6. Algorithms for the Evaluation of interaural Level Differences ("Level-Difference-Cocktail-Party-Processor")

6.1. Problem Definition

At high frequencies and naturally ear distances the evaluation of interaural phase differences alone is not sufficient for an unambiguous determination of the input direction.

One possibility to approach this problem could be, to evaluate group delays between the signal envelopes at higher frequencies, similar to corresponding observations at the human auditory system.

Precondition for this is, certainly, that the <u>signal</u> envelopes are known.. If multiple different sound sources are present, there are ear signal envelopes, which are only caused by the interference of the different sound sources signals. For example, if two signals with different frequencies interfere, envelopes with the difference frequencies between the signals are generated.

$$\cos(\Omega_a t) + \cos(\Omega_b t) = 2 \cos\left(½(\Omega_a - \Omega_b)t\right) \cos\left(½(\Omega_a + \Omega_b)t\right)$$

Now the problem appears, to distinguish between signal envelopes (the interference of different frequency lines of <u>one</u> sound source) and inter-modulation envelopes (the interference of different frequency lines of <u>multiple</u> sound sources). Information about the input direction of a sound source can be determined from the first case, but not from the second case.

Without additional information about the characteristics of the envelopes of the sound sources a distinction between signal envelopes and inter-modulation envelopes is not possible. If the signals of the sound sources are unknown, there is nearly no possibility to determine input directions from envelope information as soon as multiple sound sources are present.

If ambiguities in the interaural phase appear, interaural level difference can be analyzed, analogous to the human auditory system. The interaural level differences of a head increase with increasing frequency, form maximal ±5 dB at 500 Hz to ±40 dB above 8 kHz. Below an algorithm shall be presented, which is able to gain information about multiple sound sources from the evaluation of interaural level differences.

6.2. Evaluation of Sound Source Attributes from the Analysis of interaural Level Differences

6.2.1. One Sound Source

If only one sound source is emitting sound, the interaural damping α_a, the input direction (respectively the interaural level difference ΔL_a) and the power of the sound source at the reference point "center of the head" (see chapter 4.1) can be evaluated from the amplitudes of the ear signals. From formula 4.1/9 results:

$$\underline{l}(t) = a_m(t)\, e^{-½\alpha_a - j½\beta_a}\, e^{j\Omega_a t + j\Phi_a} \qquad \underline{r}(t) = a_m(t)\, e^{½\alpha_a + j½\beta_a}\, e^{j\Omega_a t + j\Phi_a}$$

$$|\underline{l}(t)|^2 = a_m(t)^2\, e^{-\alpha_a} \qquad |\underline{r}(t)|^2 = a_m(t)^2\, e^{½\alpha_a}$$

$$\alpha_a = \tfrac{1}{2} \ln (|\underline{r}(t)|^2 / |\underline{l}(t)|^2)$$

$$\Delta L_a = 20 \text{ dB } \lg e^{\alpha_a} = 10 \text{ dB } \lg (|\underline{r}(t)|^2 / |\underline{l}(t)|^2) \qquad (6.2.1/1)$$

$$a_m(t)^2 = |\underline{r}(t)|\,|\underline{l}(t)|$$

With the help of the free field outer ear transfer function the input direction can be evaluated from the interaural level difference and the free field power can be evaluated from the normalized power.

If the exact free-field-outer-ear-transfer-functions are not known, or if rough approximations are sufficient, averaged relationships between interaural time and level differences according appendix C (polynomial approximation according to Gaik, straight line approximation) can be taken as a basis for directional estimation and processing.

6.2.2. Two Sound Sources

For 2 sound sources the following analytic time signals and the following ear signal amplitudes appear according to formula 4.2/1:

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

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

$$|l(t)|^2 = a_m(t)^2 e^{-\alpha_a} + b_m(t)^2 e^{-\alpha_b}$$
$$+ 2\, a_m(t)\, b_m(t)\, e^{-\frac{1}{2}(\alpha_a+\alpha_b)} \cos\left[(\Omega_a-\Omega_b)t + \Phi_a - \Phi_b - \tfrac{1}{2}(\beta_a-\beta_b)\right]$$

$$|r(t)|^2 = a_m(t)^2 e^{+\alpha_a} + b_m(t)^2 e^{+\alpha_b}$$
$$+ 2\, a_m(t)\, b_m(t)\, e^{\frac{1}{2}(\alpha_a+\alpha_b)} \cos\left[(\Omega_a-\Omega_b)t + \Phi_a - \Phi_b + \tfrac{1}{2}(\beta_a-\beta_b)\right]$$

For estimating the power and input direction of two sources from the amplitudes of the ear signals a similar algorithm can be applied, as for estimations from the interaural cross product. For integration times $2T_\mu$, which are big against the period of the instantaneous frequency difference $2\pi/(\Omega_a-\Omega_b)$, the following mean value and standard deviation of the square amplitude result, for example of the right ear signal:

$$\mu_r(t) = {}^1/_{2T_\mu} \int_{t-T_\mu}^{t+T_\mu} |r(t_\mu)|^2\, dt_\mu = a_m(t)^2 e^{\alpha_a} + b_m(t)^2 e^{\alpha_b} \qquad (6.2.2/1)$$

$$\sigma_r^2(t) = {}^1/_{2T_\mu} \int_{t-T_\mu}^{t+T_\mu} (|r(t_\mu)|^2 - \mu_r)^2\, dt_\mu = 2\, a_m(t)^2 b_m(t)^2 e^{\alpha_a+\alpha_b} \qquad (6.2.2/2)$$

