5. Algorithms for the Evaluation of Interaural Phase Differences ("Phase-Difference-Cocktail-Party-Processor")

5.1. Requirements for a Cocktail-Party-Processor

A system for the analysis of ear signals shall match the following requirements:

A Cocktail-Party-Processor shall be adapted to the signal processing of the auditory system:

- The processing of ear signals shall be possible.
- Signals shall be processed inside critical bands.
- Time constants shall be adapted to the human auditory system.
- The power and input directions of at least 2 sound sources shall be evaluated from a mixture of sound sources (according to the auditory experiments, chapter 3). A loss of information related to phase and frequency information would therefore be tolerable.

The algorithm shall be able to work under a big variety of acoustical conditions:

- Processing of microphone recordings (no level difference between the receivers), head related recording as well as signals from microphone arrays,
- Processing of signals with a constant amplitude (harmonic tones) and time dependent amplitude (noise, speech),
- Processing of configurations with 1, 2 or more sources.
- Processing under different spatial environments: free field, closed rooms with reflections and reverberation, diffuse sound field.

The system shall be applicable in conjunction with different signal processing methods:

- It shall be possible to process signals from data reduction algorithms (see chapter 7).
- The information content of ear signals shall as far as possible not be decreased and it shall be possible to forward the result to further analysis steps.
- The algorithm shall be describable in a mathematically closed form..

Subsequently a system shall be presented, which is based on the analysis of the interaural cross product and which fulfills these requirements. The results of this system shall be discussed for different signal and environmental conditions (for the integration into a package of binaural signal processing see chapter 7 and the overview scheme 7.3-1, respectively)

5.2. The interaural Cross Product

Definition

According to the considerations in chapter 4 the interaural cross product $\underline{k}(t)$ between band-pass-filtered analytic time signals of the ear signals $\underline{r}(t)$, $\underline{l}(t)$ will be used as analysis method. It is defined as:

$$\underline{k}(t) = \underline{r}(t)\ \underline{l}(t)^*$$

$$\underline{k}(t) = |r(t)|\ |l(t)|\ e^{j(\Phi_r - \Phi_l)}$$

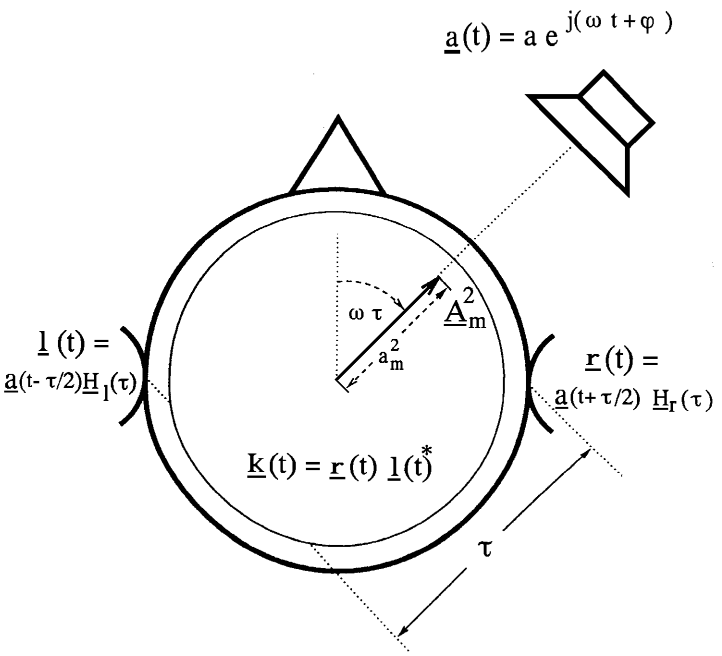


Fig. 5.1:
The interaural cross product k(t) for one sound source with constant amplitude.

The absolute value of the interaural cross product corresponds to the product of the ear signal amplitudes, the phase corresponds to the phase difference between the ear signals. Hereby information about the absolute phase and about the absolute frequency gets lost as well as information about level differences between the ear signals. The loss of phase and frequency information by the Cocktail-Party-Processor-Algorithm is not in contradiction to the auditory experiments (chapter 3), whereby the auditory system can extract input directions (interaural time/level differences) and signal power of single sound sources, but not phase, frequency or sound information of these sources.

Since this algorithm does not evaluate interaural level differences, additional algorithms for processing interaural level difference are necessary for a complete binaural analysis (see chapter 6).

One Sound Source with constant Amplitude, Source Vectors

If one sound source is present and the interaural transfer functions (formula 4.1/9) remain constant inside critical bands, the interaural cross product results into:

$$\underline{k}(t) = |a_m(t)|^2 \, e^{j\beta_a} = \underline{A}_m(t)^2 \qquad (5.2/1)$$

The absolute value of the interaural cross product is proportional to the mean power of the source signal at the reference point "center of the head" (chapter 4.1). The phase corresponds to the interaural phase, it if therefore proportional to the interaural time difference. For sound sources with a constant amplitude the locus curve results into a point in the complex plane (Fig. 5.1). Power and interaural time difference of the source can be evaluated from the locus curve. Signal phase and interaural level difference have no influence on the locus curve. The interaural cross product of such a signal a(t) at the reference point "center of the head" will be named *source vector* $\underline{A}_m(t)^2$ below.

The locus curve of the interaural cross product corresponds to a representation of the interaural cross correlation function according to formula 4.3/1 and 4.4/1 in polar coordinates with a phase being proportional to the displacement time of the cross correlation function.

This correspondence is only valid for signals with constant amplitude and constant interaural phase or for short integration times of the cross correlation function. For time variant signal

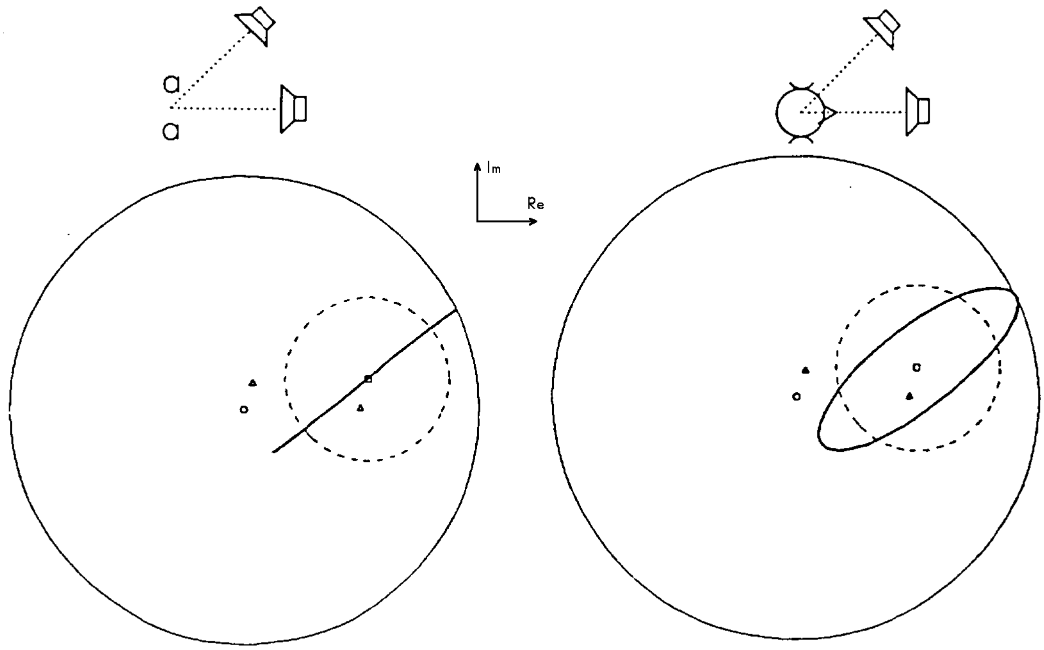


Fig. 5.2:Locus curve of the interaural cross product $\underline{k}(t)$ (formula 5.3.1/1) for 2 sound sources with constant signal amplitude, but different directional information.
Parameter of the locus curve is the time.
Signal sound source A: Sinus 500 Hz, Amplitude=1 , τ_a= 0µs; $\underline{k}_A(t)$=const.
Signal sound source B: Sinus 560 Hz, Amplitude=0.5, τ_b=400µs; $\underline{k}_B(t)$=const.
——— Time- locus-curve of the interaural cross product $\underline{k}(t)$
- - - - - - Circle with radius $|\underline{\sigma}|$ (standard deviation) around the complex mean value $\underline{\mu}$
Left: no interaural level differences $\Delta L_a=\Delta L_b$=0dB; Right: ΔL_b= 6dB

parameters the cross correlation function will only represent the mean value of power and interaural phase, whereas the interaural cross product will track the course exactly.

5.3. The Phase Difference-Cocktail-Party-Processor

5.3.1. The interaural Cross Product at two Sound Sources

If 2 sound signals a(t),b(t) are emitted from different directions, the ear signals of the particular signals superpose. If the interaural differences inside a critical band can be considered as being constant, (low frequency description according to formula 4.2/1), the analytic time signals of the ear signals result to:

$$\underline{r}(t)= a_m(t)\, e^{j\Omega_a t+j\Phi_a}\; e^{+\frac{1}{2}\alpha_a+j\frac{1}{2}\beta_a} + b_m(t)\, e^{j\Omega_b t+j\Phi_b}\; e^{+\frac{1}{2}\alpha_b+j\frac{1}{2}\beta_b}$$

$$\underline{l}(t)= a_m(t)\, e^{j\Omega_a t+j\Phi_a}\; e^{-\frac{1}{2}\alpha_a -j\frac{1}{2}\beta_a} + b_m(t)\, e^{j\Omega_b t+j\Phi_b}\; e^{-\frac{1}{2}\alpha_b -j\frac{1}{2}\beta_b}$$

With the cross products of the particular source signals and with the source vectors $\underline{A}_m^2=|a_m|^2e^{j\beta_a}$, $\underline{B}_m^2=|b_m|^2e^{j\beta_b}$ (formula 5.2/1) the interaural cross product results to::

$$\underline{k}(t)= \underline{A}_m(t)^2 + \underline{B}_m(t)^2 + 2\, \underline{A}_m(t)\, \underline{B}_m(t) \cosh\big(j(\Omega_a-\Omega_b)t + j(\Phi_a-\Phi_b) + \tfrac{1}{2}(\alpha_a-\alpha_b)\big)$$
(5.3.1/1)

Table 5.1: Influence of Sound-Source-Parameters on the Locus Curve of the interaural Cross Product (Ellipse)

Ellipse-Parameter	influencing Signal Parameter
- Center	Sum of the locus curves of the particular signals (power and input directions of the sound sources)
- Largest principal axis	Geometric average of the signal powers
- Orientation (phase zero)	Mean interaural phase difference
- Angular frequency	Difference between the instantaneous frequencies
- Phase	Difference between the phases of both signals
- Principal axis ratio	Interaural level differences

The time locus curve (Fig. 5.2) results into an ellipse. The Influences of the signal parameters on the ellipse are depicted in Table 5.1.
For microphone recordings (α=0) the locus curve results into a line..

