applied cross correlation models respectively binaural models for signal processing purposes, where at the presence of multiple sources the power of one source was estimated from cross correlation patterns.

In the following it shall be investigated, whether the results of the auditory experiments can be reproduced with the help of cross correlation functions or similar functions. This means, to investigate how far information about input directions and signal level of two sound sources can be achieved from the analysis of ear signals in one critical band.

The considerations are based on unmodified cross correlation functions, because they can be described mathematically in a self contained way. For signal processing purposes possibilities for a representation in the frequency domain are of special interest, because here fast computation methods exist for unmodified cross correlation functions.

4.4. Cross Correlation Models for a single Sound Source

If there is only one sound source a(t) with an interaural time difference τ, the Fourier-Transform of the sliding cross correlation function results, according formula 4.3/2, to:

$$\underline{S}_{rl}(f,t) = \underline{R}(f,t)\,\underline{L}(f,t)^{*} \ \star\ \underline{W}(f)$$

$$\underline{S}_{rl}(f,t) = \underline{A}_m(f,t)\,\underline{A}_m(f,t)^{*}\,e^{j\beta_a(f)} \ \star\ \underline{W}(f)$$

$$\underline{S}_{rl}(f,t) = \underline{A}_m{}^{+}(f,t)^2 \ \star\ \underline{W}(f) \qquad (4.4/1)$$

$\underline{A}_m{}^{+}(f,t)^2$ is the interaural cross power density of the signal $a_m(t)$. The cross correlation function corresponds here to the auto correlation function of the sound signal at the reference point center of the head, being shifted by the interaural time difference. Since the maximum of the auto correlation function appears for real signals at a displacement of zero, the maximum of the cross correlation function is shifted by its complex part, which is determined by the interaural time difference.

4.5. The Cross Correlation Function for multiple Sound Sources

The Cross Correlation Function from -∞ to ∞

If the signals of two sound sources a(t),b(t) from different input directions interfere (interaural parameters of source a: $\alpha_a(f)=\alpha(f,\tau_a)$; $\beta_a(f)=\beta(f,\tau_a)$; source b: $\alpha_b(f)=\alpha(f,\tau_b)$; $\beta_b(f)=\beta(f,\tau_b)$), the spectra of the ear signals and the cross correlation function result to:

$$\underline{R}(f) = \underline{A}_m(f)\,e^{+\frac{1}{2}\alpha_a(f) + j\frac{1}{2}\beta_a(f)} + \underline{B}_m(f)\,e^{+\frac{1}{2}\alpha_b(f) + j\frac{1}{2}\beta_b(f)}$$

$$\underline{L}(f) = \underline{A}_m(f)\,e^{-\frac{1}{2}\alpha_a(f) - j\frac{1}{2}\beta_a(f)} + \underline{B}_m(f)\,e^{-\frac{1}{2}\alpha_b(f) - j\frac{1}{2}\beta_b(f)} \qquad (4.5/1)$$

