1. Introduction

It is still a major problem in acoustics and communications engineering to analyze acoustical situations, where more than one sound source is present or where reflections and reverberation appear.

Oftentimes it is requested, to extract the information of only one sound source out of a mixture of multiple sound sources and to suppress the information of all interfering sound sources.

For speech processing systems, like speech recognition systems or hands-free telephones in an open-space office, oftentimes the problem appears, to extract only the speaker, who is communicating with the system, and to eliminate all other interfering speakers and also reflections and reverberation from being transferred.

For noise measurements tasks it is oftentimes requested, to localize the direction of individual noise sources and extract only their specific part of the total noise out of a multitude of noise sources, in order to act specifically on certain noise components. (Genuit [21]).

Hearing impaired people (e.g. people with only one functioning ear), have oftentimes more difficulties than normal hearing people, to concentrate in a conversation ("Cocktail-Party") on a conversational partner, because they are much more disturbed by other speakers and other sound sources than normal hearing people. Hearing aids, which could compensate this deficit, would be very helpful.

At the situations above oftentimes multiple types of disturbances appear simultaneously: Disturbances by other sound sources as well as disturbances by reflections and reverberation. A solution of these problems could be, to extract selectively the information of sound sources of a certain input direction.

A number of acoustical and communications engineering methods have already been developed to solve this problem, but for some of these problems no satisfying solution has been found up to now.

At the array technique multiple microphones are distributed linear or spatially. By an appropriate processing of the microphone signals the reception characteristics of the array can be modified, so that ideally the sound of sound sources of a certain input direction can be extracted specifically .In order to achieve a high directional selectivity the dimensions of the array have to be bigger than the biggest considered wavelength. In order to extract directions unambiguously, the microphone distances have to be smaller than the smallest considered wavelength. The direction specific processing of the whole range of audibility from 20 Hz to 20000 Hz, with wavelengths from 17 mm up to 17 m, will result into problems concerning the dimensioning of the array. Linear arrays are more suitable for directional selective filtering of bandwidth limited signals. But even for the direction selective processing of the telephone frequency range from 300 Hz up to 3400 Hz linear arrays have to be more than 1 m large, using microphone distances of less than 10 cm.

Unidirectional microphones are based on similar mechanisms than microphone arrays (directional filtering by phase extinction for interfering directions). Therefore the size of these microphones has to be quite similar than the size of microphone arrays. Manageable unidirectional microphones achieve an acceptable directional selectivity for frequencies above some 100 Hz, the directional selectivity increases with increasing frequency. The selected direction can only be changed by moving the microphone. Simultaneous or retroactive extraction of additional input directions (as requested for measurement tasks) is not possible by this kind of technique.

The adaptive filter technique separates a desired signal from interfering signals due to different spectral characteristics of the signals. Prerequisite is, that the signal characteristics of the interfering signal or of the desired signal have to be known (i.e. by placing a microphone near to the interfering sound source) or can be estimated from the recorded signals. By filters, whose transfer functions are dynamically adjusted to the signal-to-noise-ratio, disturbing spectral fractions can be eliminated from the resulting signal. The efficiency of adaptive filters is reduced, if the signal characteristics of desired and interfering signals are unknown or if the signal characteristics change rapidly or if the signal characteristics of desired and interfering signal don't differ significantly (i.e. two speakers).

By combining different techniques (i.e. array technique for the higher frequency range, adaptive filters for the lower frequency range) the properties of the techniques can also be combined and the results can be improved.

The human auditory system with only 2 ears has the ability, to enhance nearly across the whole range of audibility desired signals versus signals from interfering directions and to eliminate disturbing echoes and reverberations, and to do this with a very high signal transmission quality. Up to now technical systems for interfering signal cancellation do not support such a big bandwidth and do not reach such a high signal processing quality as the human auditory system.

Therefore there is also interest in investigating the human auditory system from a technical point of view, in order to collect information about the "algorithms" of the auditory system, which allow a signal processing with such a high quality.

The decoding of the methods of the human auditory system takes place on several layers:

Physiological acoustics collects information about the functionality of all the components of the auditory system. This includes, for example, the investigation of mechanical, electrical, physical and biological characteristics of outer ear, middle ear and inner ear, the analysis of nerve potentials at the auditory nerve and at involved parts of the brain. Objects of investigation are, for example, the conversion of acoustical stimuli into nerve excitations, the transfer of information through the auditory system, neural processing steps inside the auditory system. This shall help to understand, how the processing steps of the auditory system (e.g. within the inner ear) process a presented sound information and how this processing may influence the content of the transferred information. As a result physiological models are developed, which shall describe as exact as possible the operations inside the auditory system and the signal processing, which is performed there.

