7. A Signal Processing Framework for binaural Models

7.1. Processing of the Input Signals

The pre-processing unit of the framework has to prepare the recorded real time signals for the analysis by the Cocktail-Party-Processors. Since the Cocktail-Party-Processors process analytic time signals inside critical bands, a pre-processing unit has to contain the following processing steps: critical band filtering of the signals, generation of the analytic time signals, possibly data reduction. The schematic diagram of the pre-processing unit is depicted in Fig. 7.1.

7.1.1 Critical-Band-Filters

The signals shall be processed inside critical bands, according to the human auditory system. The processing inside critical bands has the following properties:

- Rather fast variations of the sound signal parameters can be detected. The possible time resolution for critical band filter lies, according to the time-bandwidth-product, between 10 ms at low frequencies (bandwidth 100 Hz) and some 100 µs at high frequencies (bandwidth up to 3 kHz).
- The number of frequency ranges is big enough, in order to evaluate different signal characteristics in different frequency ranges.

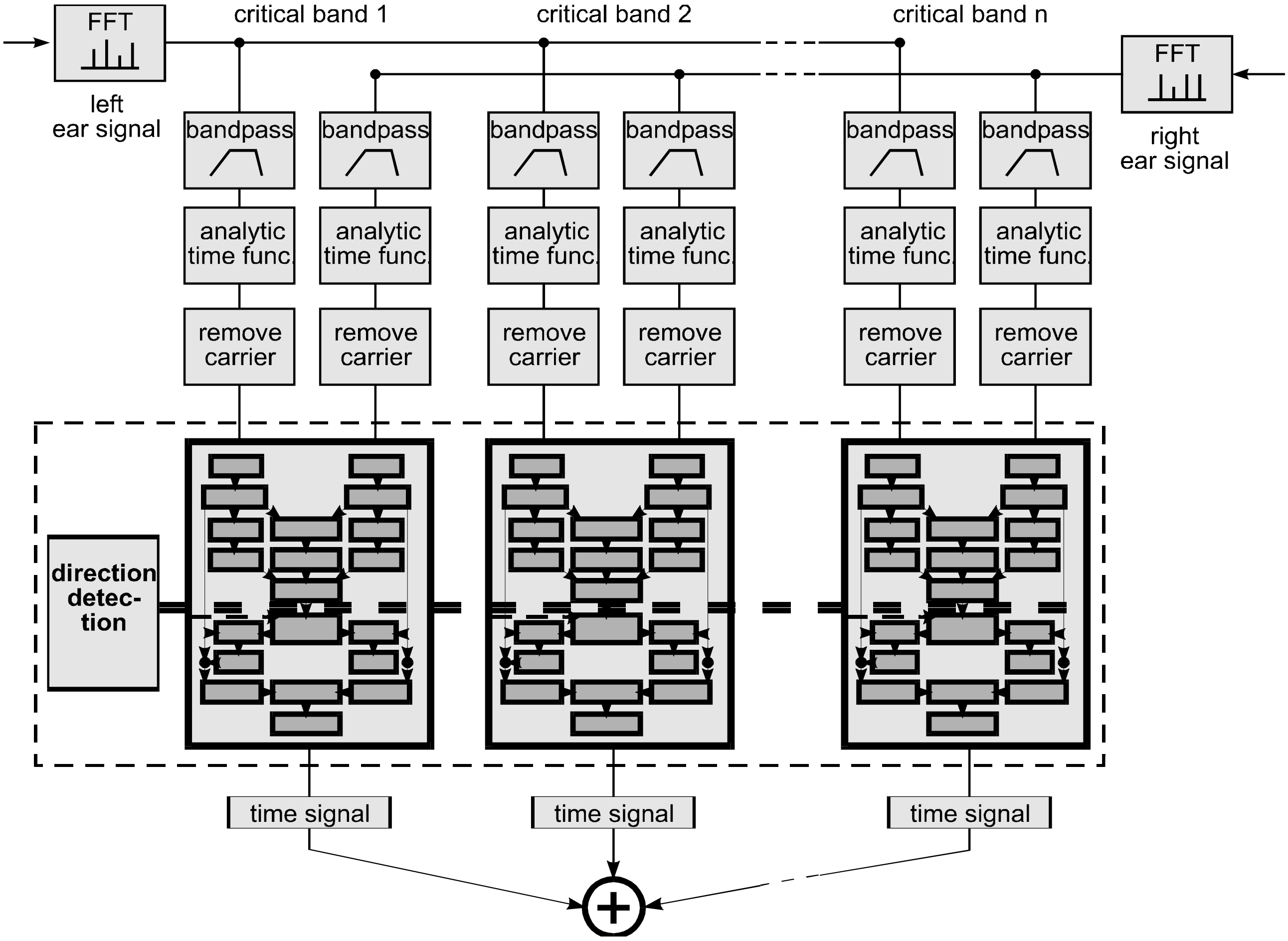

Fig. 7.1: Signal processing framework for the binaural model
The model structure inside the critical bands corresponds to Fig.7.5

- Analysis bandwidth and time resolution are adapted to the characteristics of the auditory system.

A filter method for critical band filtering has to fulfil the following requirements:

- Only the desired frequency range shall be filtered out, if possible.
- The signals shall not be changed inside the pass band.
- In order to be able to analyze cross references between the signals of different critical bands the time delays of the filters shall be as small as possible.
- In order to be able to process long signals sections efficiently, the impulse responses of the filter shall not exceed a certain length (usage of the Overlap-Add-Method).
- In order to achieve a high time resolution at a fixed bandwidth, the impulse response of the band pass filters shall be as short as possible.
- The filter algorithm shall be applicable for different signal types (real time function, analytic time signal, modulation functions) and be able to work in the time domain and in the frequency domain as well.

A delay free, non-causal filter method in the frequency domain has been chosen (phase of the transfer function=0) with an impulse response, which is symmetric to the time t=0. The transfer functions are therefore constructed that way, that if possible no discontinuities appear in low order derivations of the transfer function (see appendix D). By this means very steep decreasing impulse responses can be achieved and errors can be reduced to a minimum, which are caused by cutting the impulse response to a predefined length. Thus in spite of relatively short impulse responses very steep filter slopes can be achieved.

The following method is used for the construction of an optimized transfer function (see Slatky [37] and appendix D):

- Definition of cut-off-frequencies, construction of the raw transfer function with the help of a "function kit" (straight line segments, Cosine-functions, exponential functions, Gauss-functions et.al.),
- if necessary, smoothing of the raw transfer function (sliding average with selectable averaging window),
- Fourier-Transformation to evaluate the impulse response,
- Cutting the impulse response to a predefined length with the help of a window function (selectable from the "function kit"),
- Transformation back to the frequency domain with increased frequency resolution, controlling the results.

The filtering is performed by an adapted Overlap-Add-Algorithm. The method of the frequency transformation depends hereby on the desired signal type: real FFT for generating real time signals, complex FFT for generating analytic time signals, complex FFT of frequency shifted signals for generating of modulation functions or down sampled analytic time signals.

Hereby the following demands have to be met concerning the transfer function of the critical band filters:

- Cut-off-frequencies and bandwidths according to Zwicker [52] (appendix B),
- filter slopes corresponding to psychoacoustical measured masking functions (low frequency slope 30..100 dB/Oct, high frequency slope up to 300 dB/Oct), simulated by corresponding exponential functions,
- short impulse responses (implemented with lengths below 30 ms).

7.1.2 Generation of the analytic Time Signal

The Cocktail-Party-Processor-Algorithms (chapter.5 and 6) are based on the analysis of the analytic time signal:

- Analytic and real time signal are two description methods for the same signal context (the real time signal is the real part of the analytic time signal).
- For sampled analytic time signals the amplitudes for the frequencies $½f_{abt}<f<f_{abt}$ are equal zero. Data reduction for bandpass filtered signals is possible via frequency transformation and reduction of the sampling rate.
- Amplitude and phase of harmonic oscillations are provided independently as amplitude and phase of the analytic time signal. Signal analysis methods, which evaluate mainly the envelopes of the source signals, are eased at this (amplitude generation or conjugated complex multiplication for phase and carrier elimination).

Since the frequency transform of the analytic time signal matches to the corresponding real time signal in the range $0<f<½f_{abt}$ and becomes 0 in the range $½f_{abt}<f<f_{abt}$, the analytic time signal can be evaluated from the real time signal easily. The processed real time signal can be obtained by extraction of the real part (for analytic time signal with the same sampling rate) or with the help of Fourier-Transformation-Methods (see chapter 7.2).

7.1.3 Data Reduction

For band pass signals, which are limited to the frequency range $f_{min}..f_{max}$ ($\underline{A}(f)=0$ for $f<f_{min}$, $f>f_{max}$), the analytic time signal $\underline{a}(t)$ and the Fourier-Transform $\underline{A}(f)$ result to::

$$\underline{a}(t) = \int_{f_{min}}^{f_{max}} \underline{A}(f)\, e^{j2\pi ft}\, df$$

$$\underline{a}(t) = e^{j2\pi f_{min} t} \int_{0}^{f_{max}-f_{min}} \underline{G}(f)\, e^{j2\pi ft}\, df \qquad \text{with} \quad \underline{G}(f) = \underline{A}(f+f_{min})$$

$$\underline{a}(t) = e^{j2\pi f_{min} t}\, \underline{g}(t) \tag{7.1.3/1}$$

$\underline{a}(t)$ corresponds to the modulation of a complex carrier $e^{j2\pi f_{min} t}$ with an analytic time signal $\underline{g}(t)$, which is limited to the frequency range $0..f_{max}-f_{min}$. If the frequency f_{min} is known, the function $\underline{a}(t)$ can be described by the function $\underline{g}(t)$ without any loss of information,.