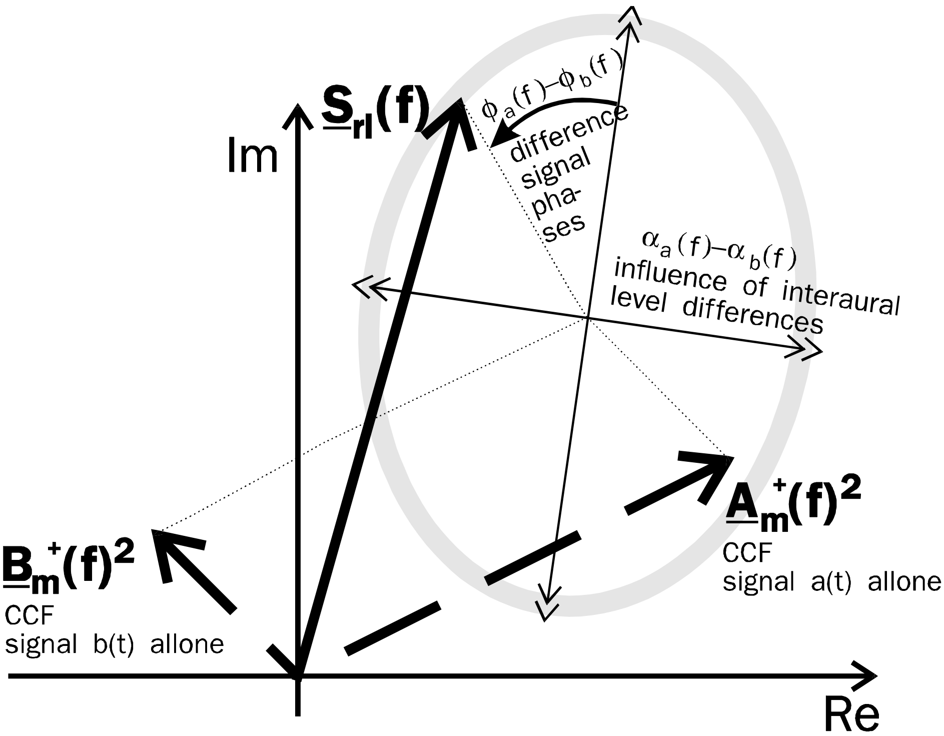

Fig. 4.7: locus curve of a frequency line of the cross correlation function $\underline{S}_{rl}(f)$ if two different sound sources are present : $\underline{A}_m^{+}(f)^2$, $\underline{B}_m^{+}(f)^2$ = cross power density of individual sound source signals

running parameter is the phase difference $\Phi_a(f)$-$\Phi_b(f)$

$$\underline{S}_{rl}(f) = |A_m(f)|^2\, e^{j\beta_a(f)} + |B_m(f)|^2\, e^{j\beta_b(f)}$$
$$+ \underline{A}_m(f)\ \underline{B}_m(f)^*\, e^{+\frac{1}{2}\alpha_a(f)\,-\frac{1}{2}\alpha_b(f)}\ e^{j\frac{1}{2}\beta_a(f)\,+j\frac{1}{2}\beta_b(f)}$$
$$+ \underline{A}_m(f)^*\, \underline{B}_m(f)\ e^{-\frac{1}{2}\alpha_a(f)+\frac{1}{2}\alpha_b(f)}\ e^{j\frac{1}{2}\beta_a(f)\,+j\frac{1}{2}\beta_b(f)}$$

With $\Phi_a(f)=\arg\{\underline{A}_m(f)\}$, $\Phi_b(f)=\arg\{\underline{B}_m(f)\}$ this results to:

$$\underline{S}_{rl}(f) = \underline{A}_m^{+}(f)^2 + \underline{B}_m^{+}(f)^2 + 2\, \underline{A}_m^{+}(f)\, \underline{B}_m^{+}(f)\ \cosh\ \tfrac{1}{2}\{\alpha_a(f)-\alpha_b(f) + j\Phi_a(f)-j\Phi_b(f)\} \qquad (4.5/2)$$

$$\underline{S}_{rl}(f) = \underline{A}_m^{+}(f)^2 + \underline{B}_m^{+}(f)^2 + 2\, \underline{A}_m^{+}(f)\, \underline{B}_m^{+}(f) \Big(\cos \tfrac{1}{2}(\Phi_a(f)-j\Phi_b(f))\ \cosh \tfrac{1}{2}(\alpha_a(f)-\alpha_b(f))$$
$$+ j\ \sin \tfrac{1}{2}(\Phi_a(f)-j\Phi_b(f))\ \sinh \tfrac{1}{2}(\alpha_a(f)-\alpha_b(f)) \Big)$$

The cross correlation function results from the sum of the cross correlation functions of the sound source signals plus a mixed term, which depends on the differences between the sound source phases and on the differences between the interaural level differences of sound source signals. (see Fig. 4.7). The locus curve of $S_{rl}(f)$, as a function of the phase difference between both source signals $\Phi_a(f)$-$\Phi_b(f)$, results into an ellipse, which center corresponds to the sum of the cross correlation functions of the individual sound source signals. The main axis depends on the geometrical mean of the cross correlation functions of the individual sound source signals and of the difference between their interaural damping.