5.3.2. The Cocktail-Party-Processor-Algorithm

With the help of the complex mean value $\underline{\mu}$ and the complex standard deviation $\underline{\sigma}$ of the locus curve an equation system can be set up, which can be used to evaluate the amplitudes and the interaural time differences of both participating sound sources. The complex mean value $\underline{\mu}$ and the complex standard deviation $\underline{\sigma}$ are defined as follows:

$$\underline{\mu} = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{k}(t_\mu)\, dt_\mu$$

$$\underline{\sigma}^2 = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} (\,\underline{k}(t_\mu) - \underline{\mu}\,)^2\, dt_\mu \qquad (5.3.2/1)$$

With the interaural cross product from formula 5.3.2/1, the complex mean value results to:

$$\underline{\mu} = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m(t_\mu)^2\, dt_\mu + 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{B}_m(t_\mu)^2\, dt_\mu$$
$$+ 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \underline{A}_m(t_\mu)\, \underline{B}_m(t_\mu) \cosh\Big(j(\Omega_a(t_\mu)-\Omega_b(t_\mu))t_\mu + j\Phi_a - j\Phi_b + ½\alpha_a - ½\alpha_b \Big)\, dt_\mu$$

When choosing the integration time $2T_\mu$, so, that it is on the one hand small compared to the change rate of the signal parameters and on the other hand large compared to the period of the signal's instantaneous -frequency-difference, the statistical parameters result as follows:

$$\underline{\mu} = \underline{A}_m(t)^2 + \underline{B}_m(t)^2$$

$$\underline{\sigma} = \sqrt{2}\; \underline{A}_m(t)\, \underline{B}_m(t) \qquad (5.3.2/2)$$

Herewith an equation system can be set up, from which the so called *source estimator* can be evaluated. A complex source estimator $\underline{A}_m'(t)^2$ or $\underline{B}_m'(t)^2$ estimates for the reference point "center of the head" from sound field parameters the corresponding source vector (formula 5.2/1). The absolute value of the source estimator is proportional to the power of the estimated sound source, the phase is proportional to the interaural phase. Herefrom also estimators for the amplitudes of the sound sources, the *amplitude estimators* $a_m'(t)$, $b_m'(t)$ can be derived:

$$(\underline{A}_m'(t) \pm \underline{B}_m'(t))^2 = \underline{\mu} \pm \sqrt{2}\,\underline{\sigma}$$

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

$$\underline{B}_m'(t) = \frac{1}{2}\sqrt{\underline{\mu} + \sqrt{2}\underline{\sigma}} - \frac{1}{2}\sqrt{\underline{\mu} - \sqrt{2}\underline{\sigma}} \qquad (5.3.2/3)$$

$$a_m'(t) = |\underline{A}_m'(t)| \qquad \beta_a' = 2\arg(\underline{A}_m'(t))$$

$$b_m'(t) = |\underline{B}_m'(t)| \qquad \beta_b' = 2\arg(\underline{B}_m'(t))$$

If the instantaneous frequencies of the sound sources are known, the interaural time difference can be determined from the interaural phase and herefrom with the help of the outer ear impulse responses the power of the sound sources in the free field $a'(t)^2$, $b'(t)^2$:

$$\tau_a' = \beta_a/\Omega_a \qquad \tau_b' = \beta_b/\Omega_b$$

$$a'(t)^2 = \left(a_m'(t) * h_m(\tau_a')^{-1}\right)^2 \qquad b'(t)^2 = \left(a_m'(t) * h_m(\tau_b')^{-1}\right)^2$$

This estimation algorithm will be called *Phase-Difference-Cocktail-Party-Processor* below. Inside the box on page 51 the method is presented comprehensively.

From the estimated powers an estimated time signal of a sound source can be evaluated by processing the original ear signals with the help of a Wiener-Filter-Algorithm (see chapter 7.2).

Properties of the Algorithm

When using the Phase-Difference-Cocktail-Party-Processor for signal processing tasks the following basic conditions have to be considered:

- Envelopes and input directions of the sound sources may only change slowly compared to the integration time. For used integration times of 10..30 ms envelope modulation frequencies of 30..100 Hz would be evaluable. The analysis of speech signals would therefore be possible without restrictions.
- The power of the sound sources is estimated for the reference point "center of the head" (chapter 4.1). It corresponds to the geometrical average over both ear signals. Knowing the free field outer ear transfer function, the free field power or the loudness of the signals can be evaluated from it. According to Remmers/Prante [34] these items depend on the sum of the ear signal powers.
- For lateral displaced sound sources the relationship between (estimated) interaural phase and (needed) interaural time difference is frequency dependent. Therefore errors can appear at the estimated directions, if the instantaneous frequencies of the signals are unknown (up to ±12% for third octave band filters). For sound sources with a constant direction the quality of the estimation can be improved by averaging the interaural time difference estimators over the time or frequency. For sources in the median plane this error does not appear.
- For natural ear distances there are ambiguities above 800 Hz for determining the interaural time difference from the interaural phase. Processing of signals of higher frequencies remains possible, as long as the ambiguities of the existing sound sources don't overlap. These ambiguities can be overcome, if additional information (for example from the evaluation of interaural level differences) can also be taken into account. (see chapter 6).

The Algorithm of the Phase-Difference-Cocktail-Party-Processor

1. Evaluation of the interaural cross product from the analytic time signals of the ear signals

   $\underline{k}(t) = \underline{r}(t)\ \underline{l}(t)^*$

2. Evaluation of statistical parameters of the interaural cross product

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

3. Estimation of sound source amplitudes and directions of two sound sources

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

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

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

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

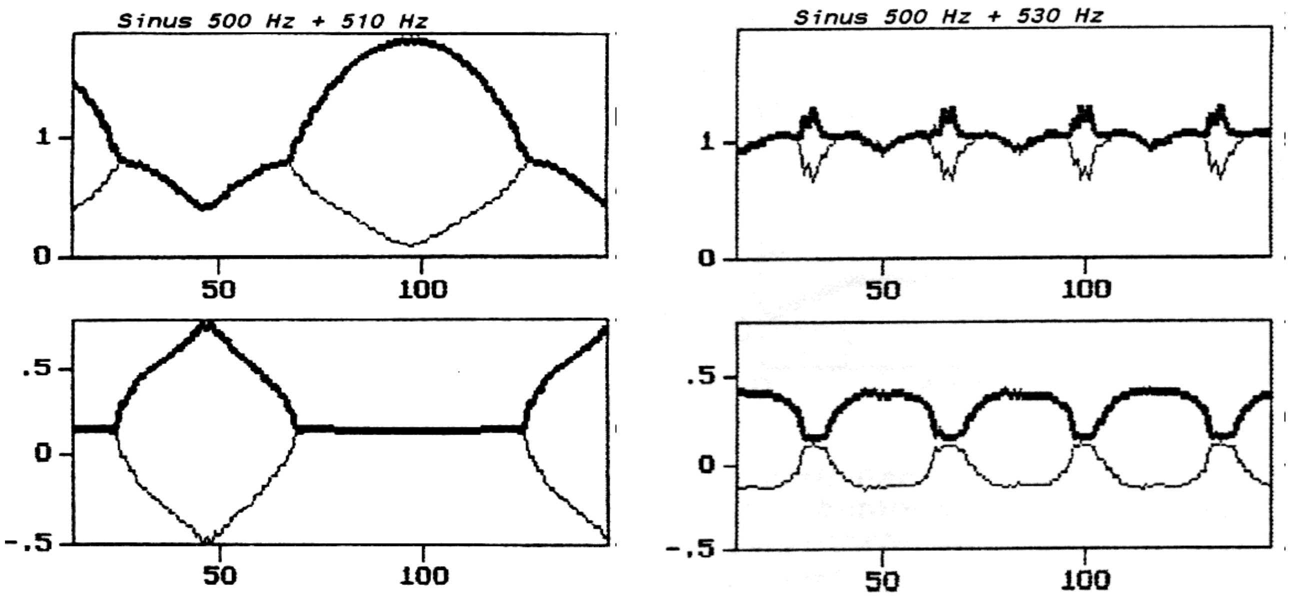
5.3.3. Signal Processing Examples

Processing of the Signals of the Auditory Experiments

Fig. 5.3 shows the estimators of the Phase-Difference-Cocktail-Party-Processor for the signals of the auditory experiments (sinus signals around 500 Hz, signals inside the main critical band). The input directions (+35°, -5°) are simulated using normalized interaural time differences according to formula 3.1/1 (350 µs and -50 µs) and the associated interaural level differences according to appendix C. Depicted are the absolute values and the interaural phases of the estimators.

For low frequency differences (signals: 500 Hz and 510 Hz) amplitude and interaural time difference of the estimators get extremely time variant.. The estimated interaural time difference moves between -500 µs and +700 µs, it covers nearly the whole naturally occurring range. The estimators match to the results of the auditory experiments very well, since: for low frequency differences very diffuse or through the room moving auditory events have been perceived. (chapter 3.2).

For signals of 500 Hz and 530 Hz the fluctuation of the estimators becomes smaller.. The estimated interaural time differences and sound source amplitudes match temporarily to the corresponding sound source parameters, but there are relatively strong deviations of up to ±200 µs for the interaural time differences and of up to ±30% for the signal amplitudes. At the corresponding auditory experiments the test persons could determine the input direction of the sound sources correctly, although with a good deal of trouble. In the Cocktail-Party-Processor-Model the troubles of the test persons could be put into relationship with the periods, where the estimated input directions and amplitudes match with the sound source characteristics.



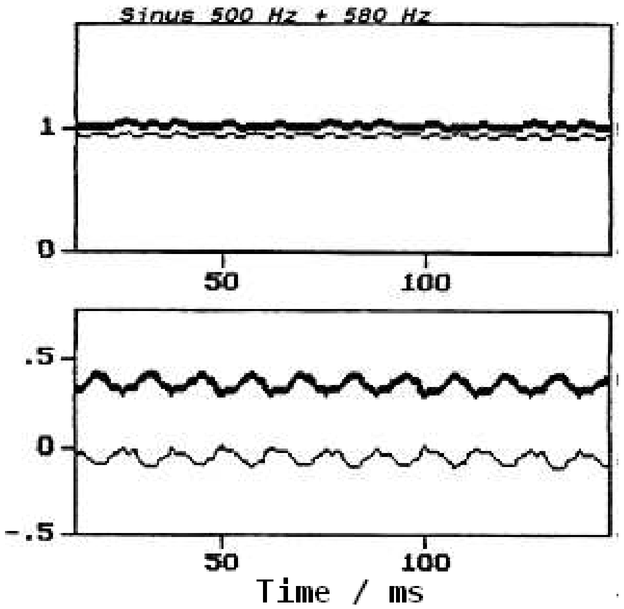


Fig. 5.3:
The Phase-Difference-Cocktail-Party-Processor at the signals of the auditory experiments.: Estimated signal amplitudes (upper diagrams) and interaural time differences (lower diagrams) in dependency of the time
Signal a: *sinus 500Hz,* *amplitude=1,*
τ_a*= -50µs* (θ*= -5°)*
Signal b: *sinus 500Hz+*Δf*, amplitude=1,*
τ_b*=350µs* (θ*=35°),* ΔL*=3dB*
Model parameters: T_μ*=28 ms*
1 estimator per ms
——— *estimator A*
——— *estimator B*
top left : Δf*= 10 Hz*
top right: Δf*= 30 Hz*
bottom left: Δf*= 80 Hz*

The Phase-Difference-Cocktail-Party-Processor matches therefore the requirements for binaural models, which arise from the auditory experiments: The processor can, as required, evaluate the directions and amplitudes of two sound sources simultaneously, and is beyond that able, to reproduce the results of the auditory experiments qualitatively.

For signals of 500 Hz and 580 Hz the estimated amplitudes, as well as the estimated interaural time difference s match to the parameters of the sound sources. In the corresponding auditory experiments the test persons could mostly determine the directions of the sound sources correctly.

Estimation of the Power of Sound Sources

The quality of estimation for stationary signals and negative Signal-to-Noise-Ratios is depicted in Fig.5.4 (presentation of two sinus signals with different frequencies and different directional parameters inside a critical band). Up to signal-to-noise-ratios of -100 dB the power of a desired source can be estimated with an error <1 dB and the direction can be estimated with an error <70 µs.