For a sampled function $a(it_{abt})$ the following sum term result (f_{abt}= sampling frequency, t_{abt}= sampling period, N=Length of the Fourier-Transformation):

$$\underline{a}(it_{abt}) = \sum_{n=n_{min}}^{n_{max}} \underline{A}(nf_0)\, e^{j2\pi i\, n/N}$$

Where: $n_{max}f_0 \geq f_{max}$; $n_{min}f_0 \leq f_{min}$;$f_0=f_{abt}/N$;

$$\underline{a}(it_{abt}) = e^{j2\pi n_{min} i/N}\, \pi/2 \sum_{n=0}^{n_{max}-n_{min}} \underline{G}(nf_0)\, e^{j2\pi i\, n/N} \qquad \text{with} \quad \underline{G}(nf_0) = \underline{A}((n+n_{min})f_0)$$

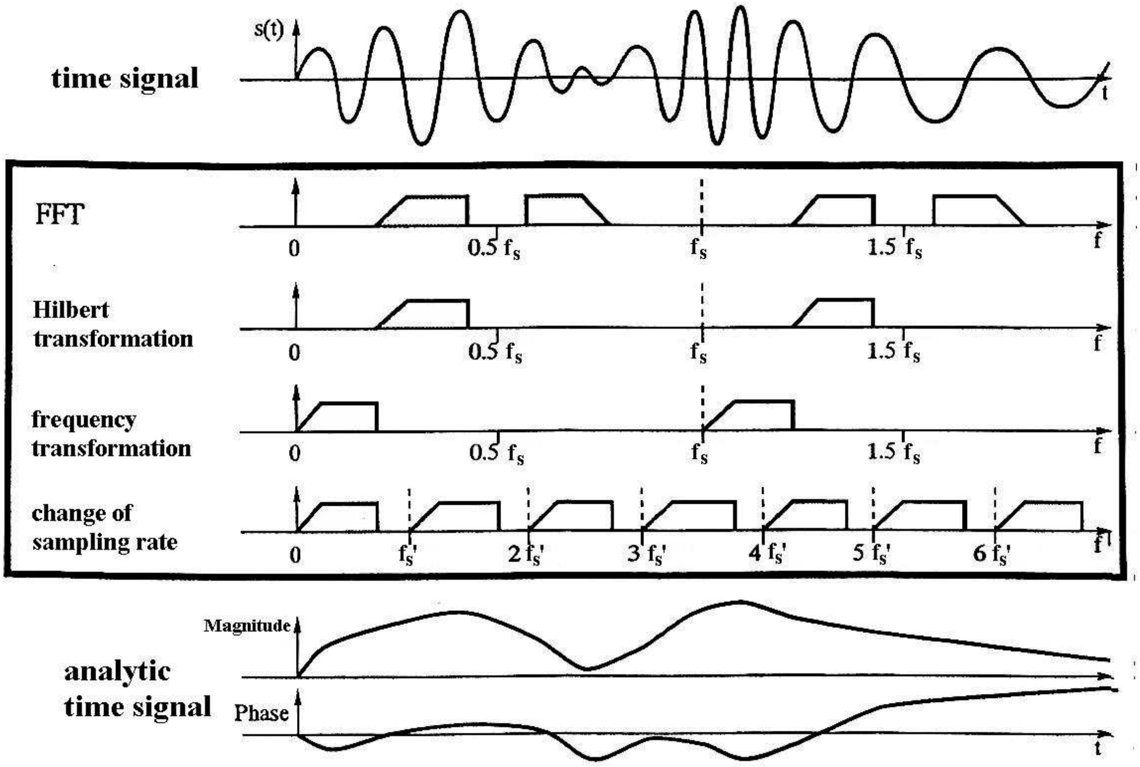

Fig. 7.2: Transformation of the input data

$$\underline{a}(it_{abt}) = e^{j2\pi n_{min} i/N}\, \underline{g}(it_{abt}) \qquad (7.1.3/2)$$

For a full description of band pass filtered signals only the knowledge of the lowest frequency and of the complex modulator function $\underline{g}(it_{abt})$ is necessary. The modulator function $\underline{g}(it_{abt})$ can be obtained by transforming the filtered time signal $\underline{a}(it_{abt})$

$$\underline{g}(it_{abt}) = e^{-j2\pi n_{min} i/N}\, \underline{a}(it_{abt})$$

The function $\underline{g}(it_{abt})$ is limited to the frequency range $0..(n_{max}\text{-}n_{min})f_{0.}$. According to the sampling theorem the information content is ensured, when the following sampling frequency is provided:

$$f_{abtg} = 2(n_{max}\text{-}n_{min})f_0 = N_g/N\ f_{abt}$$

$$t_{abtg} = N/N_g\ t_{abt} \qquad ;\ N_g{=}2(n_{max}\text{-}n_{min})$$

If the length of the Fourier-Transform N is a whole-numbered multiple of N_g, a down sampled version $\underline{g}'(kt_{abtg})$ can be constructed from $\underline{g}(it_{abt})$ by using only each (N_g/N))-th sample. Then the function $\underline{a}$ can be described completely by $\underline{g}'$, n_{min} and t

Using a bandpass filter bank with bandpasses of infinite steepness for critical band filtering the over all data rate of all filtered signals would stay constant when applying this down sampling method. Using critical band filters from appendix D and allowing an interference level of -80 dB the resulting over all data rate will be two-times up to three times the data rate of the original signal. Using conventional digital filters for 24 critical band filters with a sampling rate of 40 kHz and 16 Bit resolution (80 kB/s) the resulting over-all data rate would be about 2000 kB/s. If the sampling rate of the critical band signals would be reduced, according to the highest appearing frequency, the over-all data rate would decrease to about. 600 kB/s. Applying the down sampling method as described above data rates of about 200 kB/s can be reached with the same filter characteristics.

After processing the signal $\underline{g}'(k\,t_{abtg})$ the sampling rate has to be converted again. The signal must be transformed into the original frequency range by multiplying it with the carrier $e^{j2\pi n_{min} i/N}$. The entire processed signal can be obtained by adding all of these bandpass signals (see Fig. 7.2).

Apart from that , this method for data reduction by carrier elimination is quite similar to the methodology of the Fourier-Transform. In order to obtain amplitude and phase of a certain frequency f, the Fourier-Transform shifts the signal frequency that way, that the frequency f will become the frequency 0 Hz (demodulation). Amplitude and phase result from an integration over these that way achieved constant values, whereby alternating components, which originate from other frequencies, are averaged out.

$$\underline{A}(f) = \underbrace{\int_t}_{\substack{\text{averaging}\\ \text{at } f'=0}} \quad \underbrace{a(t)}_{\text{signal}} \quad \underbrace{e^{j2\pi ft}}_{\substack{\text{demodulation:}\\ \text{transformation to } f'=0}} \quad dt$$

7.2. Processing of the Output Signals

7.2.1 Requirements to a Re-Synthesis-Unit

A re-synthesis unit has the following functions:

- generation of processed ear signals based on the estimators of the Cocktail-Party-Processors
- distribution of the processed ear signals to the needed number of output channels
- combination of critical band ear signals to processed broadband signals.

Fig. 7.3 shows the block diagram of a re-synthesis unit

The following requirements shall be fulfilled by it:

- The method should be as fast as possible.
- The signal distortions should be as low as possible, especially non-linear distortions should not appear.
- The binaural information of the input signals should be preserved.
- Depending on the kind of continuation of the analysis different types of resulting signals should be available as real time signals, analytic time signals or frequency transformed analytic signals.