Input directions and power of the sound sources can not be determined directly from the locus curve. Absolute value and phase of the cross correlation function depend in total on 6 parameters (amplitude and interaural phases of both sound sources, signal phase differences and interaural level difference differences between both signals). Information about the participating sound sources can only be determined directly from the cross correlation function for some special cases.

- For sound source signals with heavily structured envelopes (signals with numerous breaks inside the considered frequency range, like speech) one sound signal prevails at many times. During the breaks of one signal the cross correlation function corresponds totally to the other signal, presenting the power and the interaural delay of this signal. This results into a cross correlation function with several relative maximums. The positions and amplitudes of these maximums can be considered as estimators for the interaural delay and the mean signal power of the participating sound sources. Bodden [10] used this characteristics to locate the positions of multiple speakers from correlation patterns.
- If one sound signal is known, as well as the interaural parameters of both sound sources, absolute value and phase of an unknown source can be determined from the cross correlation function (Equation 4.5/2 corresponds in this case to a complex equation with one complex unknown). This problem can also be solved by a linear equation system over 2 receiver signals (according to equation 4.5/1).

If the interaural parameters are constant for a certain frequency range, which includes multiple frequency lines, and if the signals of adjacent frequency lines are independent from each other, an equation system for multiple frequency lines can be defined, which allows to evaluate all signal parameters and all interaural parameters.

Example: Smooth signal spectra, smooth free field outer ear transfer functions, at least one signal with uniformly distributed phase. By averaging over a wide frequency range the signal phase dependent mixed term in formula 4.5/1 can be eliminated. The interaural cross correlation function of the interfering sound source signals corresponds to the sum of the interaural cross correlation functions of the individual signals. If the interaural delays of the sound sources are known, the power of the sound sources can be evaluated.

Averaging over a wide frequency range corresponds to a reduction of the frequency resolution. This is equivalent to the insertion of a short time window, meaning, it is equivalent to a sliding cross correlation function.

If the signals are independent of each other, then also the phases of different frequency lines are independent of each other. The precondition of this analysis approach would therefore be fulfilled in many cases. This outlined approach shall be investigated deeper below.

Analysis via sliding Cross Correlation Function

The sliding cross correlation function $S_{rl}(t,\tau)$ is defined as:

$$S_{rl}(t,\tau) = 1/2T \int_{t-T}^{t+T} r(t_s)\, l(t_s+\tau)\, w(t_s-t)\, dt_s \quad ; \quad w(t)= \text{window function: } w(t)=0 \text{ for } |t|>T$$

Within the frequency domain the use of a time window $w(t)$ corresponds to a convolution with the transfer function of this window. The hereby reachable frequency resolution corresponds to the bandwidth of the Fourier-Transformation of this window (~$1/2T$). Windowed signals can therefore be described in the frequency domain without any loss of information by a line spectrum with a line distance of $1/2T$. When introducing $\underline{A}_m^+(f,t)^2$ and $\underline{B}_m^+(f,t)^2$ as the cross power densities of windowed signals, where only the corresponding sound source is present, relationships similar to formula 4.5/1 result for interfering sound sources, but with time dependent parameters.

$$\underline{S}_{rl}(f,t) = \underline{A}_m^+(f,t)^2 + \underline{B}_m^+(f,t)^2$$
$$+ 2\, \underline{A}_m^+(f,t)\, \underline{B}_m^+(f,t)\, \cosh ½\{ \alpha_a(f,t)-\alpha_b(f,t) + j\Phi_a(f,t)-j\Phi_b(f,t) \} \qquad (4.5/3)$$

Solution in the Frequency Domain

Subsequently the considerations shall be focused on interaural parameters and signal parameters, which change - compared with the window size of the cross correlation function - relatively slowly.