From mean value and standard deviation of the ear signal amplitudes estimators for the ratio of specific sound source signals at the ear signals can be evaluated.. These estimators are called *Component-Estimators* a_r', b_r', a_l', b_l' below:

$$a_r' = \sqrt{\mu_r + \sqrt{2}\sigma_r} + \sqrt{\mu_r - \sqrt{2}\sigma_r} \qquad b_r' = \sqrt{\mu_r + \sqrt{2}\sigma_r} - \sqrt{\mu_r - \sqrt{2}\sigma_r}$$

$$a_l' = \sqrt{\mu_l + \sqrt{2}\sigma_l} + \sqrt{\mu_l - \sqrt{2}\sigma_l} \qquad b_l' = \sqrt{\mu_l + \sqrt{2}\sigma_l} - \sqrt{\mu_l - \sqrt{2}\sigma_l} \qquad (6.2.2/3)$$

$$a_r' = a_m(t)\, e^{+½\alpha_a} \qquad b_r' = b_m(t)\, e^{+½\alpha_b}$$

$$a_l' = a_m(t)\, e^{-½\alpha_a} \qquad b_l' = b_m(t)\, e^{-½\alpha_b}$$

When combining the corresponding estimators of both ear signals, a_r' with a_l' and b_r' with b_l' (corresponding means resulting from a similar formulation), then the desired estimators for the normalized power and interaural damping (or level difference) of the sound sources results to:

$$a_m'^2 = a_r'\, a_l' \qquad b_m'^2 = b_r'\, b_l'$$

$$\alpha_a' = \ln(a_r' / a_l') \qquad \alpha_b' = \ln(b_r' / b_l')$$

This method for estimating interaural damping and normalized power of two sound sources from ear signal amplitudes is called *Level-Difference-Cocktail-Party-Processor*. At the bottom of this page the algorithm is described in a condensed form. With the help of the free field outer ear transfer functions the free field power and input directions of the sound sources can be determined from the estimators. Time signals can be generated from it by a Wiener-Filter-Algorithm (see chapter 7.2).

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

1. Generation of the square amplitudes of the ear signals from the analytic time signals of the ear signals

$$|l(t)|^2 = \underline{l}(t)\, \underline{l}(t)^* \qquad |r(t)|^2 = \underline{r}(t)\, \underline{r}(t)^*$$

2. Evaluation of statistical parameters of the square amplitudes of the ear signals

$$\mu_l = 1/\, 2T_\mu \int_{t-T\mu}^{t+T\mu} |l(t_\mu)|^2\, dt_\mu \qquad \mu_r = 1/\, 2T_\mu \int_{t-T\mu}^{t+T\mu} |r(t_\mu)|^2\, dt_\mu$$

$$\sigma_l^2 = 1/2T_\mu \int_{t-T\mu}^{t+T\mu} \left(|l(t_\mu)|^2 - \mu_l\right)^2 dt_\mu \qquad \sigma_r^2 = 1/\, 2T_\mu \int_{t-T\mu}^{t+T\mu} \left(|r(t_\mu)|^2 - \mu_r\right)^2 dt_\mu$$

3. Evaluation of the component estimators

$$a_r' = \sqrt{\mu_r + \sqrt{2}\sigma_r} + \sqrt{\mu_r - \sqrt{2}\sigma_r} \qquad b_r' = \sqrt{\mu_r + \sqrt{2}\sigma_r} - \sqrt{\mu_r - \sqrt{2}\sigma_r}$$

$$a_l' = \sqrt{\mu_l + \sqrt{2}\sigma_l} + \sqrt{\mu_l - \sqrt{2}\sigma_l} \qquad b_l' = \sqrt{\mu_l + \sqrt{2}\sigma_l} - \sqrt{\mu_l - \sqrt{2}\sigma_l}$$

4. Estimation of the normalized power and interaural damping of the sound sources

$$a_m'^2 = a_r' a_l' \qquad \alpha_a' = \ln(a_r'/a_l') \qquad b_m'^2 = b_r' b_l' \qquad \alpha_b' = \ln(b_r'/b_l')$$

By combining phase difference and level difference processor ambiguities and error sources of both methods can be compensated, like ambiguous estimations of the input direction for higher frequencies by the phase difference algorithm, inaccurate estimations because of too low level differences at low frequencies by the level difference algorithm. Most likely are those estimated power and input directions, which are supported by both methods or where only small deviations appear between both methods (see chapter 6.3).

If one sound source with varying amplitude is presented instead of 2 sound sources with constant amplitude, then, similar to the interaural cross product, 2 estimators with the same interaural level difference are generated.

6.2.3. More than two Sound Sources

More complex sound fields can be described by the interference of multiple sound sources with different spectra, according to formula 5.4.1/1. The square amplitudes of the ear signals can be evaluated in analogy to the derivation of the interaural cross product in chapter 5.4.1. Herewith the right ear signal results to:

$$|r(t)|^2 = \sum_{q=1}^{N} e^{\alpha_a} \sum_{i=1}^{M_q} a_{mqi}^2 + 2 \sum_{q=1}^{N} \sum_{p=1}^{q-1} e^{½(\alpha_q+\alpha_p)} \sum_{i=1}^{M_q} \sum_{k=1}^{M_p} a_{mqi}\, a_{mpk} \cos\{(\Omega_i-\Omega_k)t+\Phi_{qi}-j\Phi_{pk} +½(\beta_p-\beta_q)\} \quad (6.2.3/1)$$

Correspondingly mean value and standard deviation result to:

$$\mu_r = \sum_{q=1}^{N} e^{\alpha_a} \sum_{i=1}^{M_q} a_{mqi}^2$$

$$\sigma_r^2 = 2 \sum_{q=1}^{N} \sum_{p=1}^{q-1} e^{\alpha_q+\alpha_p} \sum_{i=1}^{M_q} a_{mqi}^2 \sum_{k=1}^{M_p} a_{mpk}^2 \quad (6.2.3/2)$$

The mean value corresponds to the sum of the ear signal power of all sound sources. The standard deviation corresponds to the superposition of the harmonic oscillations of all difference frequencies between each 2 frequency lines of different sound sources.