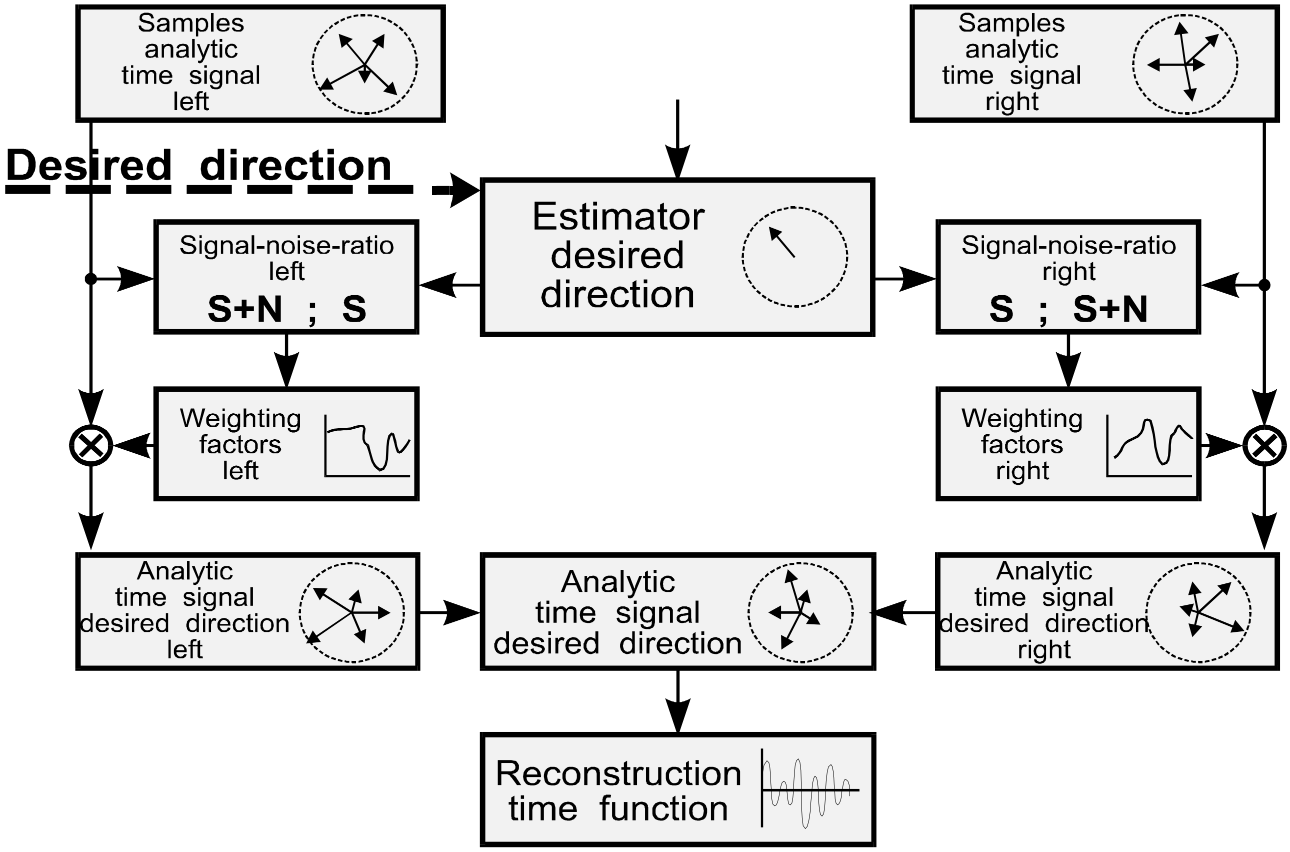
Samples analytic time signal left
Samples analytic time signal right
Desired direction
Estimator desired direction
Signal-noise-ratio left
S+N ; S
Signal-noise-ratio right
S ; S+N
Weighting factors left
Weighting factors right
Analytic time signal desired direction left
Analytic time signal desired direction
Analytic time signal desired direction right
Reconstruction time function

Fig. 7.3: Generation of time signals from the estimators of a Cocktail-Party-Processor

7.2.2 Alignment of the Input Signals to the Signal Estimators

Generation of Estimators for the Ear Signals

Output of Cocktail-Party-Processors are estimators for a "center of the head" (chapter 4.1) related signal of the desired direction. The needed estimators for the ear signals can be evaluated from this with the help of the corresponding free field outer ear transfer functions. These ear signal estimators $a_r'(t)$,$a_l'(t)$ describe for e certain time window the amplitude of the desired direction signal at the ears.

The re-synthesis unit shall modify the analytic time signals of the ear signals, which are available at the input of the model, in such a way, that inside the time window $2T_\mu$, where the ear signal estimators are valid, the power of the ear signals corresponds to the estimated power of the desired direction.

Generation of Weighting Factors

Hereto for each time window weighting factors $g_l(t)$, $g_r(t)$ are evaluated, which correspond to the quotient between estimated and existing ear signal power. By multiplying the input signals with these weighting factors, the input signals shall be aligned to the estimators (Fig. 7.4). Design and characteristics of this method correspond to the Wiener-Filter-Algorithm, which has been presented by Bodden/ Gaik [9]

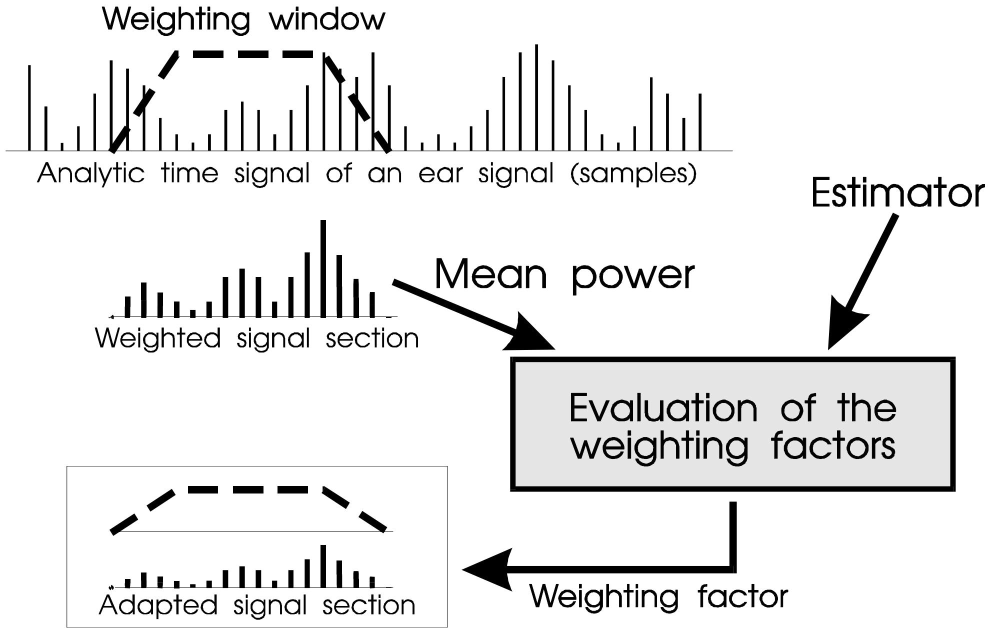

Fig.7.4: Evaluation of weighting factors from the mean power of a signal section and the estimated values

$$g_l(t)^2 = \frac{a_l'(t)^2}{\int_{T_G} \underline{l}(t_G)^2 \, dt_G} \qquad g_r(t)^2 = \frac{a_r'(t)^2}{\int_{T_G} \underline{r}(t_G)^2 \, dt_G}$$

For these time dependent weighting factors the following requirements have to be fulfilled:

- The weightings factors must not be bigger than 1. An estimated signal power, which is bigger than the power of the ear signals is a sign of estimation errors.
- In order to avoid non-linear distortions the change rate of the weighting factors has to be limited. For signals with a limited bandwidth the maximal change rate of the envelope may not exceed the maximal slope of an oscillation with a bandwidth corresponding frequency.
- In order to transmit the signals inside the critical band completely the time window for generating the weighting factors must be bigger than the period of the lowest frequency of this critical band.

As a consequence, a post-processing of the weighting factors is necessary, e.g.:limiting to a maximal value of 1 and averaging over several subsequent estimation intervals. In order to avoid, that individual estimation errors have an influence on the processed data, estimations and weightings should be performed with overlapping time windows.

The method is applied as follows (Fig. 7.4): A trapezoid or triangle formed window function is assigned to each estimator with the estimated weighting factor as the maximal value. The length of the trapezoid window (measuring point: 50% of the maximum) corresponds to the estimation interval $2T_{\mu}$. Attack and decay are dimensioned, that a sufficiently smooth fade over from the previous and to the subsequent estimator is achieved. If possible, estimations are performed with overlapping time intervals. By overlaying and averaging of the weighting factors a corresponding weighting function is generated.

The resulting weighting functions adapt well to fast signal modifications (maximal delay = 2 window sizes). They describe the signals very reliable as well, caused by a big number of separate computations (for 90% overlap the weighting factors of 10 estimations are averaged). Additional signal distortions are avoided (click free fade over due to smooth trapezoid slopes).

By multiplying with these weighting functions the input signals are adjusted to the estimated values.

Characteristics of the Re-Synthesis-Algorithm

This method for adjusting filtered input data is in principle an enhancement of a method for suppression of interfering speakers with a known signal-to-noise-ratio, which has been applied by Bodden/Gaik [9].

When using data reduction algorithms (chapter 7.1.3) the sampling rate of the weighted signals is reduced. By transposing these signals inside the frequency domain to higher frequencies (time synchronous complex multiplication with the initially removed carrier signal) signals with the original sampling rate can be generated. From these signals broadband time signals can be generated again by merging the signals from all critical bands.

This re-synthesis method has the following characteristics::

- By using the original input signals the fine structure of the ear signals survives.
- The processing of analytic time signals allows a relatively fast synthesis for data reduced signals.
- The main signal processing step is the engraving of the estimated signal power into the input data (Wiener-Filter-algorithm, see above). This method is very flexible. When using appropriate estimation methods also phase- or frequency-estimators could be engraved into the analytic time signals by a simple complex multiplication..
- By limiting the modification rate of the weighting factors and by using a sufficiently long weighting time nonlinear distortions can be avoided and the total frequency range of the critical band can be transferred.
- By modulating the input signals rather slowly with the weighting factors the fine structure of signals, the signal spectrum and the signals phases survive. This is especially important for binaural speech processing.

With this method the original mixture of sound signals is transferred with the power of the desired signal. This approach is quite similar to the vocoder technique, where an intelligible transmission of speech can be achieved by transmitting only the signal envelope in critical bands and only basic information about the carrier signals needs to be transferred (tonal signal with a corresponding pitch or noise). If, like here, the original signals are used as carrier signals, the transmitted carriers match to the desired signal for dominant sound sources (positive signal-to-noise-ratio). For low desired signals and negative signal-to-noise-ratios the carrier signal of interfering sound sources is used, which impairs the sound of the processed signal but affects speech intelligibility less.