In the frequency domain a frequency line represents a frequency range, whose width corresponds to the transfer function of the window. If signal parameters stay constant across several window sizes, the signals can be described with a better frequency resolution than the cross correlation function. If the frequency lines of the cross correlation function do not match to the frequencies inside the signal, phase differences appear between consecutive cross correlation windows, which size depends on the difference between signal spectrum and analysis frequency lines. Also phase differences can appear between both signals, if their signal spectra don't match within the analyzed interval.

The time dependency of the signal phases $\Phi_a(t)$, $\Phi_b(t)$ can be described with the help of signal specific instantaneous angular frequencies $\Omega_a(t), \Omega_b(t)$ with $\Phi(t)=\Omega(t)t+\Phi_o$ Between the signals a time dependent difference frequency $\Omega_a-\Omega_b$ will appear. The mean value of the spectra of consecutive sliding cross correlation functions will be:

$$\underline{\mu}(f,t) = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{S}_{rl}(f,t_\mu)\, dt_\mu \qquad (4.5/4)$$

$$\underline{\mu}(f,t) = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f,t_\mu)^2\, dt_\mu \; + \; 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{B}_m^+(f,t_\mu)^2\, dt_\mu$$

$$+ \; 1/T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f,t_\mu)\, \underline{B}_m^+(f,t_\mu)\, \cosh \tfrac{1}{2}\big(\alpha_a(f)-\alpha_b(f) + j\Phi_a(f,t_\mu)-j\Phi_b(f,t_\mu)\big)\, dt_\mu$$

If A_m^+ and B_m^+ are nearly constant and if the integration time T_μ is much longer than $2\pi/[\Omega_a(f)-\Omega_a(f)]$, then equation 4.5/4 results to:

$$\underline{\mu}(f,t) = \underline{A}_m^+(f,t)^2 + \underline{B}_m^+(f,t)^2 \qquad (4.5/5)$$

Further information about the sound sources can be derived from the analysis of the standard deviation of consecutive cross correlation function windows. A complex standard deviation $\underline{\sigma}$ can be defined as:

$$\underline{\sigma}^2(f,t) = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \left(\underline{S}_{rl}(f,t_\mu)-\underline{\mu} \right)^2 dt_\mu \qquad (4.5/6)$$

$$\underline{\sigma}^2(f,t) = 2/T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m^+(f,t_\mu)^2\, \underline{B}_m^+(f,t_\mu)^2\, \cosh^2 \tfrac{1}{2}\big(\alpha_a(f)-\alpha_b(f) + j\Phi_a(f,t_\mu)-j\Phi_b(f,t_\mu)\big)\, dt_\mu$$

Under the same conditions as above ($A_m^+, B_m^+=\text{const.}; T_\mu \gg 2\pi/[\Omega_a(f)-\Omega_a(f)]$) it applies:

$$\underline{\sigma}^2(f,t) = 2\, \underline{A}_m^+(f,t)^2\, \underline{B}_m^+(f,t)^2 \qquad (4.5/7)$$

From formula 4.5/5 and 4.5/7 relationships can be derived, to gain the cross power density of the original signals from mean value and standard deviation. The result are estimators $\underline{A}_m'(f,t)$, $\underline{B}_m'(f,t)$ for the frequency lines of the original signals at the reference point "center of the head".

$$2\,\underline{A}_m'(f,t) = \sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} + \sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}}$$
$$2\,\underline{B}_m'(f,t) = \sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} - \sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}} \qquad (4.5/8)$$

The phases of $\underline{A}_m'(f,t)$, $\underline{B}_m'(f,t)$ correspond to the interaural phases of the signals, the absolute values correspond to the amplitudes of the sound source signals at the reference point "center of the head". Knowing the free field outer ear transfer functions, input directions and signal spectrums of the sound sources can be evaluated,

The box on page 44 contains a comprehensive description of this algorithm for estimating interaural phases and amplitudes (and therefore input directions and spectra) of two interfering sound sources

By analyzing cross correlation functions power and input direction of two interfering sound sources can be estimated in the frequency domain. The demands of the auditory experiments, requiring the existence of Cocktail-Party-Processor mechanisms, can therefore be fulfilled.