The statistical parameters of a complex sound field can be derived, in analogy to formula 5.4.1/4, from the statistical parameters of the individual source signals μ_q,σ_q. This results to the superposition theorem:

$$\mu_r = \sum_{q=1}^{N} \mu_q\, e^{\alpha_q}$$

$$\sigma_r^2 = \sum_{q=1}^{N} \sigma_q^2\, e^{2\alpha_q} + 2 \sum_{q=1}^{N} \sum_{p<qt}^{q-1} \mu_q\, \mu_p\, e^{\alpha_q+\alpha_p} \quad (6.2.3/3)$$

Correspondingly the considerations from the interaural cross product can be assigned to the estimators of the square amplitudes of the ear signals. If, for example, multiple sound sources with a constant amplitude ($\mu_q=a_{mq}^2,\sigma_q=0$) are present, the statistical parameters result therefore to::

$$\mu_r = \sum_{q=1}^{N} a_{mq}^2 \, e^{\alpha_q} \qquad \sigma_r^2 = 2 \sum_{q=1}^{N} \sum_{p<q}^{q-1} a_{mq}^2 \, a_{mp}^2 \, e^{\alpha_q + \alpha_p}$$

In order to get information, whether and to which extend signals of a certain direction are present in such complex sound fields, additional algorithms are necessary, similar to the correction algorithms of the Phase-Difference-Cocktail-Party-Processor. These correction algorithms shall allow, to determine the sound power of a desired direction from the statistical parameters and estimators even in such complex sound fields..

6.2.4. Diffuse Sound Field

Diffuse Sound Field alone

According to chapter 5.4.4 the ear signals in a diffuse sound field (reference point "center of the head") result to ($\underline{e}_{m\theta}(t)$ signals of the mirror sound sources):

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

$$|l(t)|^2 = \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t) \, e^{-½\alpha_\theta - j½\beta_\theta} \, d\theta + \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t)^* \, e^{-½\alpha_\theta + j½\beta_\theta} \, d\theta$$

$$|r(t)|^2 = \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t) \, e^{½\alpha_\theta + j½\beta_\theta} \, d\theta + \int_{-\pi}^{\pi} \underline{e}_{m\theta}(t)^* \, e^{+½\alpha_\theta - j½\beta_\theta} \, d\theta$$

If the power density E_m' and the spectra of the mirror sound sources are identical for all input directions, then the ear signals for a symmetric head result, analogous to chapter 5.4.4, to:

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

Under these conditions the square amplitudes of both ear signals are identical and no longer time variant, the standard deviation results to zero. The statistical parameters and estimators result to:

$$\mu_r = \mu_l = \mu = |r(t)|^2 \qquad \sigma_r = \sigma_l = 0$$

$$a_r'^2 = a_l'^2 = \mu_r \qquad b_r'^2 = b_l'^2 = 0$$

$$a_m'^2 = a_r' \, a_l' = \mu \qquad b_m'^2 = b_r' b_l' = 0$$

$$\alpha_a' = \ln (a_r'/a_l') = 0 \qquad (6.2.4/1)$$

In a pure diffuse sound field only one estimator for the median plane appears, which incorporates the total power of the ear signals. Here the behavior of Level-Difference-Cocktail-Party-Processor is quite similar to the behavior of the Phase-Difference-Cocktail-Party-Processor. Since only the late reverberation has been considered here, but not the early reflections, conclusions about the hearing in closed rooms cannot be drawn from it without fail. This characteristics to project the late reverberation onto the front-direction, could possibly be used for de-reverberation algorithms.

One Sound Source in the diffuse Sound Field

For a single sound source inside a diffuse sound field the statistical parameters of the ear signals can be evaluated from the corresponding parameters of the source signals μ_q,σ_q and of the diffuse sound field μ_d,σ_d (σ_d=0) with the help of the superposition theorem according to formula 6.2.3/3.

$$\mu_l = \mu_q\, e^{-\alpha_q} + \mu_d \qquad \mu_r = \mu_q\, e^{\alpha_q} + \mu_d$$

$$\sigma_l^2 = \sigma_q^2\, e^{-2\alpha_q} + 2\, \mu_q\, \mu_d\, e^{-\alpha_q} \qquad \sigma_r^2 = \sigma_q^2\, e^{2\alpha_q} + 2\, \mu_q\, \mu_d\, e^{\alpha_q}$$

If the amplitude of the sound signal changes only slowly, compared to the integration time, then it applies $\sigma_q^2 \ll \mu_q\mu_d$, and the component estimators and the estimators for amplitude and interaural damping of the concerned sound sources apply to:

$$a_l'^2 = \mu_q\, e^{-\alpha_q} \qquad a_r'^2 = \mu_q\, e^{\alpha_q}$$

$$b_l'^2 = \mu_d \qquad b_r'^2 = \mu_d$$

$$a_m'^2 = \mu_q \qquad b_m'^2 = \mu_d$$

$$\alpha_a' = \alpha_q \qquad \alpha_b' = 0 \qquad (6.2.4/2)$$

Similar to the analysis of the interaural cross product 2 independent estimators are obtained, one for the diffuse field and one for the sound source, as an estimator for the-reverberated source signal.

For a sound source in the median plane there is no possibility to separate the source signals from the diffuse sound field. The power of the diffuse sound field is added to the power of the sound source.