Since the signal fine structure is transmitted unchanged, the Cocktail-Party-Processor-System can be used as a pre processing unit for other analysis tasks. In principle the system is also suitable for the preprocessing of microphone array signals. Since the signal phases remain unchanged and since there are no transit time deformations when using the critical band filtering as described above, the output signals of the system can be processed by array techniques in order to achieve additional signal-to-noise-ratio improvements for the desired direction. Cocktail-Party-Processor technique and array technique are useful complements for each other, because for low frequencies a Phase-Difference-Cocktail-Party-Processor with small dimensions achieves good results, but linear arrays must become very big, to achieve a good directional efficiency. A further possibility would be to pre-process dummy head recordings by Cocktail-Party-Processors, in order to achieve additional signal-to-noise-ratio improvements for binaural recordings and to concentrate the recordings to signals of distinct directions.

7.2.3 Reducing the Number of Output Channels

If less output channels than input channels are needed, the following methods can be used;

- *Combination of output channels:*
  The signals of the output channels have to be filtered with the corresponding inverse free field transfer function, in order to compensate time delays and level differences for the desired direction. When these transformed signals without phase or level differences are added up, the signals of the desired direction are amplified while signals of other directions are less amplified or attenuated. This additional benefit for the desired direction is not very big, if there are bigger interaural level differences, this method yields more benefit for stereo microphone arrangements or microphone arrays. A relatively exact estimation of the desired direction is essential, because otherwise the desired signal could possibly be attenuated at higher frequencies (comb filter effect).

- *Transferring of output channels with maximal signal-to-noise-ratio*
  The weighting factors for the channels represent the estimated signal-to-noise-ratios. From it the channels with the maximal can be selected. Since the signal-to-noise-ratio can be maximal at different channels in different critical bands, the interaural differences have to be compensated, too, before the channels of different critical bands are combined to one resulting signal. This method is very robust, false estimations for the input direction and the signal power have not much influence on the processing of the desired signal.

7.2.4 Generating the Time Function from Samples of the analytic Time Signal

As a result there are analytic time signals inside critical bands with a reduced data rate and with a power, which has been adapted to the desired signal. If a correction of phases or time has been carried out, this has to be taken into account at the following processing steps by correction the transformation time.

The real time signal can be generated with the help of an inverse method to the procedure of Fig. 7.2: For data reduced critical band signals $\underline{g}(i\, t_{abtg})$ there is only complex information about the envelope within the frequency range $0..(f_{max}\text{-}f_{min})$. Through the transformation into frequency domain a signal with a periodic spectrum results with a period of $1/(f_{max}\text{-}f_{min})$. The data rate is increased to the original values by generating a new frequency function, which corresponds to the spectrum above in the frequency range $0..(f_{max}\text{-}f_{min})$ and is zero in the frequency range $(f_{max}\text{-}f_{min})..f_{abt}$. By shifting the frequency range $0..(f_{max}\text{-}f_{min})$ by f_{min} the data are transformed to the original frequency range, resulting into the spectrum of the analytic time signal.

The real time function can be generated from the spectrum of the analytic time signal with the help of the real inverse Fourier-Transformation (mirroring the spectrum at $\frac{1}{2}f_{abt}$). The processed real time signals inside critical bands can be combined to a processed broadband signal by a simple addition.

7.3 Overall Presentation of the Cocktail-Party-Processor Model

7.3.1 Model Structure

Fig. 7.5 shows for one critical band an overview about the processing stages of the proposed Cocktail-Party-Processor model. (Total composition of the model in Fig. 7.1). The kernel of the Cocktail-Party-Processor is build by the Phase-Difference-Cocktail-Party-Processor (chapter 5) for

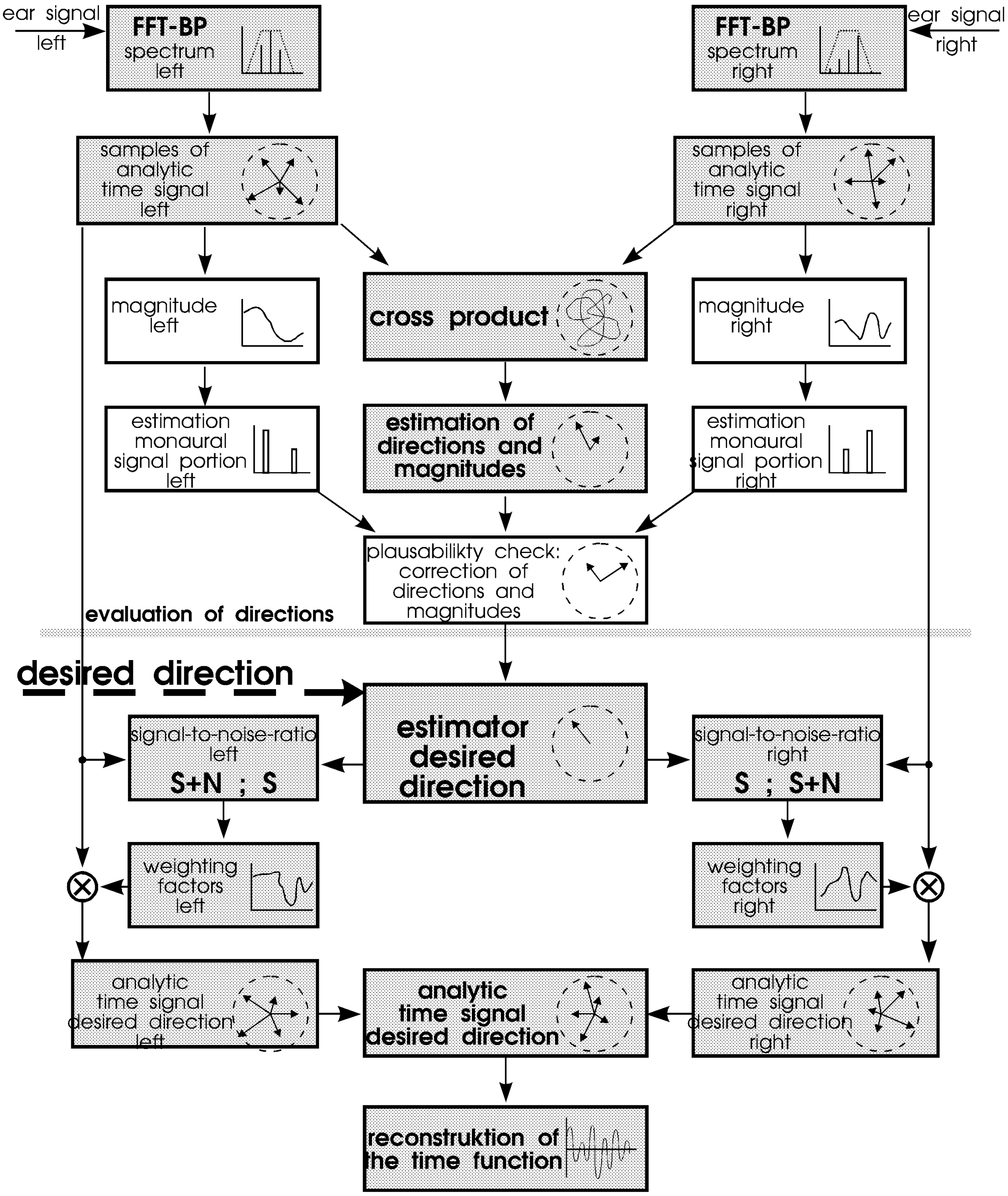

Fig. 7.5: Overall presentation of the binaural processor.
gray background: Modules of the phase difference processor
white: Additional modules of the level difference processor

analyzing the interaural cross product and by the Level-Difference-Cocktail-Party-Processor (chapter 6), which analysis the amplitudes of the ear signals. The combination of the resulting estimators takes place within a joining stage according to chapter 6.4. Although the algorithms are able to estimate input directions of sound sources (see Fig.5.7), it has to be specified from outside or from corresponding model (chapter 8), which of the possible directions shall be used as the desired direction for processing. After a desired direction has been specified the estimators can be mapped onto this direction and be corrected correspondingly (chapter 5.6 and 6.3). Pre- and post-processing of the signals are performed according to this chapter.

7.3.2 Capabilities of the Cocktail-Party-Processors

Test Conditions

The capabilities of the model shall be demonstrated at an example. Speech signals of a male and of a female speaker (2.7 s continuous text) have been recorded in the free field (anechoic chamber) mapped on different directions by using simplified free field outer ear transfer function (appendix C) and mixed together to ear signals with different interaural parameters for each speaker. In this simulation the signals of the female speaker were presented without interaural differences, the signals of the male speaker were presented from the right with an interaural time difference of 400 µs and corresponding interaural level differences (according to appendix C). Desired direction was the direction of the female speaker. Tests have been performed with 3 different signal-to-noise-ratios 0 dB, -10 dB and -20 dB. The specified signal-to-noise-ratios are related to the total energy of the 2.7 Seconds lasting speech signals.