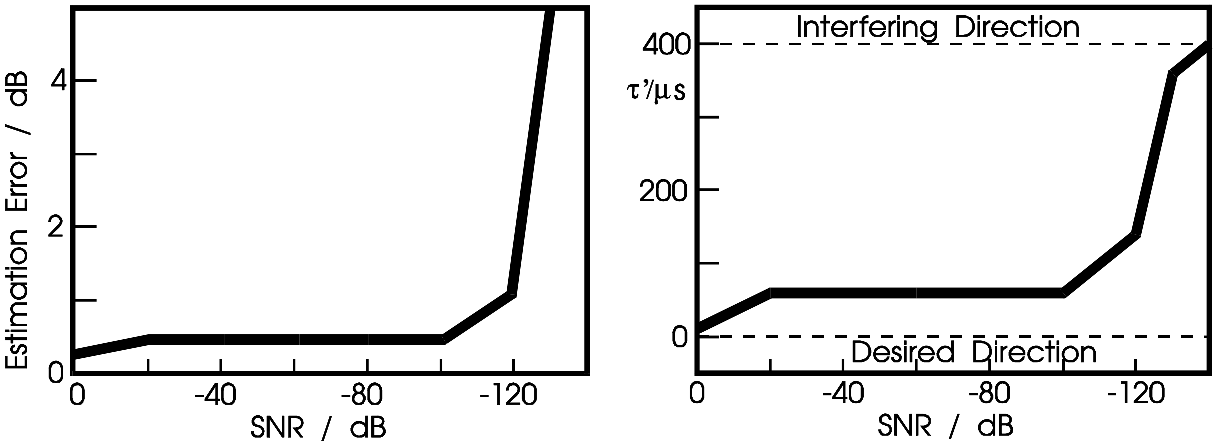


Fig. 5.4:The Phase-Difference-Cocktail-Party-Processor at stationary signals and negative signal to noise ratios (SNR).
Left figure: estimation error of the signal amplitude,
Right figure: estimated interaural time difference .
Desired signal: sinus 560 Hz, amplitude= 1..10^{-7}, τ_{nutz}= 0 µs
Interfering signal: sinus 500 Hz, amplitude= 1, $\tau_{stör}$= 400 µs; $2T_{\mu}$= 20 ms

This means, that two stationary signals from different directions can be analyzed direction selectively nearly up to the limit of computational accuracy. For this algorithm all those signals appear as stationary signals, whose amplitude, frequency and direction remains constant during the integration time of the statistical parameters (here 20 ms).

Fig. 5.5 shows the behavior of this algorithm for non-stationary signals and an averaged signal-to-noise-ratio of -10 dB. The estimator for the desired direction follows here relative exactly to the envelop of the desired signal. The power of the desired signal can be estimated with only very small errors even for occasionally appearing signal-to-noise-ratios of -24 dB. Estimators inside a range of ±70 µs around the desired direction have been considered as estimators of the desired signal.

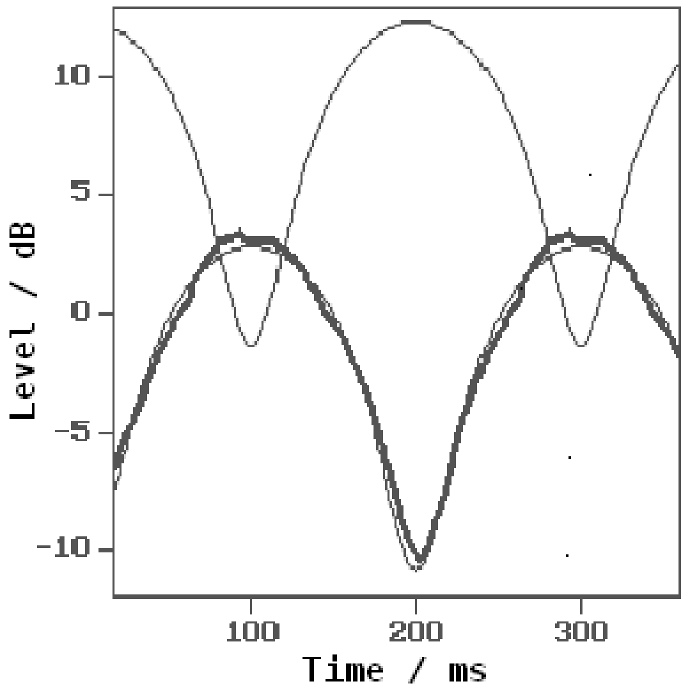


Fig. 5.5:
Estimation of the amplitudes of an amplitude modulated desired signal Which is disturbed by a stronger interfering signal

Signal-envelopes
desired signal: Sinus 560Hz,
f_{mod}=5Hz, τ=400µs
interfering signal: Sinus 500 Hz
f_{mod}=5Hz, τ= 0µs
estimated envelope of the desired signal

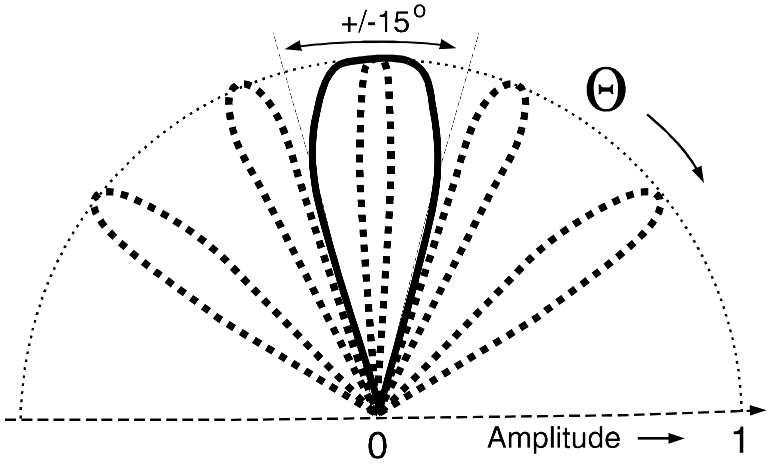


Fig. 5.6:
Direction selectivity of the Phase-Difference-Cocktail-Party-Processor

Desired direction: in front,
Lock-in-range: ±0.07π.
Test signals:
Sinus 500 Hz;
Sinus 2 kHz
Width of the directional lobes::
±15° at 500 Hz
± 4° at 2 kHz

Estimation of Directions

Since the interaural phase of a dedicated input direction is frequency dependent, these frequency dependent deviations have to be taken into account for signal processing tasks. Therefore a lock-in-range is defined around the interaural phase of the desired direction. Estimators with interaural phases inside this lock-in-range are counted as estimators of the desired direction. The width of this lock-in-range determines the directional selectivity of this method. It depends on the bandwidth of the signals and on the input direction. Fig. 5.6 shows for third-octave-band filtered signals (used lock-in-range 0.07π) the resulting estimators of the desired direction "in front", as a function of the signal input direction. For frequencies around 500 Hz the directional lobe covers an angular range of ±15° around the desired direction. For higher frequencies the directional lobe gets smaller. For frequencies above 800 Hz and naturally ear distances the interaural phase becomes ambiguous and a bundle of lobes occurs, the interaural time difference of two neighbor lobes differs by one period of the signal frequency.

For signal processing tasks delays can be added to the ear signals in order to generate an interaural time difference of zero for the desired direction. Then the frequency driven variance of the interaural phases becomes minimal, and small lock-in-ranges and small directional lobes can be used.

Fig. 5.7 illustrates the capability of this method, to estimate the interaural parameters and therefore the directions of two speakers and to assign the estimated signal power to the speaker positions. Displayed is the estimated signal energy in dependence upon the correspondingly estimated interaural time difference. The directions (i.e. the interaural time differences) can therefore be estimated quite precisely in the frequency range around 500 Hz, very pronounced energy maximums appear at the interaural time differences of the speaker directions. This figure shows also, how the estimated signal energy is distributed over the speakers. Even in the frequency range around 2500 Hz , where the interaural phases of both speakers are quite similar because of ambiguities, the method succeeds to distribute the energy onto two distinct speakers, which are very close together. Even in the presence of phase ambiguities the Cocktail-Party-Processor can determine the signal power of both sources and separate the signals direction selectively. For direction separation however a correction method from chapter 5.6 (validity range of the estimators) was used here. The solid lines show the results with correction method, the dotted lines without.

A signal processing example with this processor is documented in chapter 7.3.2.

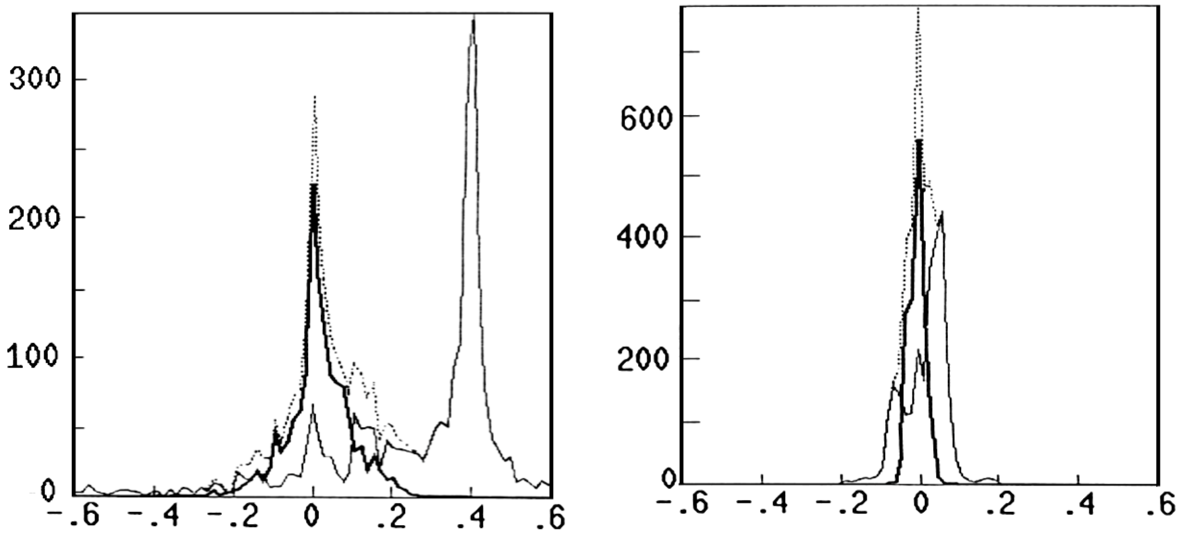


Fig. 5.7: Assignment of signal power to input directions
2 speaker with the same power in the free field, 2.7 seconds continuous text,
speaker A in front (τ_a=0µs); speaker B from 45° right (τ_b=400µs);
Y-axis: direction dependent summed up total energy of the estimators,
X-axis: corresponding interaural time difference of the estimators in ms,
——— estimator for speaker A, τ_a=0µs
——— estimator for speaker B, τ_b=400µs
············ estimator for speaker A without correction method, (see chapter .5.6)
left : critical band around 500 Hz right: critical band around 2.5 kHz
parameters of the processing according to chapter 7.3.2

The capabilities of the Phase-Difference-Cocktail-Party-Processor to evaluate the input direction and signal power of sound sources exceed for stationary signals by far the capabilities of the human auditory system. For signals with time variant amplitude the signal processing capabilities are at least of the same order, but can be enhanced yet by further post-processing methods (see chapter 5.6).

The ambiguities of the interaural phases, which appear at higher frequencies for natural ear distances, can be overcome by additional estimation algorithms (Level-Difference-Cocktail-Party-Processor, chapter 6) or by multi-microphone-arrangements with different microphone distances for different frequency ranges.

Further algorithms to evaluate the input direction and signal power of involved sound sources from statistical parameters of the interaural cross product are described in appendix E. At comparative tests, however, the above described algorithm has turned out to be the most precise and the fastest one.

5.3.4. Single Sound Source with time variant Amplitude

For a sound source with time dependent amplitude the absolute value of the locus curve $\underline{k}(t)$ changes, while the phase remains constant. The locus curve results into a line segment through the origin.

If there is an amplitude modulated sound source with the source signal $a_x^2(t)=a_o^2+a_1^2\cos(\Omega_x t+\Phi_x)$ (a_o^2=mean value, $2\pi/\Omega_x<1/2T_\mu$), and the period of the amplitude modification is small against the integration time $2T_\mu$, then mean value and standard deviation of the interaural cross product result to:

$$\underline{\mu} = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} a_x(t_\mu)^2\, e^{j\beta_a}\, dt_\mu = a_o^2\, e^{j\beta_a}$$

$$\underline{\sigma}^2 = 1/2T_\mu \int_{t-T_\mu}^{t+T_\mu} \left(a_x(t_\mu)^2\, e^{j\beta_a} - \underline{\mu}\right)^2 dt_\mu = \tfrac{1}{2}\, a_1^4\, e^{j2\beta_a}$$

The source estimators (according formula 5.3.2/3) result into two estimators with identical interaural phase. From this the following amplitude estimators arise:

$$(\underline{A}_m' \pm \underline{B}_m')^2 = (a_o^2 \pm a_1^2)\, e^{j\beta_a}$$

$$a_m'(t) = \frac{1}{2}\sqrt{a_o^2 + a_1^2} + \frac{1}{2}\sqrt{a_o^2 - a_1^2}$$

$$b_m'(t) = \frac{1}{2}\sqrt{a_o^2 + a_1^2} - \frac{1}{2}\sqrt{a_o^2 - a_1^2}$$

$$\beta_a' = \beta_b' = \beta_a$$

This example can explain the functional principle of the Cocktail-Party-Processor-Algorithm a bit more in detail. Basis is the analysis of inter-modulations, which appear when mixing signals with different spectra. If signals interfere with different instantaneous frequencies of the analytic time signals, beats occur, which means, that changes in the signal envelope and phase (inter-modulations) occur. By a non-linear algorithm (the interaural cross product) terms are generated, which are dependent of the inter-modulations of both signals, and terms, which describe a linear overlay of the signals. By generating the complex mean value $\underline{\mu}$ the terms of the linear overlay can be extracted, with the help of the complex standard deviation $\underline{\sigma}$ the "inter-modulation-terms" can be isolated.