Diffuse Sound Field with two Sound Sources

If there are 2 sound sources with a constant amplitude ($\mu_a=a_m^2$, $\mu_b=b_m^2$; $\sigma_a,\sigma_b \ll \mu_a,\mu_b$) and with the interaural damping α_a,α_b, inside a diffuse sound field (μ_d; σ_d), then the statistical parameters of the monaural square amplitudes result to (for example for the right ear signal):

$$\mu_r = a_m^2\, e^{\alpha_a} + b_m^2\, e^{\alpha_b} + \mu_d$$

$$½\sigma_r^2 = a_m^2\, b_m^2\, e^{\alpha_a+\alpha_b} + \mu_d\, (a_m^2\, e^{\alpha_a} + b_m^2\, e^{\alpha_b})$$

$$(a_r' \pm b_r')^2 = a_m^2\, e^{\alpha_a} + b_m^2\, e^{\alpha_b} + \mu_d \pm 2\sqrt{(a_m^2\, e^{\alpha_a} + \mu_d) + (b_m^2\, e^{\alpha_b} + \mu_d) - \mu_d^2}$$

If both sound sources are dominant compared to the diffuse sound field, meaning that a listener is inside the reverberation radius, then estimators of the source signals result, which are interfered by the diffuse sound field each.

$$a_r'^2 \approx a_m^2 e^{\alpha_a} + \mu_d \qquad a_l'^2 \approx a_m^2 e^{-\alpha_a} + \mu_d$$

$$b_r'^2 \approx b_m^2 e^{\alpha_b} + \mu_d \qquad b_l'^2 \approx b_m^2 e^{-\alpha_b} + \mu_d$$

$$a_m'^2 \approx a_m^2 + \mu_d \cosh(\alpha_a) \qquad e^{2\alpha_a} \approx (a_m^2 e^{\alpha_a} + \mu_d) \,/\, (a_m^2 e^{-\alpha_a} + \mu_d)$$

$$b_m'^2 \approx b_m^2 + \mu_d \cosh(\alpha_b) \qquad e^{2\alpha_b} \approx (b_m^2 e^{\alpha_b} + \mu_d) \,/\, (b_m^2 e^{-\alpha_b} + \mu_d)$$

The following case distinction applies for the interaural damping (α_b correspondingly):

$\alpha_a >> 0$: $\quad \alpha_a' \approx \alpha_a - (\mu_d/a_m^2)e^{\alpha_a}$

$\alpha_a << 0$: $\quad \alpha_a' \approx \alpha_a + (\mu_d/a_m^2)e^{-\alpha_a}$

$\alpha_a \approx 0$: $\quad \alpha_a' \approx \alpha_a$

$$|\alpha_a'| \approx |\alpha_a| - (\mu_d/a_m^2)e^{|\alpha_a|}$$

By the diffuse sound field the estimated power is increased, compared to the original value, and the estimated directions are displaced towards the median plane.

Outside the reverberation radius there is one estimator for the diffuse sound field and one estimator for the superposition of both source signals.

$$a_l'^2 \approx \mu_d \qquad a_r'^2 \approx \mu_d$$

$$b_l'^2 \approx a_m^2 e^{-\alpha_a} + b_m^2 e^{-\alpha_b} \qquad b_r'^2 \approx a_m^2 e^{\alpha_a} + b_m^2 e^{\alpha_b}$$

$$a_m'^2 \approx \mu_d \qquad e^{2\alpha_a'} \approx 1$$

$$b_m'^2 \approx \sqrt{a_m^2 + b_m^2 + 2a_m b_m \cosh(\alpha_a - \alpha_b)} \qquad e^{2\alpha_b'} \approx e^{\alpha_a + \alpha_b} \frac{\cosh\{(\alpha_a - \alpha_b)/2 + \ln(a_m/b_m)\}}{\cosh\{(\alpha_a - \alpha_b)/2 - \ln(a_m/b_m)\}}$$

6.2.5. Dominant Sources

A complex sound field with multiple sound sources frequency lines can be described by the statistical parameters of the ear signal amplitudes according to formula 6.2.3/3, for example for the right ear signal:

$$\mu_r = \sum_{q=1}^{N} \mu_q e^{\alpha_q}$$

$$\sigma_r^2 = \sum_{q=1}^{N} \sigma_q^2 e^{2\alpha_q} + 2 \sum_{q=1}^{N} \sum_{p<q}^{q-1} \mu_q \mu_p e^{\alpha_q + \alpha_p}$$

$\mu_p, \mu_q, \sigma_p, \sigma_q$ are mean values and standard deviations of the source signals.

One dominant Source

If one source (a_m^2, α_a) is dominant, meaning, that their power a_m^2 prevails the power of all other sources within the ear signals, then this source can be separated from a mixture of sources. If there is a dominant source with standard deviation $\sigma_a=0$ and if the other source signals can be described by by individual frequency lines ($\sigma_i=0$), then the right ear signal, for example, results to:

$$\mu_r = a_m^2 e^{\alpha_a} + \sum_{q \neq a} \mu_q e^{\alpha_q}$$

$$\sigma_r^2 = a_m^2 e^{\alpha_a} \sum_{q \neq a} \mu_q e^{\alpha_q} + 2 \sum_{q \neq a} \sum_{p<q} \mu_q \mu_p e^{\alpha_q + \alpha_p}$$

The component estimators result to:

$$a_r'^2 \approx a_m^2 e^{\alpha_a} \qquad b_r'^2 \approx \sum_{q \neq a} \mu_q e^{\alpha_q}$$

$$a_l'^2 \approx a_m^2 e^{-\alpha_a} \qquad b_l'^2 \approx \sum_{q \neq a} \mu_q e^{-\alpha_q}$$