Taking as an example the Phase-Difference-Cocktail-Party-Processor, the capabilities of the developed algorithms shall be demonstrated. The following system parameters have been used for the test:

- Slight data reduction when generating of the analytic time signals of the ear signals (one sample per period)
- Time constant for the evaluation of the statistical parameters of the interaural cross product: 20 ms,
- Evaluation of new estimators: once per 1 ms,
- Averaging time for estimator generation: 20 ms,
- Width of the directional lobe (lock-in-range): ±0.1π (corresponds to ±18°),
- Correction method for estimators differing from the desired direction: "Valid Estimator Range"(chapter 5.6),
- Window type for adapting the analytic time signals of the ear signals to weighting factors: Triangle,
- Method for reducing the output channels: Selection of the channel with the biggest signal-to-noise-ratio (here: left ear signal).

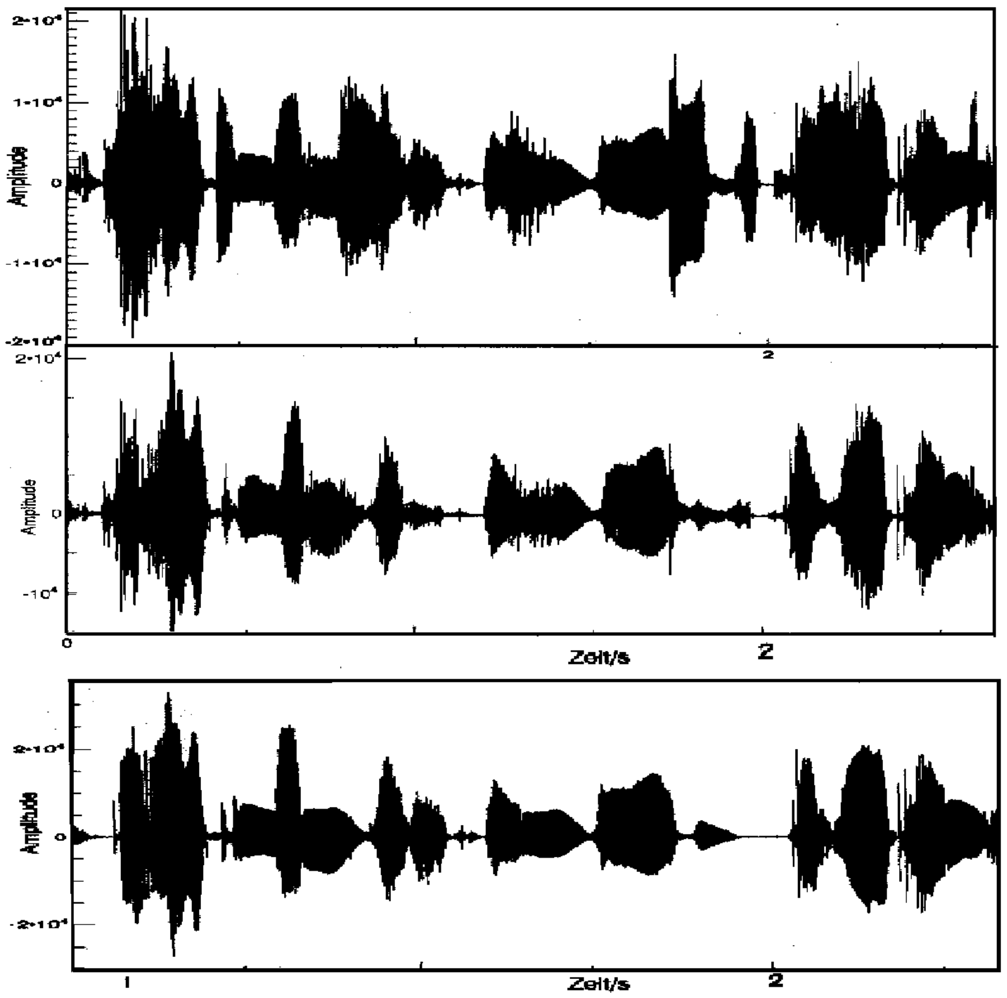

Fig. 7.6: Processing results of the Phase-Difference-Cocktail-Party-Processor for 2 speakers in the free field; 2.7 s continuous text
interaural time difference of the direction of the interfering speaker: 400 µs
interaural time difference of the direction of the desired speaker: 0 µs
signal-to-noise-ratio of the desired signal: 0 dB
top: left ear signal
middle: processed signal
bottom: undisturbed desired signal

Results

Fig.7.6, Fig.7.7 and Fig.7.8 show the results of the direction selective processing (processed signals) together with the left ear signal, which is used by the re-synthesis unit for generating the processed signal, and the undisturbed desired signal. Desired direction was the direction of the female speaker (τ=0).

Fig.7.6 shows the results of the direction selective processing if desired and interfering signal have the same level. Although the ear signals show strong influences of the interfering speaker (upper figure), these influences are nearly completely eliminated in the processed signal (middle figure). The processed signals corresponds nearly to the desired signal (lower figure).

The acoustical quality of the processed signal is judged as rather high. The loudness of the interfering speaker is reduced drastically.

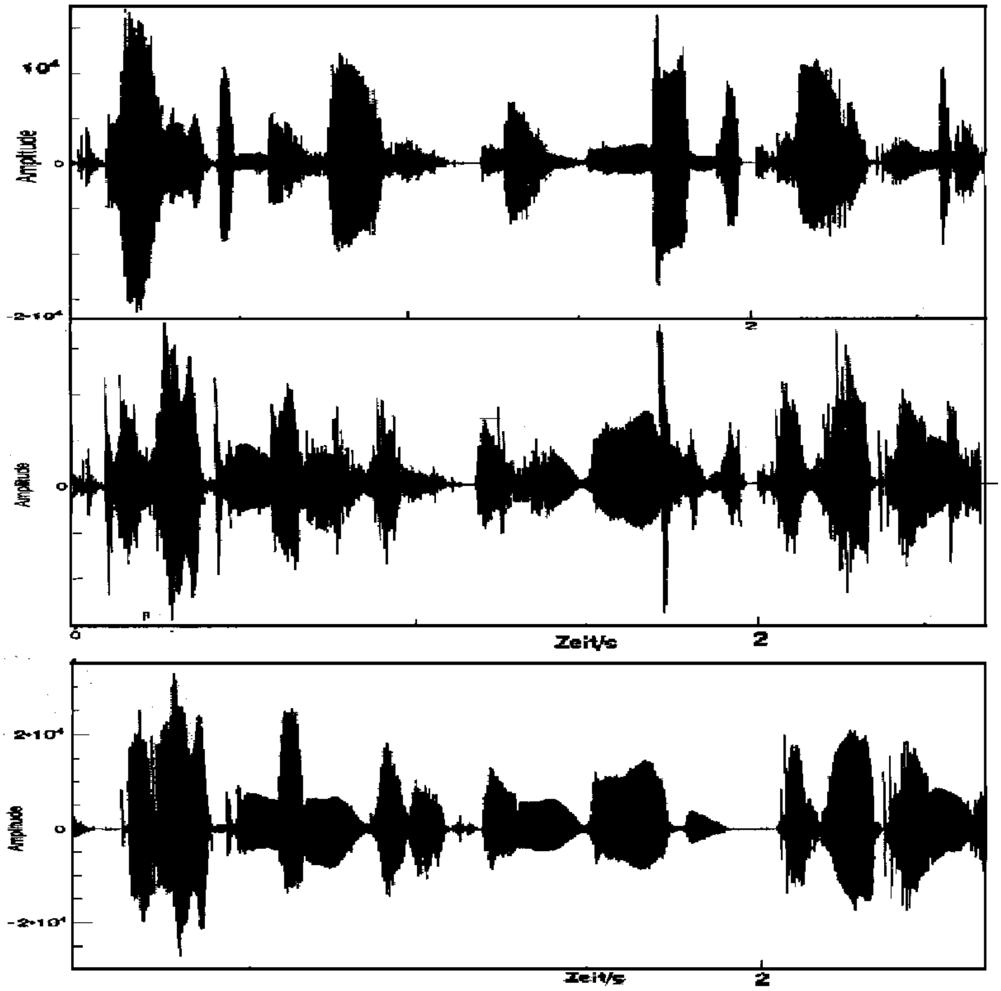

Fig. 7.7: Processing results of the Phase-Difference-Cocktail-Party-Processor for 2 speakers in the free field; 2.7 s continuous text signal-to-noise-ratio of the desired signal: -10 dB other conditions, see Fig.7.6 and text

top: left ear signal
middle: processed signal
bottom: undisturbed desired signal

If the signal-to-noise-ratio for the desired signal is negative, like in the situation of Fig.7.7, the interfering signal prevails in the ear signals (upper figure). By the direction selective processing of the Cocktail-Party-Processor the interfering signal with a 10 dB higher level, can still be suppressed. When comparing the processed ear signal (middle figure) with the ear signal of the undisturbed desired signal (lower figure), the processed signal corresponds mainly to the desired signal, even though there are already some errors at certain points.

The acoustical quality of the processed signal remains quite high The speaker of the desired direction is perceived louder than the interfering speaker.