If only one source with a time variant amplitude is present, the sound signals are interpreted as an interference of two signals with different frequency and amplitude, but from the same direction.. For the algorithm there is no difference, whether the modulations of the envelope are caused by the interference of two signals from different directions (with different instantaneous frequencies) or by the interference of two frequency lines of one source signal. Also variations of the envelope are considered as the result of the interference of two source signals.

One Sound Source with arbitrary Amplitude Characteristics

An arbitrary amplitude characteristics can be described by the interference of a couple of frequency lines:

$$\underline{a}(t) = \sum_{i=1}^{N} a_i\, e^{\,j\Omega_i t + j\Phi_{io}}$$

The interaural cross product of this sound source results to::

$$\underline{k}(t) = e^{j\beta} \sum_{i=1}^{N} \sum_{k=1}^{N} a_i\, a_k\, e^{j(\Omega_i - \Omega_k)t + j\Phi_i - j\Phi_k}$$

$$\underline{k}(t) = e^{j\beta} \left(\sum_{i=1}^{N} a_i^2 + 2 \sum_{i=1}^{N} \sum_{k=1}^{i-1} a_i\, a_k \cosh\{ j(\Omega_i - \Omega_k)t + j\Phi_i - j\Phi_k \} \right)$$

For sufficiently big integration times the statistical parameters result to:

$$\underline{\mu} = e^{j\beta} \sum_{i=1}^{N} a_i^2$$

$$\underline{\sigma}^2 = 2\, e^{j\beta} \sum_{i=1}^{N} \sum_{k=1}^{i-1} a_i^2\, a_k^2 = \underline{\mu} - e^{j2\beta} \sum_{i=1}^{N} a_i^4$$

The absolute value of the mean value is proportional to the power of the source, the phase corresponds to the interaural phase. $\underline{\mu}$ correspond to the source vector of the source. The estimators for the source signal have all the same interaural phase β, but different amplitudes.:

$$\beta_a' = \beta_b' = \beta$$
$$a_m'(t) = \frac{1}{2}\sqrt{|\mu| + \sqrt{2}|\sigma|} + \frac{1}{2}\sqrt{|\mu| - \sqrt{2}|\sigma|}$$
$$b_m'(t) = \frac{1}{2}\sqrt{|\mu| + \sqrt{2}|\sigma|} - \frac{1}{2}\sqrt{|\mu| - \sqrt{2}|\sigma|}$$

If the source signal can be described inside the considered frequency range and time period by 2 frequency lines, the amplitude estimators yield into the amplitudes of these two frequency lines.

5.4. Complex Sound Fields

In this chapter the behavior of the Cocktail-Party-Processor-Algorithm at arbitrary sound fields shall be investigated, in order to analyze applicabilities and potentially necessary enhancements of the algorithm.

5.4.1. The interaural Cross Product at arbitrary Sound Sources

Arbitrary sound fields can be described by the interference of N different sound sources. Each sound source q, p can be described inside the considered frequency range and time period by M_q respectively M_p frequency lines, each frequency line i, k with its own instantaneous frequency Ω_i, Ω_k and instantaneous phase Φ_i, Φ_k (indices: p, i for the right, q, k for the left ear signal). The ear signals result to:

$$\underline{l}(t) = \sum_{q=1}^{N} e^{-j\frac{1}{2}\beta_q - \frac{1}{2}\alpha_q} \sum_{i=1}^{Mq} a_{mqi}\; e^{j\Omega_{qi}t + j\Phi_{qi}}$$

$$\underline{r}(t) = \sum_{p=1}^{N} e^{+j\frac{1}{2}\beta_p + \frac{1}{2}\alpha_p} \sum_{k=1}^{Mp} a_{mpk}\; e^{j\Omega_{pk}t + j\Phi_{pk}} \qquad (5.4.1/1)$$

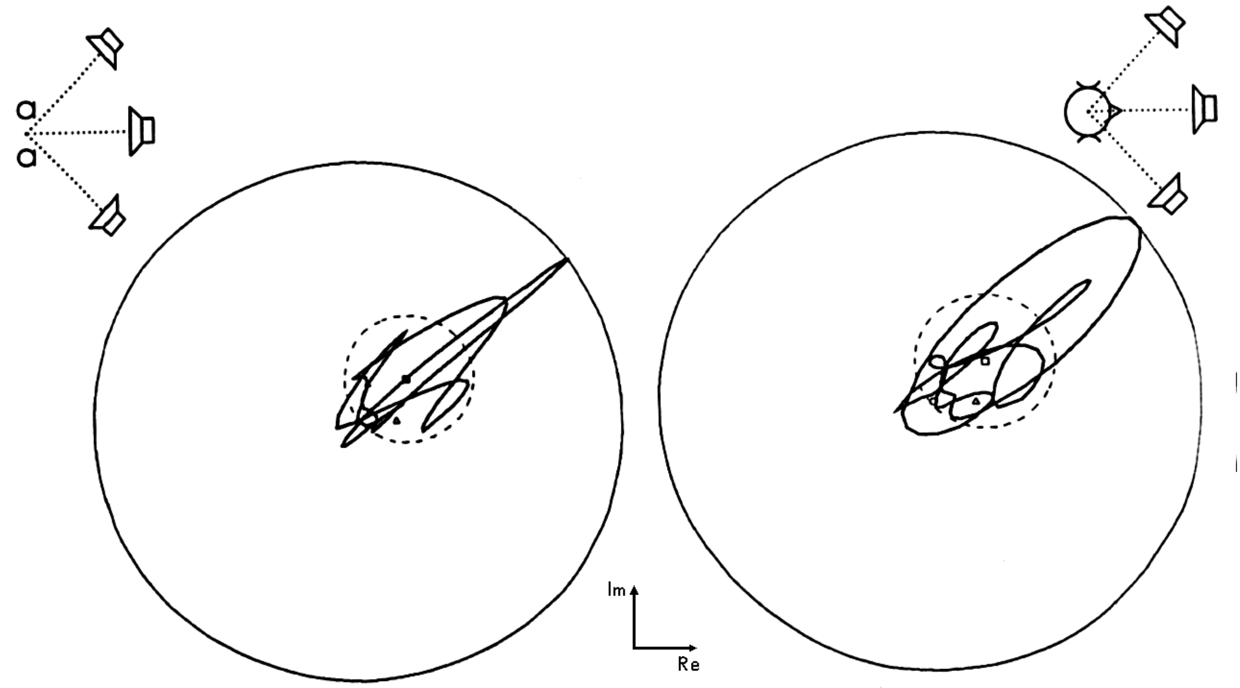


Fig. 5.8:
Locus curve of the interaural cross product for 3 sound sources with the same amplitude.
Signal A: Sinus 500 Hz, a=1, τ_a= 0µs;
Signal B: Sinus 540 Hz, b=1, τ_b=400µs;
Signal C: Sinus 580 Hz, c=1, τ_c=-400µs.
left figure: without interaural level differences, right figure: ΔL_b= 6dB, ΔL_c=-6dB

Then the interaural cross product results to::

$$\underline{k}(t) = \sum_{q=1}^{N}\sum_{p=1}^{N} e^{\,j\frac{1}{2}(\beta_p+\beta_q)+\frac{1}{2}(\alpha_p-\alpha_q)} \sum_{i=1}^{Mq}\sum_{k=1}^{Mp} a_{mqi}\, a_{mpk}\, e^{j(\Omega_{pk}-\Omega_{qi})t + j\Phi_{pk}-j\Phi_{qi}}$$

$$\underline{k}(t) = \sum_{q=1}^{N} e^{j\beta_q} \sum_{i=1}^{Mq} a_{mqi}^2$$
$$+\, 2 \sum_{q=1}^{N}\sum_{p=1}^{q-1} e^{j\frac{1}{2}(\beta_p+\beta_q)} \sum_{i=1}^{Mq}\sum_{k=1}^{Mp} a_{mqi}\, a_{mpk} \cosh\{\, j(\Omega_i-\Omega_k)t + j\Phi_{qi}-j\Phi_{pk} + \tfrac{1}{2}(\alpha_p-\alpha_q)\,\} \quad (5.4.1/2)$$

If the integration time $2T_\mu$ is big against the period of the difference frequencies $(\Omega_i-\Omega_g)/2\pi$, the statistical parameters result to:

$$\underline{\mu} = \sum_{q=1}^{N} e^{j\beta_q} \sum_{i=1}^{Mq} a_{mqi}^2$$

$$\underline{\sigma}^2 = 2 \sum_{q=1}^{N}\sum_{p=1}^{q-1} e^{j\beta_p+\beta_q} \sum_{i=1}^{Mq} a_{mqi}^2 \sum_{k=1}^{Mp} a_{mpk}^2 \qquad (5.4.1/3)$$

$\underline{\mu}$ and $\underline{\sigma}^2$ can be described by the statistical parameters of the cross product for the particular sources $\underline{\mu}_q$,$\underline{\sigma}_q$ (formula 5.3.4/1) (superposition theorem for the statistical parameters of the interaural cross product):

$$\underline{\mu} = \sum_{q=1}^{N} \underline{\mu}_q$$

$$\underline{\sigma}^2 = \sum_{q=1}^{N} \underline{\sigma}_q^2 + 2 \sum_{q=1}^{N} \sum_{p<q}^{q-1} \underline{\mu}_q \underline{\mu}_p = \underline{\mu}^2 + \sum_{q=1}^{N} \underline{\sigma}_q^2 - \underline{\mu}_q^2 \qquad (5.4.1/4)$$

The sum of the source vectors of the particular sources $\underline{\mu}_q = \underline{A}_{mq}^2$ determine the center of this locus curve. The shape of the locus curve is determined by the superposition of multiple ellipses, each corresponding to the locus curve from the combination of 2 of these sound sources.

If there are no interaural level differences (microphone recordings), these ellipses pass over to line segments, analogous to 2-sound source without interaural level differences. This line segments form a bundle of axes, which can span a part of the complex plane. The locus curves of the interaural cross product for 3 sound source are depicted in Fig. 5.8.

5.4.2. Two Sound Sources with time variant Amplitudes

The influence of time variant source signals on the locus curve of the interaural cross product depends on the change rate of amplitudes. If the change rate of the amplitudes is small against the difference between the signal frequencies, only the center and the size of the ellipse is modified, compared to the locus curve of two sources with constant amplitude (Fig. 5.2). The ratio between the main axes and the orientation of the ellipse remains unchanged (creation of a moving spiral). The estimators of the Cocktail-Party-Processor for such a case are depicted in Fig. 5.5.

If the signal amplitude changes faster than the "angular frequency" of the ellipse Ω_a-Ω_b, the shape of the locus curve changes; "loops" are formed (see Fig. 5.9). In this case the preconditions of formula 5.4.1/2 remain valid. The interaural cross product results to (sources a and b)::

$$\underline{k}(t) = e^{j\beta_a} \sum_{i=1}^{Ma} a_{mi}^2 + e^{j\beta_b} \sum_{k=1}^{Mb} b_{mk}^2$$
$$+ e^{j½(\beta_a+\beta_b)} \sum_{i=1}^{Ma} \sum_{k=1}^{Mb} a_{mi}\, b_{mk} \cosh\{ j(\Omega_i-\Omega_k)t + j\Phi_i - j\Phi_k + ½(\alpha_a-\alpha_b) \} \qquad (5.4.2/1)$$

The center of the locus curve results to the sum of the source vectors of the particular sources, like for two sound sources with constant amplitude. For each combination across the frequency lines of both sources an ellipse arises around this center point. The main axis ratio and the orientation of all these ellipses are equal. But size, angular frequency and zero phase are dependent on each particular combination of frequency lines.