As a result there is one estimator, which describes the signal of the dominant sound source and one estimator, which corresponds to a mixture of all other sound sources.

$$a_m'^2 = a_m^2 \qquad\qquad \alpha_a' = \alpha_a$$

$$b_m'^4 = \sum_{q\neq a}\mu_q^2 + 2\sum_{q\neq a}\sum_{p<q}\mu_q\,\mu_p\cosh(\alpha_q-\alpha_p) \qquad\qquad e^{\alpha_b'} = b_r' / b_l'$$

Two dominant Sources

If 2 dominant sources a_m^2, b_m^2 appear, the following statistical parameters result under the same preconditions as above, for example for the right ear signal:

$$\mu_r = a_m^2 e^{\alpha_a} + b_m^2 e^{\alpha_b} + \sum_{q\neq a,b}\mu_q e^{\alpha_q}$$

$$\tfrac{1}{2}\sigma_r^2 = a_m^2 b_m^2 e^{\alpha_a+\alpha_b} + \left(a_m^2 e^{\alpha_a} + b_m^2 e^{\alpha_b}\right)\sum_{q\neq a,b}\mu_q e^{\alpha_q} + 2\sum_{q\neq a,b}\sum_{p<q}\mu_q\,\mu_p\, e^{\alpha_q+\alpha_p}$$

Correspondingly the following component estimators result from it:

$$a_r'^2 \approx a_m^2 e^{\alpha_a} + \sum_{q\neq a,b}\mu_q e^{\alpha_q} \qquad\qquad b_r'^2 \approx b_m^2 e^{\alpha_b} + \sum_{q\neq a,b}\mu_q e^{\alpha_q}$$

$$a_l'^2 \approx a_m^2 e^{-\alpha_a} + \sum_{q\neq a,b}\mu_q e^{-\alpha_q} \qquad\qquad b_l'^2 \approx b_m^2 e^{-\alpha_b} + \sum_{q\neq a,b}\mu_q e^{-\alpha_q}$$

Because of the influence of the weak sources the estimators for the source signals deviate from the signals characteristics of the dominant sources. The deviations depend on the power ratio between the dominant sources and the other sound sources.

In order to extract arbitrary signals from a desired direction out of a mixture of sources, a method is required, which estimates the possible power ratio between dominant sources and other sources from the deviations between estimated direction and desired direction and which corrects in this way the possible sound power of the desired direction.

Discussion

If more than 2 dominant sources are present, the sources can no longer be extracted correctly. The estimators are then primarily determined from a mixture of the dominant sound sources.

Problems in constructing a signal estimator can appear, if different sources are dominant in only one ear (for example at big interaural level differences). The ear specific component estimators can than correspond to different sources for each ear, so that no existing source signal can be extracted by combining the estimators.

> Example: For 3 sound signals from the directions -45°,0°,45° interaural level differences of -10 dB, 0 dB, 10 dB result in the frequency range around 3 kHz. If the sound power is identical for each sound source, the sound sources 1 and 2 are dominant in the left ear signal and the sound sources 2 and 3 in the right ear signal. The power of all component estimators would then be identical, the estimated interaural level difference would then be 0 dB. It would be impossible to determine the directions of the lateral signals.

For more than 2 sound sources it is therefor necessary to use correction methods, which weight appearing deviations of the estimators from a desired direction and which correct the estimators, if necessary; or correct the result by using results of other information sources.

6.3. Mapping of Estimators to a desired Direction

Error Sources

If the estimated directions do not correspond to a desired direction, it has to be evaluated, similar to the weighting methods of the phase difference algorithm (chapter 5.6), whether a signal of the desired direction is present and which power it can have, possibly.

The error causes, which can provoke deviations of the estimated direction from a desired direction, even if a signals of the desired direction is present, are in many cases similar to the error causes of the Phase-Difference-Cocktail-Party-Processor (see chapter 5.6):

- more than 2 sound sources,
- modulated sound sources,
- reflections and reverberation,
- errors in specifying the desired direction or errors in determining the interaural damping, which corresponds to the desired direction,
- inadequate integration time for evaluating the statistical parameters,
- frequency dependent interaural level differences, especially at high frequencies.

The impacts of these error causes on the results of the method shall be investigated below, in order to derive correction and valuation methods for deviating source estimators from it. Similar to chapter 5.6 a weighting function $W_x(\alpha)$.shall be defined, which can be used to generate a corrected estimator for the desired direction $s_m'^2$.

Frequency Dependencies

At high frequencies the interaural damping within a critical band is no longer frequency independent.. Within the free field outer ear transfer functions of a critical band variations of the interaural level difference of up to ±5 dB can appear (in the approximated transfer functions according to appendix C still up to ±2 dB). So any deviations of this size have to be rated as correct values. Then the weighting function of an estimator $W_x(\alpha)$ and the estimator of the desired direction $s_m'^2$ applies to ($x_m'^2 = a_m'^2$ or $b_m'^2$):

$$W_x(\alpha_{soll}) = 1 \qquad s_m'^2 = x_m'^2 \qquad \text{for} \quad |\alpha_x' - \alpha_{soll}| < \left(\frac{\Delta\alpha}{\Delta f}\right)_{max} \frac{f_o - f_u}{2}$$

Valid Estimator Range

The sum of the power of the component estimators has to correspond to the mean ear signal power. The ear signal power of an estimator of the desired direction has therefore to range between the mean value of the square amplitude and the computed component estimators. The possible interaural damping range is given by combining the interaural attributes of these corner marks. It is rather unlikely, that signals with an interaural damping outside this range are present in the current sound situation. The deviation of the estimated interaural damping from the interaural damping of the desired direction can therefore be a measure for the probability that signals of the desired direction are present.