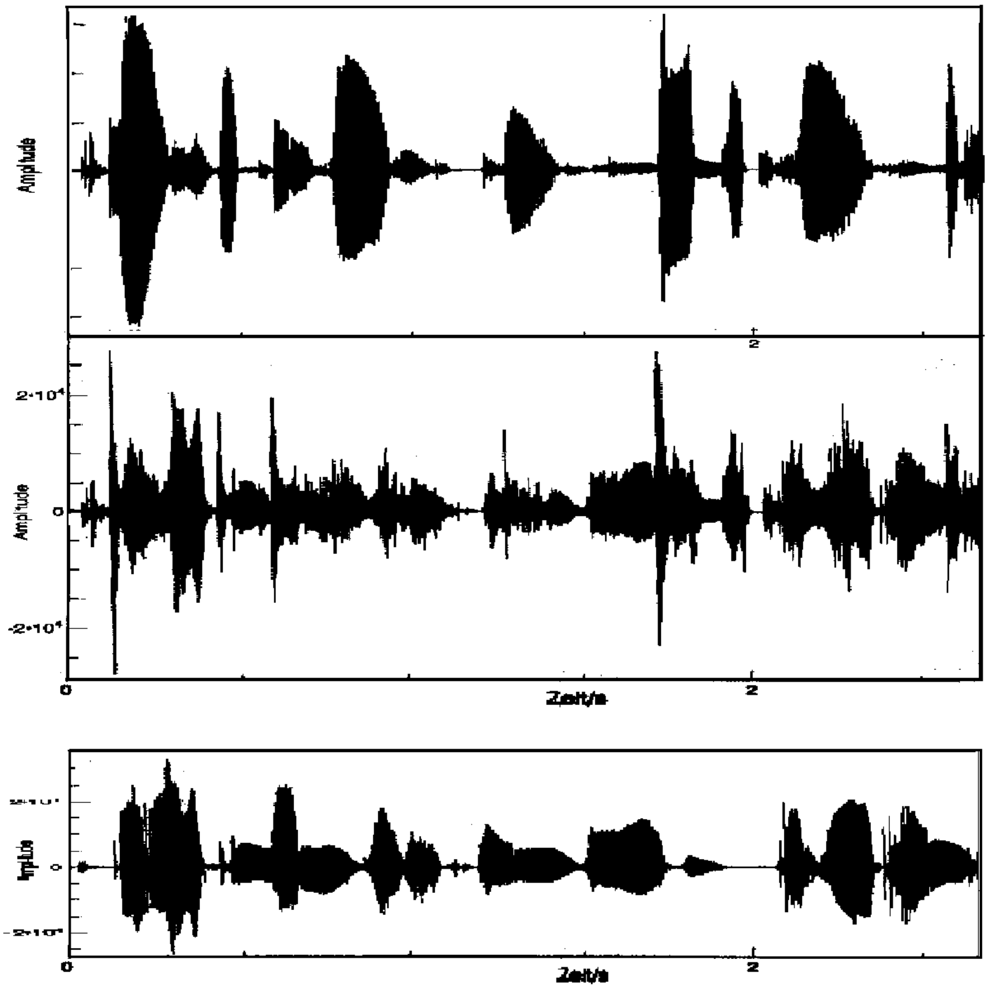

Fig. 7.8: Processing results of the Phase-Difference-Cocktail-Party-Processor for 2 speakers in the free field; 2.7 s continuous text signal-to-noise-ratio of the desired signal: -20 dB other conditions, see Fig.7.6 and text

top: left ear signal
middle: processed signal
bottom: undisturbed desired signal

For very low signal-to-noise-ratios of -20 dB, like in Fig.7.8, the desired signal is nearly unrecognizable in the ear signals (upper figure). Nevertheless, the Cocktail-Party-Processor attenuates the interfering signal thus far, that the structure of the desired signal (lower figure) can be recognized from the processed signal (middle figure). The structure of the processed signal is a little bit similar to the structure of the unprocessed left ear signal for a signal-to-noise-ratio of 0 dB.

When presenting the unprocessed ear signals acoustically the perceived loudness of the desired speaker is very low. Within the processed signal the desired speaker has nearly the same loudness than the interfering speaker.

For further reduced signal-to-noise-ratios the interfering source prevails in the processed signals, too. At a signal-to-noise-ratio of -30 dB the desired signal is no longer detectable in an acoustical presentation. But after processing by the Cocktail-Party-Processor the desired signal can be perceived acoustically again.

Survey

The examples above show, that the used algorithms are able to improve the signal-to-noise-ratio for a desired direction even under unfavorable conditions.

The reason for the possibility, to process signals with a very low signal-to-noise-ratio, is mainly founded in the fact, that the algorithms base on a 2-source-model, that therefore all sound field parameters are interpreted as the result of the interfering of two sound sources or two frequency lines. Therefore small variations of the sound field parameters of a dominant sound source, which are caused by a weak interfering sound source, can still be evaluated in terms of two sound sources..

For processing of these about 2.7 seconds long signals in 24 critical bands an array processor (Stardent Titan) needed about 7 minutes computing time (real time factor≈150) in a not computing time optimized test version, whereas the signal processing inside the 5 lowest critical bands was performed in real time each. By optimizing the algorithms (deactivating of test routines, and graphic outputs, introduction of function tables, external critical band filtering, stronger data reduction, especially in the higher critical bands) real time realizations might be possible, at least for each critical band. Especially by applying data reduction methods inside of critical bands, as discussed in chapter 7.1.3 (sampling of the complex modulation function instead of the original signals) a reduction of computing time by the factor 3 would be possible without a loss of information. The model is designed in such a way, that each critical band can be processed independently of each other and that inside each critical band the processing can be distributed onto maximal 14 independent processes, which all can be executed on independent processors (see appendix F). Providing a corresponding hardware effort, a real time setup of the model would be possible, when using corresponding signal processor or transputer boards and a computing time optimized model setup.

The desired direction for the Cocktail-Party-Processors has to be specified externally yet.. One possible further development task could be, to introduce methods, which select the desired direction of the Cocktail-Party-Processors automatically. In the following chapter possibilities shall be discussed to construct a processing unit for the directional control of the Cocktail-Party-Processors, which is adopted to the directional control mechanisms of the human auditory system.

8. Control of the Cocktail-Party-Processor

The capabilities of the human auditory system in perceiving directions shall form the basis for controlling the desired direction of Cocktail-Party-Processors. In this context the perception of directions in complex sound fields and dynamical effects of directional perception are of special interest. The criteria of the auditory system for selecting the direction of attention can act as an prototype for a directional control unit of Cocktail-Party-Processors, which detects the desired direction automatically and keeps tracking it.

8.1. Detection Criteria for directional Information:
Hearing in enclosed Rooms (Franssen-Effect)

Inside of enclosed rooms localization and directional processing of sound signals is quite difficult for the human binaural system. If the listener is sufficiently far outside the reverberation radius of all sound sources, the human binaural system is, according to the experiments of Franssen [16], no longer able, to localize stationary signals correctly and process them direction selectively (Franssen-Effect).

> Inside a room (auditorium) there are 2 loudspeakers at different positions. At the beginning of the presentation loudspeaker 1 emits a noise signal with a steep attacking slope. Subsequently the power of this loudspeaker remains constant. The listeners can localize this loudspeaker easily. During the stationary part of the envelope the signal is very smoothly faded over from loudspeaker 1 to loudspeaker 2. Although loudspeaker 2 emits all the sound at the end, the listener's auditory events remain at the position of loudspeaker 1. This (mis-)localization remains, even if the test supervisor plugs off the cables of loudspeaker 1 demonstratively.

This leads to the following conclusions about the signal processing of the human auditory system in reverberant environment for sound sources outside the reverberation radius:

- The human auditory system is not able, to localize stationary signals in situations like this (otherwise loudspeaker 2 would have been localized).
- The human auditory system is not able, to process stationary signals direction selectively in situations like this (otherwise the loudness of the existing auditory event would have been attenuated after fading over).
- But the human auditory system can very well localize the corresponding sound source during fast signal changes or at signal onsets (correct localization of loudspeaker 1 at the beginning of the experiment).

In complex sound fields, like in big enclosed rooms, there is only for a few moments directional information available, which can be interpreted by the auditory system. If in all frequency bands the power of the reflections is bigger than the power of the direct sound of the sound source, then the auditory system can neither determine the direction and the signal characteristics of the sound source, nor process the direct sound of this source direction selectively.

For localization and direction selective processing the auditory system is therefore dependent on -at least- short time frames, in which the direct sound of one sound source prevails and is evaluable. Possibly these time frames have also to be used, to evaluate signal-to-noise-ratios or to evaluate the power of the echoes and reverberation, respectively. These time frames are characterized by attacking slopes of the envelope - at least in some frequency bands. Wolf [49] has used these properties to build up a localization system for sound sources in enclosed rooms.

If the direct sound prevails, the variance (related to the displacement) of the interaural cross correlation function is small. The signals can be localized. If reverberations prevail, the variance increases. Based upon this Allen/Berkley/Blauert [1] constructed a de-reverberation system, which attenuates or suppresses by variance-controlled weighting factors those signal periods, where the interaural variance is big and therefore reverberations dominate.

During attacking signal slopes with prevailing direct sound the Cocktail-Party-Processors of the signal processing model above deliver estimators with the interaural parameters of the direct sound with only a small directional variance. (estimators at dominant sources, chapter 5.5 and 6.2.5). Therefore the variance of the directional estimation can be taken as a detection criterion for time frames with evaluable directional information.

This detection criteria corresponds to applying the slope method of Wolf [49] onto Cocktail-Party-Processor-Models. Applied to Cocktail-Party-Processors improvements in detection could possibly be expected, since those attacking slopes, which are caused by the interference of several sources with different instantaneous frequencies and which come along with bigger directional variances, would not be detected (for example interfering of mirror sound sources of signals with variant short time spectrum and different time delays)..