If the statistical parameters $\underline{\mu}_a, \underline{\mu}_b, \underline{\sigma}_a, \underline{\sigma}_b$ of all sources are given, the statistical parameters of the cross product result according to formula 5.4.1/2 to:

$$\underline{\mu} = \underline{\mu}_a + \underline{\mu}_b$$

$$\underline{\sigma}^2 = \underline{\sigma}_a^2 + \underline{\sigma}_b^2 + 2\, \underline{\mu}_a \underline{\mu}_b$$

From this the following source estimators conclude:

$$(\underline{A}_m' \pm \underline{B}_m')^2 = \underline{\mu}_a + \underline{\mu}_b \pm \sqrt{4\, \underline{\mu}_a\, \underline{\mu}_b + \underline{\sigma}_a^2 + \underline{\sigma}_b^2}$$

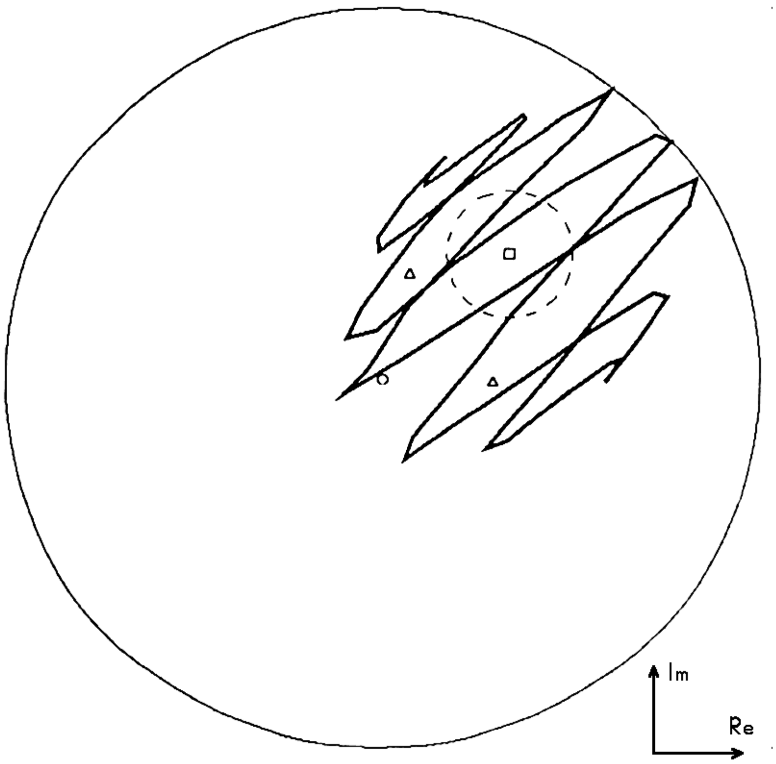


Fig. 5.9:
Locus curve
of the interaural cross product
for 2 modulated sound sources.

Signal A: Sinus 505Hz,
f_{mod}=10Hz; a=1; τ= 0µs

Signal B: Sinus 565 Hz,
f_{mod}=10Hz, b=1, τ=400µs

If the standard deviations of the particular signals are small against the geometric mean of all signal's power, this results into::

$$(\underline{A}_m' \pm \underline{B}_m')^2 \approx \left(\sqrt{\underline{\mu}_a}+\sqrt{\underline{\mu}_b}\right)^2 \pm \frac{\underline{\sigma}_a^2+\underline{\sigma}_b^2}{2\sqrt{\underline{\mu}_a\,\underline{\mu}_b}}$$

$$(\underline{A}_m' \pm \underline{B}_m')^2 \approx \left(\sqrt{\underline{\mu}_a}+\sqrt{\underline{\mu}_b}\right)^2 \pm \frac{\underline{\sigma}_a^2+\underline{\sigma}_b^2}{8\sqrt{\underline{\mu}_a\,\underline{\mu}_b}\;\left(\sqrt{\underline{\mu}_a}+\sqrt{\underline{\mu}_b}\right)}$$

$\underline{\mu}_a$ and $\underline{\mu}_b$ are the desired source vectors of both sound sources. As a consequence, the relative estimation error results into:

$$f_{Sch} = \frac{\underline{A}_m' - \sqrt{\underline{\mu}_a}}{\sqrt{\underline{\mu}_a}} = \frac{\underline{\sigma}_a^2+\underline{\sigma}_b^2}{8\,\underline{\mu}_a\,\underline{\mu}_b\left(\sqrt{\underline{\mu}_a/\underline{\mu}_b}\pm 1\right)} \quad (5.4.2/2)$$

The source estimators are composed of the demanded power estimators and of an error term, which depends on the standard deviations and therefore on the modulation depth of the signals. As a consequence, the directions of the estimators differ from the sound source directions. If the input directions of the sources are known, the difference between estimated and real direction can be a measure for the accuracy of the estimated amplitude values. (see also chapter 5.6).

5.4.3. More than two Sound Sources with constant Amplitudes

For sources with constant amplitudes the formula 5.4.1/2 becomes less complex ($M_q=M_p=1$):

$$\underline{k}(t) = \sum_{q=1}^{N} a_{mq}^2 \, e^{j\beta_q} + 2 \sum_{q=1}^{N} \sum_{p=1}^{q-1} a_{mq} \, a_{mp} \, e^{j½(\beta_p+\beta_q)} \cosh\{ j(\Omega_p-\Omega_q)t + j\Phi_q - j\Phi_p + ½(\alpha_p-\alpha_q) \} \qquad (5.4.3/1)$$

The sum of all source vectors $\underline{A}_{mi}^2$ corresponds to the center of this locus curve. The shape is caused by the superposition of ellipses, each representing the locus curve of a 2-source-configuration of two of the involved sound sources.

If there are no interaural level differences (microphone recordings) the ellipses pass over to line segments, analogous to 2-sound source without interaural level differences. This line segments form a bundle of axes, which span a part of the complex plane (formula 5.4.3/2, Fig. 5.8).

$$\underline{k}(t) = \sum_{q=1}^{N} a_{mq}^2 \, e^{j\beta_q} + 2 \sum_{q=1}^{N} \sum_{p=1}^{q-1} a_{mq} \, a_{mp} \, e^{j½(\beta_p+\beta_q)} \cos\{ (\Omega_p-\Omega_q)t + \Phi_q - \Phi_p \} \qquad (5.4.3/2)$$

For constant signal amplitudes and a sufficiently long integration time mean value and standard deviation of the interaural cross product result to:

$$\underline{\mu} = \sum_{q=1}^{N} \underline{A}_{mq}^2 \qquad (5.4.3/3)$$

$$\underline{\sigma}^2 = 2 \sum_{q=1}^{N} \sum_{p=1}^{q-1} \underline{A}_{mq}^2 \, \underline{A}_{mp}^2 \qquad (5.4.3/4)$$

5.4.4 Diffuse Sound Fields

Sound fields inside closed rooms can be described by mirror sound source models according to Fig. 4.4. The sound field characteristics of early reflections corresponds sooner to single sound sources, thus late reverberations can approximately be described as a diffuse sound field. In order to apply the Phase-Difference-Cocktail-Party-Processor to closed rooms (e.g. for de-reverberation purposes), it has to be investigated, how the algorithm reacts in a diffuse sound fields.

The diffuse Sound Field alone

An ideal diffuse sound field can be described by an unlimited number of sound sources with the same power density $E_m'^2=|A_{mi}'|^2=\text{const.}$, which are statistically distributed over all solid angles (see. Fig. 4.6: diffuse sound field within late reverberations). With $\underline{e}_{m\theta}(t)$ as the mean signals of the mirror sound sources the mean ear signals (reference point "center of the head") and the mean interaural cross product result to:

$$\underline{l}(t) = \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t) \; e^{-½\alpha_\theta \; -j½\beta_\theta} \, d\theta \qquad \underline{r}(t) = \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t) \; e^{+½\alpha_\theta \; +j½\beta_\theta} \, d\theta \qquad (5.4.4/1)$$

$$\underline{k}(t)= \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t)\, e^{+½\alpha_\theta + j½\beta_\theta}\, d\theta \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t)^*\, e^{-½\alpha_\theta + j½\beta_\theta}\, d\theta$$

If the mean power density is identical for all solid angles, then $|\underline{e}_{m\theta 1}(t)|\approx|\underline{e}_{m\theta 2}(t)|=E_m'$. If, additionally the mean spectrum of all mirror sound sources is identical, then also $\arg\{\underline{e}_{m\theta 1}(t)\}\approx\arg\{\underline{e}_{m\theta 2}(t)\}$. If, additionally the head is symmetrical, , then $\alpha(-\theta)=-\alpha(\theta)$; $\beta(-\theta)=-\beta(\theta)$. With all these conditions, which correspond to the description of late reverberations in symmetrical rooms by mirror sound sources, according to chapter 4.2, the interaural cross product results to:

$$\underline{k}(t)= \quad |E_m'|^2 \left| \int_{-\pi}^{\pi} e^{+½\alpha_\theta + j½\beta_\theta}\, d\theta \right|^2$$

$$\underline{k}(t)= 2\, |E_m'|^2 \left| \int_{0}^{\pi} e^{+½\alpha_\theta + j½\beta_\theta}\, d\theta \right|^2 \qquad (5.4.4/2)$$

In this case the phase and the standard deviation of the interaural cross product become to zero. The interaural cross product is no longer time dependent. The statistical parameters and the source estimators result to:

$$\underline{\mu} = |\underline{k}(t)| \qquad \underline{\sigma}^2 = 0$$

$$\underline{A}_m'^2 = |\underline{\mu}| \qquad \underline{B}_m'^2 = 0 \qquad (5.4.4/3)$$

The result is only one estimator for the front direction (median plane). This also corresponds to the considerations in chapter 4.2. The results of the model differ from the common hearing experience in reverberant rooms. But it has to be considered, that here only the diffuse part of the late reverberations has been examined and early reflections, which influence the hearing impression essentially, have not be examined here. Early reflections can be treated like a limited number of additional sound sources (see above).

One Sound Source in the diffuse Sound Field

If one sound source is located in a diffuse sound field, the free field ear signals of sound source $\underline{r}_q,\underline{l}_q$ and diffuse sound field $\underline{r}_d,\underline{l}_d$ interfere.

$$\underline{r}(t) = \underline{r}_d(t) + \underline{r}_q(t)$$

$$\underline{l}(t) = \underline{l}_d(t) + \underline{l}_q(t)$$

$$\underline{k}(t) = \underline{k}_d(t) + \underline{k}_q(t) + \underline{r}_d(t)\underline{l}_q(t)^* + \underline{r}_q(t)\underline{l}_d(t)^* \qquad (5.4.4/4)$$

If the periods of all difference frequencies are small against the integration time, the statistical parameters of the cross product can be evaluated as a function of the parameters of the particular sound source fields:

$$\underline{\mu} = \underline{\mu}_d + \underline{\mu}_q$$

$$\underline{\sigma}^2 = \underline{\sigma}_q^2 + \underline{\sigma}_d^2 + 2\, \underline{\mu}_d\, \underline{\mu}_q = \underline{\sigma}_q^2 + 2\, \underline{\mu}_d\, \underline{\mu}_q \qquad (5.4.4/5)$$

If the amplitude of the sound source is constant, the following statistical parameters and source estimators result from it::

$$\underline{\mu} = \underline{\mu}_d + \underline{A}_q^2$$

$$\underline{\sigma}^2 = 2\, \underline{\mu}_d\, \underline{A}_q^2$$

$$4\,(\underline{A}_m' \pm \underline{B}_m')^2 = \underline{\mu}_d + \underline{A}_q^{\,2} \pm \sqrt{2\,\underline{\mu}_d}\;\underline{A}_q$$

As a result there are 2 source estimators, one for the sound source and one for the diffuse field.

$$\underline{A}_m'^2 \approx \underline{A}_q^2 \qquad\qquad \underline{B}_m'^2 \approx \underline{\mu}_d \qquad (5.4.4/6)$$

If the sound source is located in the median plane, two estimators with the same direction are obtained. The power of the diffuse field and of the sound source add up. This remembers to room acoustical effects, where a systematic construction of rooms with a reverberation time smaller than 0,8 s can improve loudness and speech intelligibility. However, sound and envelope of the sound source signals change by adding diffuse field parameters.