If a_r' and a_l' are the component estimators with the biggest power, then an estimator shows the biggest deviation interaural damping from the mean value, if the component estimators are combined to signal estimators in a wrong way.. The interaural damping of these estimators α_a'' and α_b'' results with the notation of chapter 6.2.2 as follows:

$$e^{\alpha_a''} = \sqrt{a_r'/b_l'} \qquad e^{\alpha_\mu} = \sqrt{\mu_r'/\mu_l'}$$

$$\frac{e^{2\alpha_a''}}{e^{2\alpha_\mu}} = \frac{\mu_r + \sqrt{2}\sigma_r}{\mu_l - \sqrt{2}\sigma_l}\frac{\mu_l}{\mu_r} = \frac{1 + 2a_r'b_r' / (a_r'^2+b_r'^2)}{1 - 2a_l'b_l' / (a_l'^2+b_l'^2)} \approx 1 + 2\frac{b_r'}{a_r'} + 2\frac{b_l'}{a_l'}$$

$$\alpha_a''-\alpha_\mu \approx b_r'/a_r' + b_l'/a_l'$$

With the same considerations the estimator b" results to:

$$\alpha_b''-\alpha_\mu \approx - b_r'/a_r' - b_l'/a_l'$$

This interaural damping difference between estimator and mean value can be used to construct a weighting function. The probability, that signals of the desired direction are present, is set to 0.5, if the interaural damping of an estimator α_x' deviates from the interaural damping of the desired direction α_{soll} by the difference stated above. With these considerations a weighting function can be generated,with the help of a cosine window:

$$|\alpha_x''-\alpha_\mu| \leq \frac{\min(a_r',b_r')}{\max(a_r',b_r')} + \frac{\min(a_l',b_l')}{\max(a_l',b_l')}$$

$$W_x(\alpha) = ½ + ½ \cos\left(\frac{\alpha_x'-\alpha}{\alpha_x''-\alpha_\mu}\,\frac{\pi}{2}\right) \qquad \text{for } |\alpha_x'-\alpha| < 2|\alpha_x''-\alpha_\mu|$$

$$W_x(\alpha) = 0 \qquad \text{else} \qquad (6.3/1)$$

An estimator of the desired direction $s_m'^2$ results form a rough estimator of the Phase-Difference-Cocktail-Party-Processor $x_m'^2$ by::

$$s_m'^2 = W_x(\alpha_{soll})\ x_m'^2$$

Three Source Model

The presented model describes the statistical parameters of the ear signals as the result of the interference of two sound sources. If more than 2 sound sources are present, this approach is no longer sufficient for a complete description of the sound situation. The estimators will deviate from the parameters of the sound sources.

Besides the sound source of the desired direction (power s_m^2, damping α_{soll}) two further sound sources shall be present (power a^2,b^2; interaural damping α_a,α_b). It is assumed, that the estimators a_m',α_a',b_m',α_b' result each from interferences of the signal of the desired direction with the signals of another direction. Therefore the portion of signals of the desired direction in the estimated power g has to be determined. The statistical parameters must be identical, if the situation is described by estimators or by ear signals. Therefore the right ear signal results, for example, to:

$$\mu_r = a_r'^2 + b_r'^2 = a_r^2 + b_r^2 + s_r^2$$

$$a_r'^2 = a_r^2 + g\, s_r^2 \qquad b_r'^2 = b_r^2 + (1-g)s_r^2$$

$$½\sigma_r^2 = a_r'^2 b_r'^2 = a_r^2 b_r^2 + a_r^2 s_r^2 + b_r^2 s_r^2 = a_r'^2 b_r'^2 + s_r^2\left[g\, a_r'^2 + (1-g)b_r'^2 - (g^2-g+1)s_r^2\right]$$

$$g\, a_r'^2/s_r^2 + (1-g)b_r'^2/s_r^2 = g^2-g+1$$

The left ear signal results correspondingly to:

$$g\, a_l'^2/s_l^2 + (1-g)b_l'^2/s_l^2 = g^2-g+1$$

The factor g and the power of the desired direction s_m^2 results to::

$$g\, a_m'^2\, e^{\alpha_{soll}-\alpha_a'} + (1-g)b_m'^2\, e^{\alpha_{soll}-\alpha_b} = g\, a_m'^2\, e^{\alpha_a'-\alpha_{soll}} + (1-g)b_m'^2\, e^{\alpha_b'-\alpha_{soll}}$$

$$g = \frac{b_m'^2 \sinh(\alpha_b'-\alpha_{soll})}{b_m'^2 \sinh(\alpha_b'-\alpha_{soll}) - a_m'^2 \sinh(\alpha_a'-\alpha_{soll})}$$

$$s_m^2 = \frac{g\, a_m'^2}{g^2-g+1}\, e^{\alpha_a'-\alpha_{soll}} + \frac{(1-g)\, b_m'^2}{g^2-g+1}\, e^{\alpha_b'-\alpha_{soll}} \qquad (6.3/2)$$

s_m^2 can act as an estimation for the emitted power of the desired direction, if three sound sources are active, and represents a corrected estimator for the case that deviations between the estimated directions and the desired direction appear.

6.4. Combining Estimators from interaural Time and Level Differences

Method

Hence 2 Cocktail-Party-Processor algorithms are available now, each with own methods for correcting estimation errors.

- The Phase-Difference-Cocktail-Party-Processor according to chapter 5 evaluates from the interaural cross product estimators for the power and interaural phase of 2 sound sources.
- The Level-Difference-Cocktail-Party-Processor evaluates from the amplitudes of the ear signals estimators for the power and interaural damping (interaural level difference) of 2 sound sources by using a similar processing method.