The use of cross correlation functions in the frequency domain for direction selective filtering has been already postulated by Gaik [17]. Gaik evaluated the current value respectively the mean value of cross correlation functions and considered a signal coming from a desired direction as being existent, if the displacement of the cross correlation function matched to a weighting window around the desired direction.

The proposed algorithm can be considered to some extend as a substantial enhancement of Gaik's algorithm. By analyzing the standard deviation additionally, the characteristics of two sound sources can be evaluated simultaneously. The second estimator can be used not only for parallel processing of two signals but also for intercepting interfering signals. With this method the parameters of a desired sound source can be evaluated in case of heavily negative signal-to-noise-ratios, when the displacement of the cross correlation function's maxima correspond nearly to the displacement of the interfering signals.(therefore lying outside the weighting window of the desired direction) and the influence of the signal of the desired direction expresses only in some slight variations of the parameters of the interfering sources.

A Cocktail-Party-Processor in the frequency domain can be build up with the help of this algorithm, if the following preconditions are fulfilled.

- quasi-stationary signals: signal parameters don't change during the integration time $2T_\mu$
- spectral differences between all sound source signals within the considered frequency ranges, meaning different signals for all sound sources.
- sufficiently long integration time T_μ for the evaluation of statistical parameters. This implies, that T_μ must be bigger than the period of the difference between the instantaneous frequencies of the signals (precondition for the simplification at the derivation of this algorithm),
- sliding cross correlation function with short time window T, in order to be able to detect changings of the signal parameters quickly and well.,
- as a consequence, the frequency resolution will be small (consequence from the last item according to the uncertainty relationship).

A Cocktail-Party-Processor in the Frequency Domain

1. Computation of the interaural short time cross correlation function from the Fourier-transformed ear signals.

   $\underline{S}_{rl}(f,t) = \underline{R}(f,t)\, \underline{L}(f,t)^* * \underline{W}(f)$

2. Computation of statistical parameters of the frequency lines of the cross correlation function

   $$\underline{\mu}(f,t) = 1/2T_\mu \int_{t-T\mu}^{t+T\mu} \underline{S}_{rl}(f,t_\mu)\, dt_\mu \qquad \underline{\sigma}^2(f,t) = 1/2T_\mu \int_{t-T\mu}^{t+T\mu} (\, \underline{S}_{rl}(f,t_\mu) - \underline{\mu}\,)^2\, dt_\mu$$

3. Estimation of amplitude and interaural phase of two sound sources from the statistical parameters of the cross correlation function

   $$\underline{A}_m'(f,t) = \frac{1}{2}\sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} + \frac{1}{2}\sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}} \qquad \underline{B}_m'(f,t) = \frac{1}{2}\sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} - \frac{1}{2}\sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}}$$

4. Evaluation the interaural time differences and power of the estimated source signals

   $$\tau_a'(f,t) = \arg\{\, \underline{A}_m'(f,t)\, \} / 2\pi f \qquad \tau_b'(f,t) = \arg\{\, \underline{B}_m'(f,t)\, \} / 2\pi f$$

   $$A'(f,t)^2 = |\, \underline{A}_m'(f,t) / \underline{H}_m(f,\tau_a')\, |^2 \qquad B'(f,t)^2 = |\, \underline{B}_m'(f,t) / \underline{H}_m(f,\tau_b')\, |^2$$