For lateral sound sources the diffuse field does not influence the estimator of the sound source. Diffuse field and sound source can be directionally separated, which can lead to a certain de-reverberation in reverberant rooms. When turning towards the source, the addition of power appears again and the source estimators are influenced by the diffuse field again.

Two Sound Sources in the diffuse Sound Field

For two sound sources in the diffuse sound field the statistical parameters of the interaural cross product ($\underline{\sigma}_q^2=2\underline{A}_q^2\underline{B}_q^2$ and $\underline{\mu}_q=\underline{A}_q^2+\underline{B}_q^2$) result, according to formula 5.4.1/3, to:

$$\underline{\mu} = \underline{\mu}_d + \underline{A}_q^2 + \underline{B}_q^2$$

$$\underline{\sigma}^2 = 2\,\underline{A}_q^2\,\underline{B}_q^2 + 2\,\underline{\mu}_d\,(\underline{A}_q^2 + \underline{B}_q^2)$$

Then the source estimators apply from:

$$(\underline{A}_m' \pm \underline{B}_m')^2 = \underline{A}_q^{\,2} + \underline{B}_q^{\,2} + \underline{\mu}_d \pm 2\sqrt{\underline{A}_q^{\,2}\underline{B}_q^{\,2} + \underline{\mu}_d\,\underline{A}_q^{\,2} + \underline{\mu}_d\,\underline{B}_q^{\,2}}$$

If the sources are outside the reverberation radius, the diffuse field prevails and one source estimator corresponds to the superposition of both sound sources, similar to sum localization, while the other estimator corresponds to the diffuse field.

$$(\underline{A}_m' \pm \underline{B}_m')^2 \approx \underline{A}_q^{\,2} + \underline{B}_q^{\,2} + \underline{\mu}_d \pm 2\sqrt{\underline{\mu}_d\,(\underline{A}_q^{\,2} + \underline{B}_q^{\,2})}$$

$$\underline{A}_m'^2 \approx \underline{A}_q^2 + \underline{B}_q^2$$

$$\underline{B}_m'^2 \approx \underline{\mu}_d$$

If both sound sources are inside the reverberation radius, the source estimators of both sources are influenced by the diffuse field::

$$4\,(\underline{A}_m' \pm \underline{B}_m')^2 \approx \underline{A}_q^{\,2} + \frac{1}{2}\underline{\mu}_d + \underline{B}_q^{\,2} + \frac{1}{2}\underline{\mu}_d \pm 2\sqrt{(\underline{A}_q^{\,2} + \frac{1}{2}\underline{\mu}_d)(\underline{B}_q^{\,2} + \frac{1}{2}\underline{\mu}_d)}$$

$$\underline{A}_m'^2 \approx \underline{A}_q^2 + \tfrac{1}{2}\underline{\mu}_d$$

$$\underline{B}_m'^2 \approx \underline{B}_q^2 + \tfrac{1}{2}\underline{\mu}_d$$

The diffuse field leads to a displacement in the directional estimation of the sound sources. However, if the head is oriented towards a sound source, no directional displacement appears, but the power of the corresponding estimator, will be the sum of diffuse field and sound source power.

5.5. Dominant Sources

For a bigger number of sound sources and frequency lines, each with the complex mean value $\underline{\mu}_q$ and the standard deviation $\underline{\sigma}_q$, mean value and standard deviation of the cross product result, according to formula 5.4.1/4, to:

$$\underline{\mu} = \sum_{q=1}^{N} \underline{\mu}_q \qquad \underline{\sigma}^2 = \sum_{q=1}^{N} \underline{\sigma}_q^2 + 2 \sum_{q=1}^{N} \sum_{p<q}^{q-1} \underline{\mu}_q \underline{\mu}_p$$

One dominant Source

If one source is dominant, the power of this source $\underline{A}_m^2$ prevails the power of all other sources $|\underline{A}_m^2| > \Sigma|\underline{\mu}_i|$, $i \neq q$. Representing all sources' frequency lines by particular sources with $\underline{\sigma}_a=0$ results to:

$$\underline{\mu} = \underline{A}_m^2 + \sum_{q\neq a} \underline{\mu}_q \qquad \underline{\sigma}^2 = 2\,\underline{A}_m^2 \sum_{q\neq a} \underline{\mu}_q + 2 \sum_{q\neq a} \sum_{p<q} \underline{\mu}_q \underline{\mu}_p$$

One source estimator corresponds to the dominant source and the second source estimator corresponds, analogous to sum localization, to the superposition of the other sources.

$$\underline{A}_m'^2 \approx \underline{A}_m^2 \qquad \underline{B}_m'^2 \approx \sum_{q\neq a} \underline{\mu}_q$$

Two dominant Sources

If the power of 2 Sources prevails the power of the rest of the sound field, the statistical parameters of the cross product result to:

$$\underline{\mu} = \underline{A}_m^2 + \underline{B}_m^2 + \sum_{q\neq a,b} \underline{\mu}_q$$

$$\tfrac{1}{2}\underline{\sigma}^2 = \underline{A}_m^2 \underline{B}_m^2 + (\underline{A}_m^2 + \underline{B}_m^2) \sum_{q\neq a,b} \underline{\mu}_q^2 + \sum_{q\neq a,b} \sum_{p<q} \underline{\mu}_q \underline{\mu}_p$$

The source estimators correspond to the source vectors of the dominant sources, which are impacted by the remaining sound field. Depending on the power of the remaining sound field the source estimators deviate more or less from the source vectors of the dominant sources.:

$$\underline{A}_m'^2 \approx \underline{A}_m^2 + \sum_{q\neq a,b} \underline{\mu}_q \qquad \underline{B}_m'^2 \approx \underline{B}_m^2 + \sum_{q\neq a,b} \underline{\mu}_q$$

Conclusion

The algorithm is also able to estimate power and direction of a sound source in complex sound fields, provided that the desired sound source is disturbed by maximal one strong source or the desired source has a bigger power than all other sources together.

If multiple interfering sources are present and the signal-to-noise-ratio is negative, the estimated directions and power deviate from the existing sound sources. Therefore a correction of the estimators becomes necessary. Such correction algorithms have to realize possible reasons for estimator-deviations and construct a corrected estimator of the desired direction out of it.

5.6. Mapping of Source Estimators onto a desired Direction

Error Reasons

The presented algorithm is able to interpret the mean value and the standard deviation of the interaural cross product as the result of a sound situation with 2 sound sources and to determine power and input directions of these sound sources.

For sound situations with more than two sound sources or modulated sound signals the parameters of all sources can not be determined, the estimators differ prom the existing sound source parameters. In order to use this method as a Cocktail-Party-Processor at least the power, which belongs to a certain input direction, has to be evaluated. If the estimated direction deviates from the desired direction, the question appears, whether signals from the desired direction are present at all and how their portion can be determined.

Reason for such deviations can be the presence of multiple sound sources.

- More than 2 sound sources can be present. In this case the source estimators of several existing sources are combined to pseudo-sources. These estimators satisfy the equation system, but they must not correspond to the power and directions of existing sources.
- Source signals can be modulated. This corresponds to the presentation of several sound sources from the same input direction, respectively.
- Reflections and reverberation can be appear. This corresponds to the presence of a big number of mirror sound sources.

Deviations between the estimated input direction and existing directions can also be caused by inaccuracies in applying this method:

- Ambiguities of the interaural phases at high frequencies. Different input directions might be mapped onto the same interaural phase.
- Too inaccurate specification of the desired direction
- Computing inaccuracies.
- Too long integration time compared to the characteristics of the signal parameters. The signal parameter change during the integration time.
- Too short integration time compared to the difference frequencies of the signals. Terms, which depend on the difference of the instantaneous frequencies, are not averaged out and influence as non-considered terms the result of the equation system..
- Different conversions interaural Phase - interaural time difference - input angle for different frequencies inside a critical band. Since this conversion depends on the instantaneous frequencies of the signals, different interaural phases can appear for the same input direction, but different frequencies. This variance of the interaural phases is dependent on the relative bandwidth of the critical band.

Subsequently the influence of these error reasons shall be investigated. In case of deviations between estimated and desired direction, it shall be estimated, with which probability signals of the desired direction are present and how big their power might be. When an estimator $\underline{X}_m'^2$ (=$\underline{A}_m'^2$ or $\underline{B}_m'^2$) has been computed, the signal portions inside this estimator, which might originate from the desired direction, are estimated with the help of a weighting function $W_x(\beta_{soll})$. Then the estimated power from the desired direction $s_m'^2$ results to (see also <ins>Fig. 5.10</ins>):

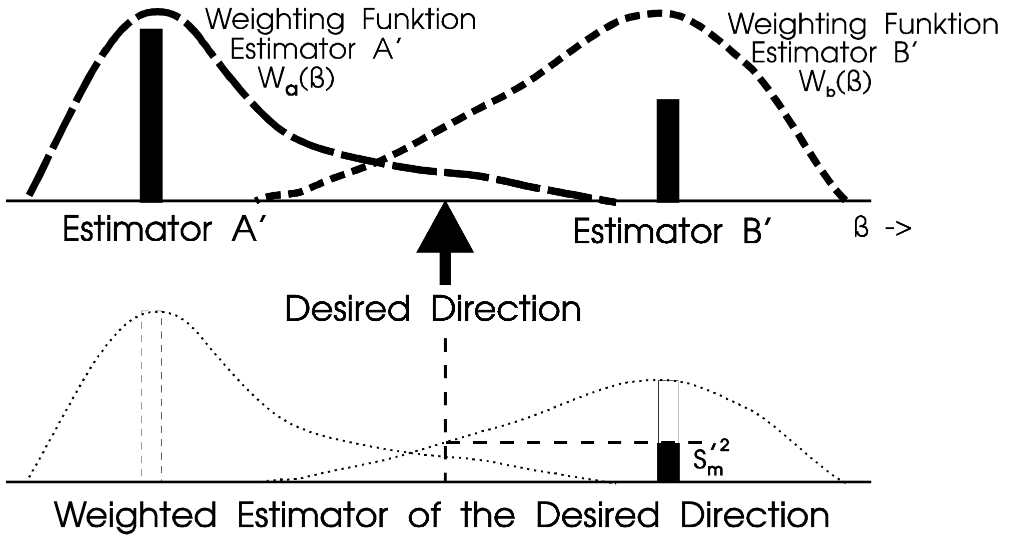


Fig. 5.10: Mapping of source estimators onto a desired direction with the help of weighting functions

$$s_m'^2 = W_x(\beta_{soll})\ x_m'^2 \qquad \text{with } x_m'^2 = a_m'^2 \text{ or. } b_m'^2 \qquad (5.6/1)$$

A weighting function is also used by Gaik [17] and Bodden [10] for appraising the patterns of an interaural cross correlation function in case of deviations between an estimated direction and a desired direction.

Dependencies between interaural Time Difference and interaural Phase

Inside a critical band with the cut-off-frequencies f_u, f_o and the center frequency f_m the interaural phases spread for a constant interaural time difference τ over a dispersion range of:

$$\beta(f_o)-\beta(f_u) = \tau\, 2\pi(f_o-f_u) = \beta(f_m)\ (f_o-f_u)/f_m$$

The magnitude of the deviations corresponds of the relative bandwidth of the critical band filter (about. ±10% at third octave wide critical bands). The effect of this error grows proportionally with the displacement from the median plane. For frontal sound (τ=0) this error does not appear.