Compared with the pure variance method improvements could be expected, too, because a couple of signals with variant interaural cross correlation functions will now become interpretable. For example interaural beats, which appear at the interference of two signals with different instantaneous frequencies, would now be interpreted as two sound sources with invariant directional estimators. The methods of the variance analysis are, however, taken over for the weighting of the estimators. Variant estimators with a broad weighting function (compare chapters 5.6 and 6.3) give an indication, that these estimators are the result of the interfering of reflections and reverberation, which should be ignored for the directional analysis.

With the help of the Cocktail-Party-Processor method signal onsets (attacking slopes) of the direct sound can be tracked even into the area of early reflections.. As long as only one early reflection is present, the sound situation is a pure 2-source-problem, which can be exactly solved for both sources with the help of the Cocktail-Party-Processor-Algorithms: If the total power of all reflections is smaller than the power of the direct sound, the direct sound can be evaluated as dominant source deeply into the "reflection hill".

8.2. Dynamical Effects of Direction Detection: The Precedence-Effect

The Precedence-Effect describes the perception of directions when presenting signals from different directions in a close temporal sequence., for example direct sound from one direction and a reflection from another direction.

The perception depends on the temporal sequence of "direct sound" and "reflection". If the time between two directional information is less than the maximal interaural time difference (<1ms), sum localization appears, and an "averaged" direction is perceived. (see chapter 4.2). If the time between direct sound and reflection is bigger than the echo threshold, the directions of both signals are perceived. The echo threshold depends on the spectrum and the dynamical characteristics of the signals. In the zone between sum localization and echo threshold "normally" only the direction of the direct sound is perceived ("law of the first wave front"). The spectral characteristics of the "reflection"-signals are mapped onto the direction of the direct sound.

Investigations of Cliffton [11], Wolf [49], Blauert/Col [8] give the result, that the Precedence-Effect does nor always follow the "law of the first wave front", but depends on the used signals, their temporal sequence and their history. The same signal configuration of direct sound and reflection can one time lead to the perception of one input direction ("law of the first wave front"), but the other time after a certain history (direction of the reflection is previous direct sound direction) lead to the perception of both direction. This means:

- The human auditory system is in principle able, to perceive in signal configurations of the Precedence-Effect (fast sequence of two different directional information) the directions of the direct sound as well as the direction of the reflection and to process the both related signals direction selectively.

Since the manifestation of the Precedence-Effect depends on the history and since the essential signals can date back some seconds, the Precedence-Effect seems to be the result of interventions of "higher perception layers" onto the determination of the desired direction and seems therefore to be a consequence of controlling the direction of attention. The Precedence-Effect can therefore give insight in the controlling mechanisms of the auditory system and act as a model for the control of Cocktail-Party-Processors.

The Precedence-Effect could be interpreted as follows: The auditory system determines a direction, to which attention shall be paid and to which the auditory system internal Cocktail-Party-Processor is oriented. As long as the binaural system cannot deliver a reliable directional information, the present desired direction is kept. If a reliable estimation of the input direction is available (for example at attacking slopes of signal onsets) the control unit must decide, whether the detected direction shall be taken as a new desired direction or not. Experiments concerning the Precedence-Effect can so give insight into the decision criteria of the binaural system for taking over a direction as the new direction of attention.

8.3. Description of the Precedence-Effect by a binaural Cocktail-Party-Processor-Model

"Law of the first Wave Front"

In experiments concerning the Precedence-Effect for example clicks of some milliseconds duration are presented from different directions with different delays, in order to investigate the influence of direct sound and reflections onto the perception.

Before the first reflection arrives the sound field is only determined by the direct sound. Input direction and signal characteristics can be determined easily. The interaural cross product (or the interaural cross correlation function) show only a small directional variance. The direction of the direct sound can be kept in mind as a reliable localized direction.

When the first reflection arrives, the signals of direct sound and reflection interfere inside each related critical band. The variance of the cross product increases. But the Cocktail-Party-Processor-

Algorithm as described above, is able to determine the power and input directions of direct sound and reflection.

If there was no reliable directional information before starting the presentation, the directional control unit takes over the analyzed direction of the direct sound period as the new desired direction. Since the system is now oriented onto the direct sound direction, the direction of the reflection is now suppressed as a unwanted direction. Due to the bigger variance of the cross product the directional information of the reflection is now classified as unreliable, and a new orientation of the system will not happen. The result is the "law of the first wave front".

Exceptions of the "Law of the first Wave Front"

During the presentation of the direct sound at the beginning of the Precedence-Effect signals there is always a time frame with low interaural variance and therefore a surely localized auditory event, which could lead in principle to a new orientation of the system..

If identical sequences of direct sound and reflection are presented, like at the experiments to the exceptions of the "Law of the first Wave Front", the information about the desired direction seems to consolidate. If then a signal pair is presented, where the directions of direct sound and reflection are swapped, both directions are perceived, the direction of the new reflection as the former desired direction and the new direction of the direct sound as an additional reliably localized direction as well.

If the new combinations of direct sound and reflection are repeated, 2 auditory event directions are perceived until the system control takes over the new direct sound direction as the new desired direction. After that the directional information of the reflection is suppressed again and the result corresponds again to the "Law of the first Wave Front".

Blauert/Col [8] repeated experiments to this effect by swapping the directions of direct sound and reflection continuously. At the first swaps of the directions of direct sound and reflection are perceived both. But after some swaps the perception went to the "Law of the first Wave Front" again.

The time constants for directional control seem to adapt to the sound field situation. At the beginning of the experiment, described above, each reliably localized direct sound direction is taken over as new desired direction, after a swap the direction of the reflection is also perceived as a signal from the old desired direction. When continuously changing the direct sound direction the auditory system seems to shorten its "memory times" and change its strategy to "continuously new orientation". As a consequence, the directions of the reflections are no longer perceived.

Echo Threshold

If direct sound and reflections do no longer interfere in particular critical bands, the variance is low, when the reflection appears, and the direction of the reflection can be considered as a reliable input direction.

Inter-Relationships between Critical Bands

According to investigations of Blauert/Divenyi [6] the Precedence-Effect also acts across the borders of critical bands. According to Blauert et,al. [7] direct sound in the frequency range around 1 kHz can suppress the directional information of reflections in other frequency ranges, but the reciprocal influence can not be observed. For other frequency ranges the impacts across critical band ranges are reduced..

The frequency ranges for direct sound and reflections do mostly not overlap at these investigations, there must have been invariant directional information in the frequency range of the direct sound and in the frequency range of the reflections, so that in principle for both directions a new orientation of the direction of attention might be possible.

Reliable directional information is therefore weighted frequency dependently. They only lead to a new-orientation of the system, if there is no contradictory information from other more important frequency ranges.

8.4. Consequences for controlling for Cocktail-Party-Processors

The input signals for a control unit for Cocktail-Party-Processors are the estimators for input directions and signal power, weighted by a probability function.

The most important task for a control unit is the detection of time periods with a low variance of the interaural cross product (direct sound detection). Signal directions and signal power, which have been detected during this time period, should be forwarded to upper model layers.

If there are several low variance estimators (e.g. from different critical bands) the control unit has to decide, which direction shall be taken over as the desired direction of the system. Criteria for this decision can be: Reliability of direction detection, importance of the related frequency range, history (frequently detected directions are rated as very reliable information) as well as the support of certain directions by other information sources.(for example optical information, intentional decisions).

When a new input direction is taken over as a new desired direction, the variance of this input direction is used to build up threshold criteria for accepting new input directions. A desired direction is kept or enforced, if it is confirmed by additional reliable directional information (for example by low variance estimators from several critical bands). If there is no confirmation for a certain time period, then the acceptance threshold decreases and the direction of other reliable information, for example low-variance-estimators for other input directions, can be taken over as the new desired direction.

The time period, for which a desired direction decision remains valid, can be modified signal dependently. If a detected input direction is considered as important, but changes frequently, the system can also switch over to "permanent new orientation".

After the decision for a desired direction has been made, detected directions, which do not correspond to the desired direction, are not forwarded and ambiguous analysis results are examined, whether they could possibly contain portions of the desired direction (see chapter 5.6 and 6.4). But, nevertheless, strong low-variance signals can be forwarded to upper layers, even if they do not match to the desired direction.

The frequency dependence of the Precedence-Effect can be incorporated by a frequency dependent weighting of the estimators from different critical bands..

8.5. From Processor Control to a Precedence Processor

Information about the power of the direct sound and of the background noise can be evaluated during attacking slopes or during time periods with invariant interaural cross product estimators. These estimators can be considered as a reference for evaluating signal and noise levels during time periods, where, caused by reflections and reverberation, no reliable estimation is possible.

As so called Precedence-Processor, which records sound source and noise parameters during attacking slopes and extrapolates this information far into the reverberation field (quasi as a pre-adjustment for time periods with unreliable directional estimation), such an algorithm could be introduced as third Cocktail-Party-Processor type besides Phase-Difference-Cocktail-Party-Processor and level difference Cocktail-Party-Processor.

9. Conclusion and Perspective

Central issue of the here present paper is the investigation of algorithms for a direction selective processing of sound sources.

The basis for the design of signal processing algorithms have been psychoacoustical experiments, which have given information about the signal processing of the human auditory system at the presence of multiple sound sources. The question of these experiments has been, up to which grade of similarity signals can be processed direction specifically. The investigation have been carried out on the one hand in the frequency range below 800 Hz, where the auditory system determines the direction by evaluating interaural phases, and on the other hand in the frequency range above 1.6 kHz, where the detection of directions is based on the analysis of interaural group delays and of interaural level differences.