4.6. The complex Cross Product

For the algorithm above the possible rate of change of the signal parameters is limited by the integration time for evaluating the statistical parameters. The integration time should be as short as possible. The integration time can be reduced, if the time window for evaluating the cross correlation function is preferably short. Considering the length of the time window of the Fourier-Transformation as the time resolution of this method, the time resolution will become maximal, if the frequency resolution of this method is minimal. This is the case, if the frequency resolution matches to the bandwidth of the analyzed signals (e.g. a critical band width). The same relationship exists for sampled signals, too, if the sampling frequency corresponds to the bandwidth of the signal, and therefore for each sample interaural relationships are evaluated. In this case a frequency line of the Fourier-Transformation corresponds to a sample of the analytic time signal of a correspondingly bandwidth restricted signal. Fourier-Transformations of subsequent time windows generate then subsequent samples of these analytic time signal. Each frequency line of a Fourier-transformed cross correlation function corresponds then to a sample of the interaural cross product of the analytic time signals of the ear signals in this frequency range. These considerations result into the following analysis method for a Cocktail-Party-Processor

- Critical band filtering, Evaluation of the analytic time signals of the ear signals,
- Sampling of the analytic time signal correspondent to the bandwidth of the considered signals (critical band width), in order to reduce the sampling rate by this means.
- Evaluation of the interaural cross product from these samples,
- Computation of mean value and standard deviation of the interaural cross product,

- Evaluation of ("center of the head" normalized) signal amplitudes and interaural phase differences of two sound sources from a mean value and standard deviation driven equation system,
- Evaluation of the original power of the signal sources and reconstruction of the time signals of both sound sources.

There are restrictions for this kind of analysis. One limiting factor is the time constant T_μ for the evaluation of the statistical parameters of the cross product:

- The solution above is only valid, if the "cosh" mixed term in equation 4.5/1 disappears. For this purpose the time constant for the evaluation of the statistical parameters must be longer than the period of the frequency differences between the signals: $2T_\mu > 2\pi/[\Omega_a - \Omega_b]$. This frequency difference characterizes the difference between the phase-change-rates of the signals. It can be considered as a measure for the dissimilarity of the source signals.

  Similar restrictions are valid for the ability of humans, to separate signals of different directions (see auditory experiments chapter 3). Only if there is a minimal frequency difference between the signals, test person could localize two active sound sources. Around 500 Hz the minimal frequency difference was about 30 Hz, this corresponds to a period of about 30 ms.

- Spectrum and interaural parameters of the considered signals may not change during the analysis time $2T_\mu$.
- The method cannot give any information about signal phases. A reconstruction of the source signals from a mixture of signals is only possible for the signal power. Phase errors and frequency deviations are therefore possible.

  This corresponds to the results of the auditory experiments. For frequency differences below a critical band width test persons could localize the sound sources, but they could not extract the sounds of both signals from the mixture of signals.

- The method doesn't give any information about interaural level differences. Within the higher frequency range, where the interaural phase becomes ambiguous, ambiguities in the estimated signal direction can appear. This can lead to errors in choosing the appropriate free field outer ear transfer function, and the signal power can no longer be reconstructed correctly from the normalized absolute values of the estimators. (for more information about correcting this error see chapter 6.)
- At sound situations with 2 sound sources direction and power of each sound source can therefore be evaluated. At three or more sound sources the estimated power and directions do not correspond to the given sound source arrangement (for more information about correcting these estimation errors see chapter 5.)

The limitations, which are caused by the integration time $2T_\mu$, mark the absolute limits of this method, but other limitations can be removed.

- Interaural level differences can be analyzed by a similar algorithm (applying the statistical method as described above onto the auto correlation function and the amplitudes of the ear signals).
- At three or more active sound sources or at diffuse sound fields correction methods for the direction estimation can be developed. At sound situations with multiple sound sources the direction estimators become time variant. From the analysis of these time dependencies correction factors for the signal power estimators can be derived.

Subsequently a binaural signal processing model shall be presented, which uses all these signal processing steps and which can be used for solving the Cocktail-Party-Processor problem. The characteristics of these algorithms, the reactions on certain signals and possibilities for optimizing this system will be discussed.