Sources with interaural phases β_{soll} inside this dispersion range have to be interpreted as correctly estimated (weighting factor=1). Therefore a corresponding weighting function $W_x(\beta_{soll})$: can be constructed like (β_x'=interaural phase of one source estimator):

$$W_x(\beta_{soll}) = 1 \qquad \text{for } |\beta_x'-\beta_{soll}| < \frac{f_o-f_u}{2\,f_m}\,|\beta_x'| \qquad (5.6/2)$$

Valid Estimator Range

The sum of the estimated source vectors must be equal to the complex mean value of the interaural cross product:

$$\underline{A}_m'^2 + \underline{B}_m'^2 = \underline{\mu}$$

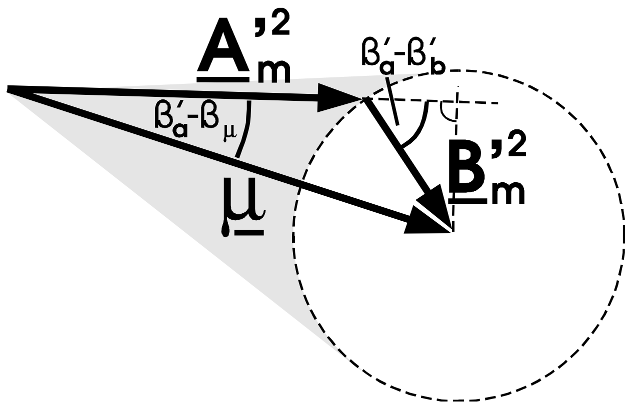


Fig. 5.11:

Relationship between source estimators and the mean value of the interaural cross product

gray background: possible interaural phase range for the estimator $\underline{A}_m'^2$

$\underline{A}_m'^2, \underline{B}_m'^2, \underline{\mu}$ build a triangle in the complex plane (Fig. 5.11). The power of the weakest estimator $|B_m'^2|$ limits the possible interaural phase difference $|\beta_a'-\beta_\mu|$ between the estimator with the biggest power $|A_m'^2|$ and the complex mean value $\underline{\mu}$:

$$|\beta_a'-\beta_\mu| \approx \arctan\left(|B_m'^2| \sin(|\beta_b'-\beta_a'|) \;/\; |A_m'^2| \right)$$

This phase difference is maximal, if there is a right angle between $\underline{A}_m'^2$ and $\underline{B}_m'^2$ in the corresponding locus curve representation. The possible interaural phase range is not limited for the weaker source estimator or for estimators with equal amplitudes.

$$|\beta_a'-\beta_\mu| \leq \arctan\left(|B_m'^2| / |A_m'^2| \right) \qquad \text{for} \quad |A_m'^2| > |B_m'^2|$$

This means, that the interaural phases β of all sound sources, which are represented by estimator $\underline{A}_m'^2$, must be located inside this interaural phase range. This limitation is not valid for estimator $\underline{B}_m'^2$.

$$|\beta-\beta_a'| \leq |\beta_a'-\beta_\mu| \qquad\qquad |\beta-\beta_b'| \leq \pi/2$$

The quotient of the statistical parameters of the cross product results to:

$$\frac{\sigma}{\sqrt{2}\underline{\mu}} = \frac{A_m' B_m'}{\underline{A}_m'^2 + \underline{B}_m'^2} = \frac{B_m'/A_m'}{1+\underline{B}_m'^2/\underline{A}_m'^2} \approx \frac{B_m'}{\underline{A}_m'} \qquad \text{for} \quad |A_m'^2| \gg |B_m'^2|$$

$$\approx 1 / 2\cos(\beta_\mu-\beta_a) \qquad \text{for} \quad |A_m'^2| \approx |B_m'^2|$$

The maximal allowed interaurale phase range around the estimator-phases can be described with the help of the statistical parameters of the interaural cross product for $|A_m'^2| \gg |B_m'^2|$:

$$|\beta-\beta_a'| \leq |\beta_a'-\beta_\mu| \leq \arctan\left(\tfrac{1}{2}|\sigma/\mu|^2 \right) \qquad \text{for} \quad |A_m'^2| > |B_m'^2|$$

$$|\beta-\beta_b'| \leq \pi/2 \qquad = 2\arctan\left(\tfrac{1}{2}|\underline{\sigma}/\underline{\mu}|^2 \; |\underline{A}_m^{2'}/\underline{B}_m^{2'}|^2 \right)$$

Out of this the upper limit for possible deviations of the source parameters from the interaural phases of the estimators can be determined as:

$$|\beta-\beta_a'| \leq \arctan\left(\tfrac{1}{2}|\underline{\sigma}/\underline{\mu}|^2 \max(|\underline{A}_m'^2|,|\underline{B}_m'^2|) \;/\; |\underline{A}_m'^2| \right) \qquad (5.6/3)$$

$$|\beta-\beta_b'| \leq 2 \arctan\left(\tfrac{1}{2}|\underline{\sigma}/\underline{\mu}|^2 \max(|\underline{A}_m'^2|,|\underline{B}_m'^2|) \;/\; |\underline{B}_m'^2| \right)$$

Therefore it is unlikely, that signal portions exist from directions outside these ranges around the estimated phases. Weighting will be done by a cosine window, which weights estimators with interaural phases according formula 5.6/3 with the probability 0.5 and reduces the weight for bigger interaural phase deviations. ($\beta_x', X_m'^2$= interaural phase and power of the weighted estimator):

$$W_x(\beta_{soll}) = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{\beta_{soll}-\beta_x'}{\beta_{max}-\beta_x'}\,\frac{\pi}{2}\right) \; ; \qquad |\beta_{max}-\beta_x'| = \arctan\left(\frac{|\underline{\sigma}|^2}{2|\underline{\mu}|^2}\,\frac{\max(|\underline{A}_m'^2|,|\underline{B}_m'^2|)}{|\underline{X}_m'^2|}\right)$$

$$\text{for } |\beta_{soll}-\beta_x'| \le 2\,|\beta_{max}-\beta_x'|$$

$$W_x(\beta_{soll}) = 0 \qquad \text{otherwise} \qquad (5.6/4)$$

Combination of Source Vectors

Both the sum of the source estimators $\underline{A}_m'^2, \underline{B}_m'^2$ (formula 5.3.2/2) and the sum of all source vectors of the sound sources $\underline{A}_{mq}^2$ (formula 5.4.3/2) have to result into the mean value of the interaural cross product.

$$\underline{\mu} = \underline{A}_m'^2 + \underline{B}_m'^2 \qquad\qquad \underline{\mu} = \sum_q \underline{A}_{mq}^2$$

If more than 2 sound sources are present, an assignment between the source vectors of the sound sources and the source estimators must exist. One (arbitrary) possibility would be, to assign all source vectors with interaural phases bigger than the interaural phase of the mean value $\underline{\mu}$, to the source estimator with also bigger interaural phase and to assign all source vectors with smaller interaural phases to the other source estimator.

$$\underline{A}_m'^2 = \sum_q \underline{A}_{mq}^2 \qquad \text{with} \quad \mathrm{sign}(\beta_\mu - \beta_q) \ge 0$$

$$\underline{B}_m'^2 = \sum_p \underline{A}_{mp}^2 \qquad \text{with} \quad \mathrm{sign}(\beta_\mu - \beta_p) \le 0$$

If the desired direction does not match to the direction of the corresponding source estimator $\underline{X}_m'^2$, then a signal of the desired direction $\underline{S}_m'^2$ can only be present, if there is at least one source vector of another signal $\underline{G}_m'^2$, for which the sum with the source vector of the desired direction results into a source estimator. An estimation of the maximal possible power form a desired interaural phase β_{soll} can be achieved, if the interaural phase difference between the desired direction and the direction of the claimed signal $\underline{G}_m'^2$ gets maximal. If the desired direction $\beta_{soll,}$ lies outside the phase range between source estimator and mean value, the maximal possible power appears, if the interaural phase of the claimed signal $\underline{G}_m'^2$ corresponds to the mean value $\underline{\mu}$.. If β_{soll} lies inside the phase range between source estimator and mean value, then the maximal power appears, if the direction of the claimed signal forms in the complex plane a right angle to the direction of the desired direction.

$$\underline{X}_m'^2 \overset{!}{=} \underline{S}_m'^2 + \underline{G}_m'(\beta_\mu)^2 \qquad \text{for} \quad \mathrm{sign}(\beta_{soll}-\beta_x') = \mathrm{sign}(\beta_\mu-\beta_x')$$

$$\underline{X}_m'^2 \overset{!}{=} \underline{S}_m'^2 + \underline{G}_m'(\beta_x'-\pi/2)^2 \qquad \text{for} \quad \mathrm{sign}(\beta_{soll}-\beta_x') = -\mathrm{sign}(\beta_\mu-\beta_x')$$

$\underline{X}_m'^2$, $\underline{S}_m'^2$ and $\underline{G}_m'^2$ form an triangle in the complex plane. With the help of the sinus theorem the power ratio between desired and estimated direction $|\underline{S}_m'^2/\underline{X}_m'^2|$ can be determined. With the same direction selected for $\underline{G}_m'^2$ as above this corresponds to the maximal possible power ratio. From this a weighting function $W_x(\beta)$ can be constructed, to correct those estimators, where the interaural phases differ from the desired direction.

$$W_x(\beta_{soll}) \le \frac{|\underline{S}_m'^2|}{|\underline{X}_m'^2|} = \frac{\sin(\beta_g-\beta_x')}{\sin(\beta_g-\beta_{soll})} \qquad \beta_g = \beta_\mu \text{ or } \beta_x' \pm \pi/2 \qquad (5.6/5)$$

Mapping of a 3-Source-Configuration to a desired Direction

A 3 sound source configuration is the simplest configuration, where the interaural phases of the source estimators deviate from the interaural phases of existing sound sources. Deviations of the source estimators from the desired direction can be interpreted as a result of the interfering of 3 sound sources: the source of the desired direction (source vector $\underline{S}_m{}^2 = s_m{}^2 e^{j\beta_{soll}}$) and two sources from other directions ($\underline{A}_m{}^2 = a_m{}^2 e^{j\beta_a}$ and $\underline{B}_m{}^2 = b_m{}^2 e^{j\beta_b}$). From the comparison between found deviations and a 3-source-configuration correction methods for these derivations can be developed.

Mapping the source estimators onto 3 input directions will give an estimate for the possible power of the desired direction. For more than 3 sources or only 2 sources the estimated power of the desired direction would be smaller. For 3 sources the statistical parameters of the cross product result to:

$$\underline{\mu} = \underline{A}_m{}'^2 + \underline{B}_m{}'^2 = \underline{A}_m{}^2 + \underline{B}_m{}^2 + \underline{S}_m{}^2$$

$$\tfrac{1}{2}\,\underline{\sigma}^2 = \underline{A}_m{}'^2\,\underline{B}_m{}'^2 = \underline{A}_m{}^2\underline{B}_m{}^2 + \underline{A}_m{}^2\underline{S}_m{}^2 + \underline{B}_m{}^2\underline{S}_m{}^2$$

Each source estimator $\underline{A}_m{}'^2, \underline{B}_m{}'^2$ shall now be described by the interference of a source vector $\underline{A}_m{}^2, \underline{B}_m{}^2$ with the signal of the desired direction $\underline{S}_m{}^2$ (g is a real factor, which is not yet known):

$$\underline{A}_m{}'^2 = \underline{A}_m{}^2 + g\,\underline{S}_m{}^2 \qquad \underline{B}_m{}'^2 = \underline{B}_m{}^2 + (1-g)\,\underline{S}_m{}^2$$

$$\tfrac{1}{2}\underline{\sigma}^2 = \underline{A}_m{}'^2\,\underline{B}_m{}'^2 = (\underline{A}_m{}'^2 - g\,\underline{S}_m{}^2)\,(\underline{B}_m{}'^2 - (1-g)\,\underline{S}_m{}^2) + (\underline{A}_m{}'^2 + \underline{B}_m{}'^2 - \underline{S}_m{}^2)\,\underline{S}_m{}^2$$

$$0 = \underline{S}_m{}^2 \left(g\,\underline{A}_m{}'^2 + (1-g)\,\underline{B}_m{}'^2 - (g^2-g+1)\,\underline{S}_m{}^2\right)$$

$$s_m{}^2(g^2-g+1) = g\,a_m{}'^2 e^{\,j\beta_a' - j\beta_{soll}} + (1-g)\,b_m{}'^2 e^{\,j\beta_b' - j\beta_{soll}}$$

From this the unknown items g and $s_m{}^2$ can be determined. From the imaginary part follows:

$$g = \frac{b_m'^2 \sin(\beta_b' - \beta_{soll})}{b_m'^2 \sin(\beta_b' - \beta_{soll}) - a_m'^2 \sin(\beta_a' - \beta_{soll})} \qquad (5.6/6)$$

$$1-g = \frac{a_m'^2 \sin(\beta_a' - \beta_{soll})}{a_m'^2 \sin(\beta_a' - \beta_{soll}) - b_m'^2 \sin(\beta_b' - \beta_{soll})}$$

From the real part the possible power of a signal from the desired direction can be evaluated:

$$s_m^2 = \frac{a_m'^2 b_m'^2 \sin(\beta_b' - \beta_a')}{g\,b_m'^2 \sin(\beta_b' - \beta_{soll}) - a_m'^2 \sin(\beta_a' - \beta_{soll})} \qquad (5.6/7)$$

From this the weighting functions result as:

$$W_a(\beta_{soll}) = \frac{s_m'^2}{a_m'^2} = \frac{\sin(\beta_a' - \beta_b')}{(1-g)\,a_m'^2/b_m'^2 \sin(\beta_a' - \beta_{soll}) - \sin(\beta_b' - \beta_{soll})}$$

$$W_b(\beta_{soll}) = \frac{s_m'^2}{b_m'^2} = \frac{\sin(\beta_b' - \beta_a')}{g\,b_m'^2/a_m'^2 \sin(\beta_b' - \beta_{soll}) - \sin(\beta_a' - \beta_{soll})}$$

$s_m{}^2$ is a measure for the possible power of a sound signal of the desired direction.