In the frequency range below 800 Hz test persons could determine the input directions of two sound sources correctly, even for relatively low signal differences, for example for sinus signals of 500 Hz and of 530 Hz, and for example for two independent noise signals with 7% relative bandwidth and 500 Hz center frequency. For these signals a majority of the signal power is concentrated in one critical band. The test persons were able, to determine the relative pitches of the localized sources (higher/lower than other sources) and to assign a corresponding loudness to the localized signals as well. But the sound of the sound sources not be determined correctly. This was only possible, if the (center) frequency difference exceeded a critical band width.

In the frequency range above 2 kHz two sound sources with different directions could only be localized correctly for frequency differences above a critical band width. Then a correct assignment of pitch, loudness and sound to the localized direction was possible, too.

Assuming, that the critical band is the smallest analysis band, which is used by the auditory system for evaluating directions, then a directional selective analysis without "Cocktail-Party-Processor-Mechanisms" would only be possible, if the signals of different directions are located in different critical bands. This means, that the auditory system would use Cocktail-Party-Processor-Mechanisms in the low frequency range, where the auditory system evaluates interaural phase delays, but not in the high frequency range, where interaural group delays are evaluated. Therefore the characteristics of this auditory system internal processor would be as follows: evaluation of interaural phases, determination of input directions and loudness of (at least) 2 sound sources in parallel, no directional selective analysis of signal sounds, this implies, that there is no direction selective separation of phase or spectral information.

Technical models, which shall reproduce the properties of the auditory system, must therefore include similar functionality than the proposed auditory system internal Cocktail-Party-Processor. Functions, which would fulfill these requirements, are for example: the interaural cross correlation function (analysis of real signals) and the interaural cross product (analysis of complex analytic time signals). For both functions Cocktail-Party-Processor-Algorithms are presented, which can determine input directions and signal power for two interfering sound sources from the ear signals. The cross correlation function uses a method in the frequency domain, the cross product in the time domain.

Major issue of the here presented paper is the description of a Cocktail-Party-Processor, which is based on the analysis of the interaural cross product. This method can be described mathematically easily, can be computed rather fast, and reacts quickly on signal changes. With the help of the interaural cross product interaural beats are analyzed, which arise, when signals with different spectra interfere. The statistical parameters of the interaural cross product are used, to determine power and input direction of two sound sources, which can generate such kind of beats.

The specialty of this binaural signal processing method, the Phase-Difference-Cocktail-Party-Processor is, that a 2-source-approach is used for modeling the binaural interactions. Input direction and power of two sources can be analyzed simultaneously and the parameters of one source can even then be evaluated, if an interfering source with much bigger power is present. Using this method, improvements of the signal-to-noise-ratio of up to 20 dB can be achieved, and even for speech signals with signal-to-noise-ratios of -30 dB audible improvements of the signal-to-noise-ratio can be reached.

For complex sound fields (more than 3 sound sources, reflections and reverberation) improvements can be achieved by this method, if the power of the desired source exceeds the power of other sources for dedicated time periods or in dedicated frequency ranges or if the desired source is at least the second strongest sound source (dominant source). By using additional estimation algorithms it is possible, however, even for weak desired sources and complex sound fields, to suppress other individual dominant sources, to estimate the possible power of the desired source and to track it, But the estimation errors grow here. Inside such complex sound fields also the human auditory system is no longer able, the localize weak sound sources correctly and process them direction selectively. (see reduced BILD in reverberant environment, Franssen-Effect).

The used Cocktail-Party-Processor-Algorithm is also able, to process single channel signals and interpret it as the interference of two signals with different spectrum. This property is utilized for constructing the Level-Difference-Cocktail-Party-Processor. Here the envelopes of both ear signals are analyzed separately, and estimators for signal power and input direction of two sound sources are evaluated by an appropriate interaural combination of these monaural estimators.

The properties of this processor are quite similar to the Phase-Difference-Cocktail-Party-Processor: Power and input direction of two stationary sound sources can be estimated even at negative signal-to-noise-ratios. The parameters of dominant sources in complex sound fields can be determined. Even here post-processing-methods can be found, which allow to estimate the possible power of weak desired sound sources in complex sound fields.

For broadband analyses at the natural ear distances both Cocktail-Party-Processors have to be combined. The Phase-Difference-Cocktail-Party-Processor produces very accurate results for low frequencies, where the interaural phase is unambiguous. The Level-Difference-Cocktail-Party-Processor is especially suitable for analyzing higher frequencies, where sufficient level differences appear.. Within the medium frequency range the Level-Difference-Cocktail-Party-Processor can be used to correct the errors of the Phase-Difference-Cocktail-Party-Processor, as there are ambiguous directional estimations. The results of both algorithms are combined in a corresponding processor unit, which selects from ambiguous estimators the most probable ones. At this point also the results from other information sources can be included into directional analysis and signal processing (for example optical information, previous knowledge, Precedence-Processor).

Psychoacoustical findings could be used to develop an auditory system adapted strategy for controlling Cocktail-Party-Processors. In analogy to the auditory experiments of Gaik [21] discrepancies between Phase-Difference-Cocktail-Party-Processor and Level-Difference-Cocktail-Party-Processor could be solved by generating monaural estimators (for example by splitting contradictory estimators).

Findings about the Precedence-Effect, which describes the dynamical behavior of the auditory system for detecting directions, could be applied for controlling the desired direction of the Cocktail-Party-Processor. Hereby the control strategy of the auditory system could be reproduced by technical means.

For processing complex sound fields with reflections and reverberation a Precedence-Processor could be constructed, which collects information about the desired sound source and the interfering sound field during time periods with reliable directional information (direct sound time periods) and which takes this information as a basis for the processing of time periods without reliable directional information.

It might be possible to enhance the acoustical analysis by integrating additional information sources like optical information (Detection of directions and of lip movements) and by including knowledge about the sound source characteristics (for example about possible signal power and signal spectra <male/female voice, sound characteristics>).

The presented Cocktail-Party-Processor-Algorithms process analytic time signals relatively fast, and can also process data reduced signals.. With some further improvements of the algorithms, real time realizations might be possible. By applying appropriate pre- and post-processing methods (non-causal critical band filter without runtime distortions, smooth modulation of the signals by the resulting estimators of the processors) processed signals of relatively high quality can be produced.. These signals can be used as input signals for further signal processing methods.

The signal processing framework of the Cocktail-Party-Processors represents an adaptive direction selective filter, whose transfer function is continuously adapted to the spectrum of the signals of a certain input direction.

Since there are no runtime distortions and since the signal phases are not changed, the algorithms could also be used for improving the directional selectivity of microphone arrays. For this all output channels of the Cocktail-Party-Processor have to get the same (direction specific) damping, in order to keep the level ratios between the receivers of the array unchanged. The directional selectivity could be improved by the processors especially in the low frequency range. Applying the Cocktail-Party-Processor in combination with two directional microphones the directional selectivity could be further improved especially in the low frequency range.

This algorithms could also be applied as a pre-processing unit for dummy head systems. The Cocktail-Party-Processor could evaluated the spectra of the interfering sources and eliminate the disturbing spectral portions without restricting the binaural analysis possibilities. This means, that there would be furthermore the possibility to listen into the remaining signals, quite similar than to unprocessed signals.

Instead of a dummy head arbitrary objects could be equipped with microphones, for example telephone chassises. Precondition for applying these processors would be the knowledge of the transfer functions between the microphones and the free sound field, analogous to the free field outer ear transfer functions of the head. By using multi microphone arrangements or by constructing properly adapted housings optimal conditions for the processors can be formed, like unambiguous phase relationships for the Phase-Difference-Cocktail-Party-Processor or smooth transfer functions for the Level-Difference-Cocktail-Party-Processor.

With an appropriate microphones placement the Phase-Difference-Cocktail-Party-Processor could also be used for broadband processing. For this purpose the microphone placement has to ensure unambiguous phase relationships for the whole frequency range (for example microphone distances of 20 cm, 5 cm and 1 cm). With such a kind of mini-array nearly for the whole audible frequency range narrow directional lobes can be achieved.

Possible applications for such Cocktail-Party-Processor methods could be use cases, where a directional selective signal processing is necessary or requested, especially, where a desired signal in disturbed by signals of other directions, for .example:

- as receiver unit for speech systems (speech recognition systems, hands free telephones, recording systems in rooms), for example on microphone equipped objects, in order to exclude interfering speakers and noise from being transferred.

- as hearing aid algorithm for noise suppression for clients with only one functional ear ("artificial binaural system"), for example hearing-aid spectacles with a mini-array and Phase-Difference-Cocktail-Party-Processor in order to enhance the, elsewhere heavily reduced, communication possibilities in disturbed environments.

- as acoustical signal analysis system (noise source localization, direction selective processing of noise sources, acoustical quality assessment of rooms/concert halls), for example as a pre-processing system for microphone arrays or dummy heads, in order to diagnose noise sources direction selectively and to judge certain sound sources subjectively.

- for simulating the human acoustical perception (investigation of disturbance/annoyance of certain spatial sources, simulation of binaural transmission systems for example for audio applications), in order to collect and realize the effects of direction selective information on affected persons.

- as binaural model for reproducing the human acoustical signal processing, for explaining psychoacoustical phenomena, in order to obtain new modeling opportunities for multiple-sources- phenomena and to provide a new basis to utilize further auditory system capabilities for technical applications.