The Phase-Difference-Cocktail-Party-Processor provides at low frequencies unambiguous results for a 2-source-situation. If the receiver distance matches to natural ear distances, the interaural phase becomes ambiguous for frequencies above 800 Hz. There is no possibility, to determine the "unwrapped phase" or interaural wave number directly from this method and to evaluate unambiguous estimators out of it.

Optimal for the Level-Difference-Cocktail-Party-Processor are higher frequencies, where the interaural level differences are big enough to provide reliable and exact directional estimators. For low frequencies and low interaural level differences the estimation errors of this method grows. This means, that the optimal frequency ranges of both processors complement one another:

- For low frequencies the Phase-Difference-Cocktail-Party-Processor alone can provide reliable results.
- For medium frequencies of some kHz the results of the Phase-Difference-Cocktail-Party-Processor become ambiguous, but the unambiguously describable direction range has still a reasonable size. Here the Level-Difference-Cocktail-Party-Processor has to determine the rough directions and make a directional decision between the ambiguities of the Phase-Difference-Cocktail-Party-Processor.
- For high frequencies only the results of the Level-Difference-Cocktail-Party-Processor can be evaluated, since the ambiguities of the Phase-Difference-Cocktail-Party-Processor lie so close to each other, that a reliable evaluation of these results is impossible.

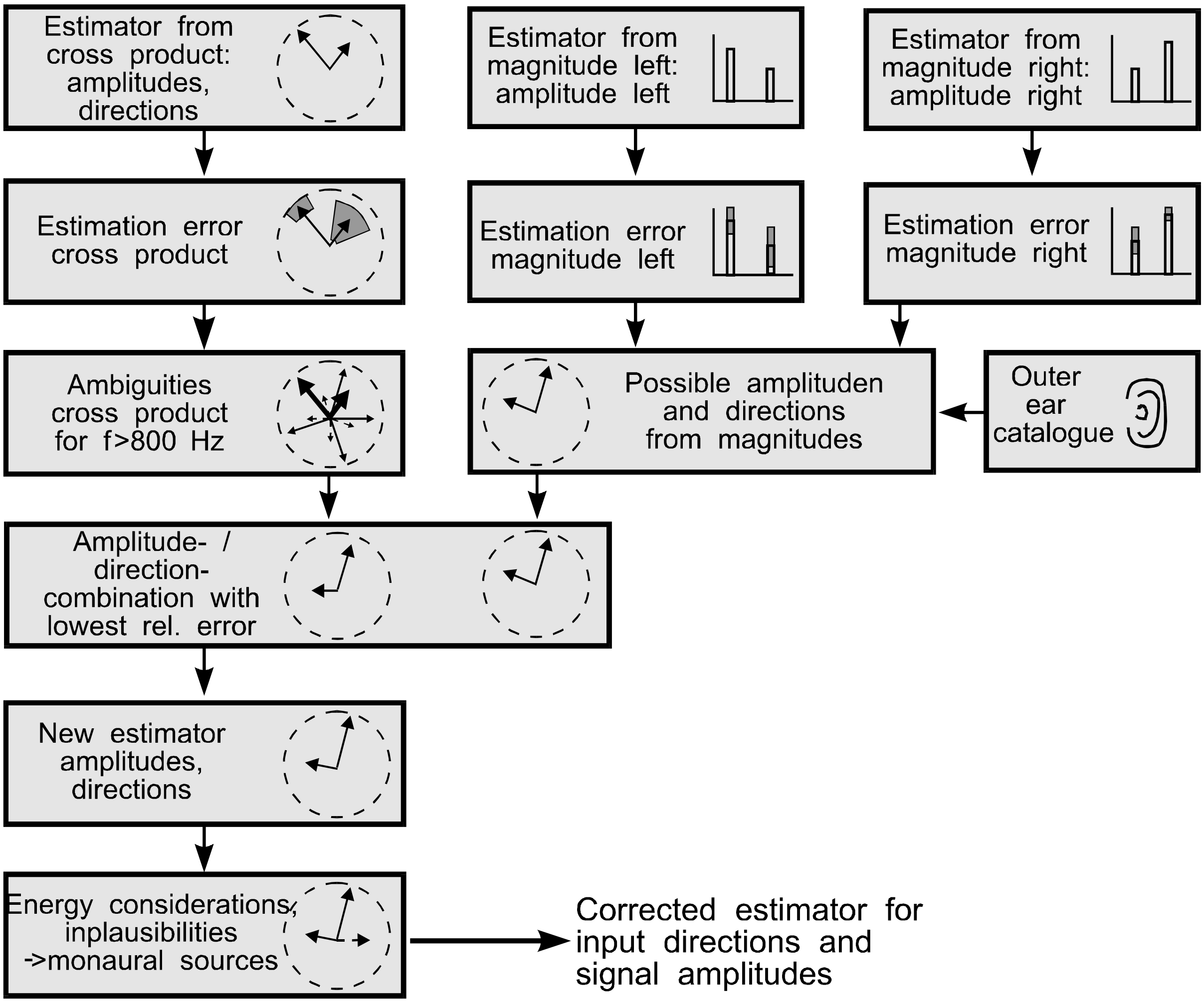

Fig. 6.1: Processing unit for combining phase difference and level difference information.

By using information of additional information sources (for example the precedence processor, chapter 8) the evaluation strategies can be enhanced beyond that.