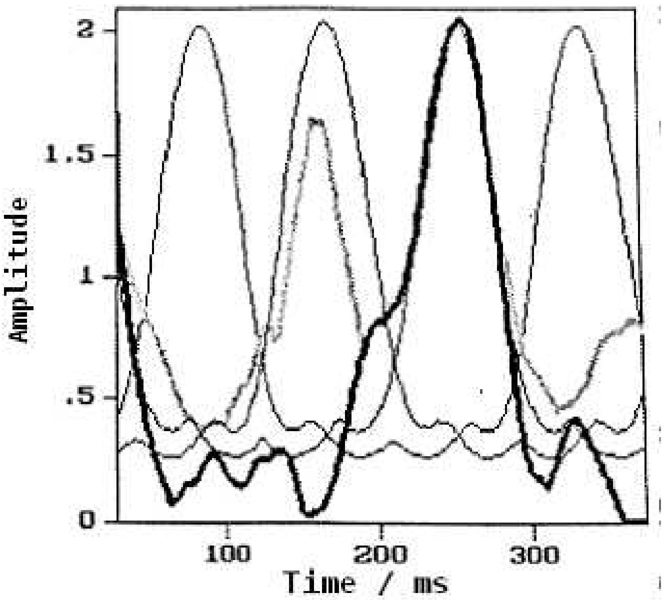


Fig 5.12:
Effect of correction methods at 3 sound sources

Signals: tone complex, 16 Hz bandwidth,
5 frequency lines.
Desired signal: 500-516Hz, τ= 400µs
Interfering signal1: 540-565Hz, τ= 0µs
Interfering signal 2: 570-585Hz, τ=-400µs
Signal-to-noise-ratio : -3 dB

signal envelopes
estimator of the desired signal without correction method
estimator of the desired signal without correction method "valid estimator range"

Signal Processing Examples

Using the example of the correction method "Valid Estimator Range" the effect of correction methods shall be demonstrated. Fig. 5.13 shows the estimators of a desired source with and without applying a correction method for amplitude modulated signals with mean signal-to-noise-ratios of -20 dB and -30 dB (corresponding to Fig. 5.5).

Without a correction bigger estimation errors of 10 dB appear for a mean signal-to-noise-ratio of

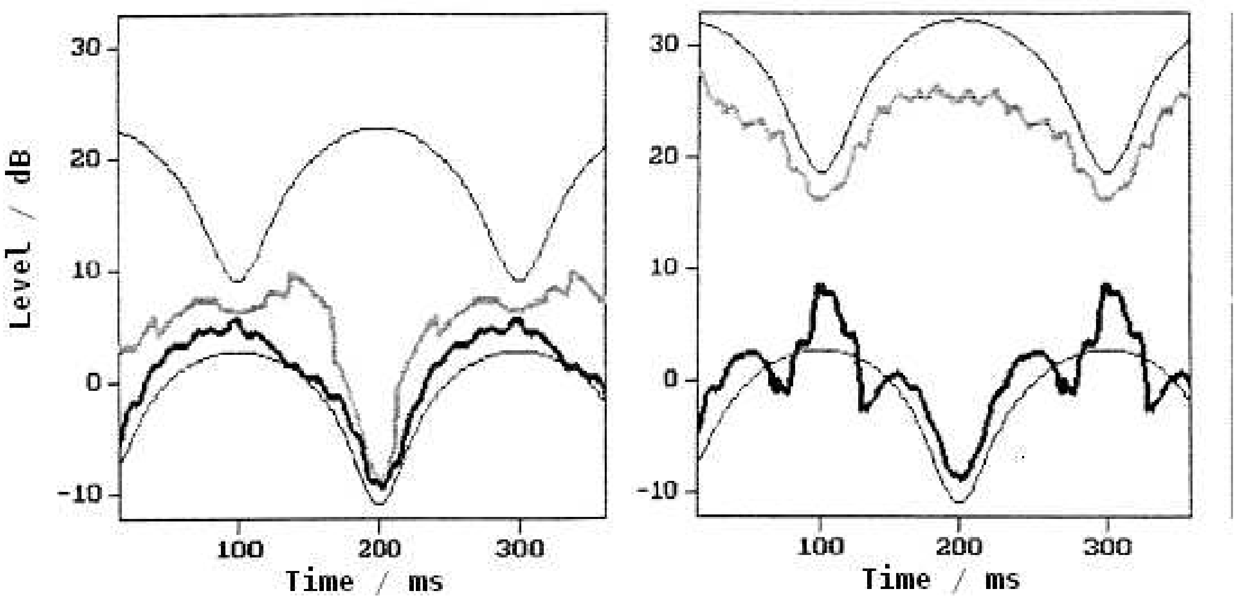


Fig. 5.13: The correction method at modulated sound signals.
Model parameter: T_μ=28 ms, 1 estimator/ms, estimator-averaging 30 ms, lock-in-range 0.1π
Correction method: "valid estimator range"
desired signal: Sinus 560Hz + Sinus 565Hz, amplitude ratio 1:0.8, τ=400µs
interfering signal: Sinus 500Hz + Sinus 495Hz, amplitude ratio 1:0.8, τ=0µs
envelopes of desired and interfering signal
estimated envelope of the desired direction without correction method
estimated envelope of the desired direction with correction method
Mean signal-to-noise-ratio : left figure: -20 dB right figure: -30 dB

-20 dB. For a signal-to-noise-ratio of -30 dB the estimation of the desired signal is no longer possible. The estimated envelopes correspond nearly to the interfering signal.

By applying the correction method "Valid Estimator Range" the estimation error can be reduced to circa 3 dB for signal-to-noise-ratios of -20 dB. For signal-to-noise-ratios of -30 dB at least a rough estimation of the envelope of the desired direction becomes possible (error ≤7 dB).

For 3 active sound sources (Fig. 5.12) the signal of the desired direction can no longer be estimated without a correction of the estimators, if the interfering signals prevail. In this case the estimated amplitudes correspond nearly to those of the interfering signals. The correction method "Valid Estimator Range" detects in those times, where the uncorrected estimator follows the interfering signals, these estimation errors and corrects the estimated values correspondingly. By this the estimation errors are reduced considerably for cases with prevailing interfering sound sources.

5.7. Conclusion

Now a signal processing method is available, which allows to process the signals of two sound sources in the free sound field direction specifically. Applied to ear signals this method is ideally suited for the processing of low frequency signals, because there are unambiguous relationships between input directions and interaural phase differences.

The algorithm interprets fluctuations of the signal envelopes as the result of the interference of two sound sources and evaluates source estimators, which describe the interfering components. The source estimators describe the effect of the sound field on the interaural envelope, they analyze it in terms of particular frequency lines or as a group of frequency lines, independently from their spatial distribution. Source estimators can also be dominant frequency lines of one sound source or the sum of several sound sources (e.g. all mirror sound sources of a diffuse sound field). For one sound source with a fluctuating envelope 2 source estimators with the same interaural phase will appear.

More than 2 sources (or frequency lines) are mapped to 2 source estimators.. If two sources (or frequency lines) are dominant, they dominate the source estimators, too, then other sources have only an effect as error terms for these estimators. If the input directions are know, then errors in the directional estimation can be a sign for the existence of further sound sources. If no or more than two dominant sources exist, then the estimation errors increase, then the source estimators describe no longer existing sound sources, but the characteristics of the sound field.

Estimation errors can provide additional information about involved sound sources. If the input direction of a desired source is known, the signal power of this direction can be estimated from the estimator power and the deviation of the estimated direction from the desired direction. Even in complex sound fields the possible course of the power of a desired sound source can be extrapolated by this, and the signal-to-noise-ratio can be enhanced. In very complex sound fields, for example multiple sound sources in reverberant environment with many early reflections, the capabilities of this method in enhancing the signal-to-noise-ratio are reduced. But in such situations also the auditory system is no longer able to analyze the sound field correctly, as the Franssen-Effect [16] shows.

For very complex sound fields additional analysis procedures for the interaural cross product might be necessary besides the analysis of mean value and standard deviation (moments of higher order, pattern-recognition-methods, main axis analysis etc.). In this manner additional conclusions about the sound field and involved sound sources might be possible (for example about the count of sound sources, diffusivity of the sound field, interaural level differences, times of dominance for specific sound sources).

One problem is, however, the ambiguity of the interaural phase of the cross product. For natural ear distances and frequencies above ca. 800 Hz the signal period becomes smaller than the maximal interaural time difference. Different input direction can than get the same interaural phase and are mapped on each other.

> Example: The input directions 0° and 40° lead to normalized interaural time differences of 0 and 400 µs (formula 3.1/1). For frequencies of 5 kHz, 10 kHz and 15 kHz these are mapped onto the same interaural phase, for 2.5 kHz, 7.5 kHz, 12.5 kHz and 17.5 kHz they are mapped onto by π displaced interaural phases, with a similar effect.

Under these conditions it is for high frequencies nearly impossible, to infer from the interaural phases onto the input direction. (At 16 kHz the signal period corresponds to only 10% of the maximal interaural time difference, this corresponds to an input direction difference of 6° for the front direction). But with increasing frequency the directional selectivity of this method increases, too. Therefore different input directions can still be detected via different interaural phase leftovers to multiples of 2π, and signal processing remains possible, even though in a reduced form.

> In the example above for 500 Hz-signals interaurale phases appear of 0 and 0.4π, for 16 kHz of 0 and 6.4π. After adjusting the phases by multiples of 2π the phases become identical. In both cases the algorithm would be able to separate the sources.

A further problem is the frequency dependency of the interaural phases for lateral sound. For third-octave-filters the interaural phases can spread by about ±10% for the same input direction. For an input direction of 40° there is a range of dispersion around 500 Hz of about $\pm 0.04\pi$, but $\pm 1.28\pi$ inside the critical band around 16 kHz. An analysis if the interaural phase would be impossible here. Since this error does not appear at a interaural time difference of 0 µs, this error can be eliminated by projecting the desired direction onto the front-direction, for example by shifting the ear signals by the corresponding interaural time difference.

It remains a problem, that for high frequency the exact estimation of the desired direction would be necessary. But nevertheless the free field outer ear transfer functions provide steep minimums and maximums for high frequencies, which lead to additional phase shifts and additional uncertainties. An exact transformation input direction - interaural time difference - interaural phase becomes hardly possible.

In order to solve this problem additional information sources have to be used, like the auditory system does, as there are interaural group delays or interaural level differences. The following chapter ("Level-Difference-Cocktail-Party-Processor") described methods, to determine input directions and power of involved sound sources with the help of interaural level differences and to correct via a second method the errors of the Phase-Difference-Cocktail-Party-Processor.

For technical applications these problems could be solved by a corresponding system design, for example with the help of multi-microphone-arrangements with different microphone distances for different frequency ranges it should be possible, to prevent ambiguities, extreme values of the transfer function and inaccurate direction estimators.