When comparing this method with other methods of binaural analysis, analogous to chapter 4.5/4.6, this methods corresponds to a combination of the results of a binaural cross correlation function and monaural auto correlation functions. Specific for the described processors are the used functions (interaural cross product, amplitudes of the ear signals), which allow an easy and very fast computation of the results.. A further specialty is the evaluation procedure. By evaluating the standard deviation information about the time dependent behavior of the patterns is included into the evaluation. Through this the possibility arises, to estimate the parameters of two sources simultaneously. As a result of this even for 2-source situations with negative signal-to-noise-ratios accurate results can be achieved, where models without evaluation of time dependent pattern structures (pure instantaneous value or mean value evaluations like cross correlation models), which follow a 1- source-approach internally, often get problems to evaluate the desired direction correctly..

Dealing with Ambiguities

Figure 6.1 shows the concept of a processing unit, which combines the information from phase difference and level difference evaluation. This unit is based on the following considerations:

- At the human head (or at microphone equipped items) interaural time and level differences only appear in "natural combinations", which are specific for a distinct frequency and input direction.
- A fused auditory event is only possible, if these "natural combinations" appear (Gaik [20])..
- If natural combinations appear between the estimators of the Phase-Difference-Cocktail-Party-Processor and the Level-Difference-Cocktail-Party-Processor these results have a bigger probability than results which lead to unnatural combinations.

For the construction of a layer, which connects both processors, this means, that :after checking the possible estimators on plausibility and rating them in each processor type individually, the possible combinations of phase and level difference estimators are rated, e.g. by weighting functions $W_x(\beta_{soll})$, $W_x(\alpha_{soll})$. The pair of estimators with the fewest deviation from a natural combination will be taken as basis for the subsequent processing.

Unnatural Combinations of Time and Level Differences

Resulting combinations of interaural time and level differences, which do not correspond to outer ear transfer functions, indicate that errors have appeared. From the reactions of the auditory system on such inconsistent information, possibly a method can be developed to cope with such errors, which might be useful for the improvement of Cocktail-Party-Processors.

According to investigations of Gaik [20] 2 auditory events appear at such unnatural combinations, one auditory event, which is slightly displaced from the direction of the interaural time difference and one monaural auditory event at the ear, where the level is increased compared to the natural combinations. Therefore the auditory system seems to interpret situations with inconsistent interaural parameters as the result of the interference of several sound sources.

For signals with unnatural combinations the two processors provide the following result:

- The Phase-Difference-Cocktail-Party-Processor evaluates one or two estimators, whose directions correspond to the interaural phase differences. The evaluated directions are ambiguous at higher frequencies. The number of estimators depends on the signal envelope (constant or modulated). The power ratio of the estimators describes the modulation rate of the envelope.
- The Level-Difference-Cocktail-Party-Processor also evaluates 2 estimators, whose power ratio also describes the modulation rate of the envelope.

For the results of each processor mean value and standard deviation of the directional estimation are determined, and a function is generated, which indicates - including all ambiguities - with which probability and maximal power an estimator of a certain direction can appear (compare chapter 5.6 and 6.3). If the estimated directions match for both processors, the estimators are combined to an overall estimator. Elsewhere the estimators of both processors are combined in such a way, that the standard deviation ranges for the directional estimation overlap or the combined probability for the directional estimation from both methods becomes maximal.

For unnatural combinations of interaural parameters there are inconsistencies between the directional estimation of both processors, there is no estimator combination with a high probability. According to the results of the psychoacoustical investigations the estimator with the lowest probability (e.g. with the highest variance) should be re-interpreted in this case, on the one hand to an estimator, which supports the estimations of the other processor and whose direction therefore has to be changed, and on the other hand to a monaural compensation estimator, which shall keep the statistical parameters consistent, although the direction of the first estimator has been changed.

6.5. Perspective: Combination of different spatial Analysis and Processing Methods

For the spatial processing of ear signals humans have a couple of directional information available, which exceed the capabilities of the both presented processor types: interaural phases, time and level differences, directional estimators from rising slopes (see chapter 8 and Wolf [49]), optical information, prior knowledge.

In order to coordinate and process this information within enhanced binaural models the information content of each information source should be rated. A rating criterion could be the variance of this information or the probability, with which an information can be considered as correct (for example by using similar methods than proposed in chapter 5.6 and 6.3).

The direction, which is supported by many information sources, or the direction with the highest probability could be defined as a new direction of attention by a central processing layer and be taken as the the new desired direction for the Cocktail-Party-Processor-Algorithms. If estimations are available, how probable the appearance of a certain direction is, the power, which is estimated by the Cocktail-Party-Processors, could be weighted by this probability and so be mapped to the desired direction.

If, as proposed here, all directional and signal power information, would be available for a longer time frame at a central processing unit, then, for example, the parallel processing of two different directions and signals could be modeled and remembrance effects as well, which mean the retroactive selection of a different information source.

In chapter 8 an attempt is made, to derive construction principles for a central processing unit of a Cocktail-Party-Processor from psychoacoustical findings about the dynamic behavior of the auditory system (Precedence-Effect). This shall lay the foundations for the modeling of a central unit, which decides about the desired direction of a Cocktail-Party-Processor without any external presetting. This unit could, for example, be used to follow a moving speaker inside a room and process his sound signals.

This would result into a processing framework for Cocktail-Party-Processors, which evaluates interaural phase differences and interaural level differences in two parallel layers and which can estimate the signal power and direction of two sound sources simultaneously. In complex sound situations a desired direction for processing should be specified, either through external definition or via a central processing unit, which has been outlined roughly above. For a complete construction of a signal processing system an additional a pre-processing unit is necessary, which prepares the received sound signals for the directional analysis and an additional post-processing-unit, which converts the estimators for the mean signal power into processed time signals. More information about that can be found in the following chapter.