Psychoacoustics investigates the effect of sound on the human perception. In auditory experiments it is investigated, which sound information is perceived by test persons. The neural processing of the sound signals is mainly considered as a "black box". The interest is focussed on the results of the auditory system's processing. For these investigations oftentimes simple, easy reproducible and easy describable signals are used (Sinus, white noise, periodical signals, clicks). Hereby the specific reaction of the auditory system on single influence parameters is measured, in order to explain at the end also the perception of complex sound situations by combining many of these single influence parameters. Resulting psychoacoustical models describe as exact as possible the reaction of the auditory system and the human perception on sound stimuli, quasi as a kind of "transfer function" between sound and perception. In conjunction with physiological results and the resulting lower layer processing of sound signals conclusions can be drawn about the processing in upper layers of the auditory system.

Signal processing tries to apply psychoacoustically found effects for technical purposes (Detection of input directions, de-reverberation, direction selective filtering). Target of this onset is, to find out, which kind of signal processing could be the basis of the psychoacoustical perception, in order to develop algorithms for technical signal processing from it. Result of this development are signal processing models, which can processes signals in a similar (or better) quality than the human auditory system, but which are adapted to technical systems. It is not the goal, to reproduce the processing methods of the auditory system in detail, but to develop techniques, which lead to similar good results.

Common to all research approaches is the methodology. Experiments give an insight into the characteristics and capabilities of the auditory system. Then the experimental results are analyzed in order to discover regularities and trends in them and to deduce generalized conclusions about the capabilities of the auditory system. From these findings concepts for possible processing methods can be derived, which could explain the experimental results. From them a model of the auditory system could be formed.. Feeding auditory system and model with the same input signals, the model should reproduce the auditory processing of the sound signals and simulate the results of the auditory experiments. Technical signal processing might leave off the last step, the description of the auditory system, and turn to applying of the model conception on other, not necessarily auditory system related, technical use cases.

Similar principles form also the basis for the present paper:

After a short overview about psychoacoustical findings, which forms the framework for the following modeling (chapter 2), auditory experiments are presented, which investigate the perceptions of test persons, if multiple sound sources are present. The results of these experiments allow to draw conclusions about possible signal processing methods of the auditory system (chapter 3).

In chapter 4 the principles of the following model building are presented: the description of the transfer function between sound source and auditory system, the nomenclature and the mathematical description of the processed signals as well as the characteristics of the ear signals and the characteristics of interaural differences in the presence of multiple sound sources. Out of a comparison of the findings from the auditory experiments with the results of existing binaural models the necessity arises, to introduce signal processing methods, which go beyond a simple evaluation of interaural differences. For cross correlation models an algorithm in the frequency domain is presented, which allows the direction specific separation of two sound sources with similar spectrum, as required from the results of the auditory experiments. When changing from the interaural cross correlation function in the frequency domain to the interaural cross product in the time domain the performance of this method can be improved and time delays can be reduced significantly.

The interaural cross product is the basis of the Cocktail-Party-Processor algorithms, which are described in the chapters 5 and 6.

The "Phase-Difference-Cocktail-Party-Processor" (chapter 5) evaluates from the interaural cross product the input directions and signal magnitudes of the two involved sound sources. The properties of this processor are presented, as well as its behavior in complex sound fields (more than

2 simultaneous sound sources, diffuse sound field). In addition algorithms are described, how to estimate the sound field magnitude for a certain input direction.

The "Level-Difference-Cocktail-Party-Processor" (chapter 6) evaluates by using a similar method than the Phase-Difference-Cocktail-Party-Processor the interaural level differences and the magnitudes of two sound sources from the ear signal magnitudes and the interaural level differences. The properties of this processor are discussed, even in complex sound fields, and algorithms are presented, to estimate the sound field magnitude for a certain input direction. Level- and Phase-Difference-Processor can be combined to a common analysis unit, which can correct inaccuracies and ambiguities of each of the processors by combining their results accordingly.

In chapter 7 the signal processing framework for the Cocktail-Party-Processors is presented, and signal processing results are discussed. The signal processing framework includes the filtering and preprocessing of the input signals for the binaural analysis, as well as the post-processing of the processing results for getting direction filtered broadband signals. In this context possibilities for data reduction and for saving of computational time are discussed, too. These binaural pre- and post-processing-steps represent a periphery interface, which can also be used as for other signal processing purposes.

The control of the Cocktail-Party-Processors can be realized by a so called "Precedence-Effect-Processor" (chapter 8), which can, similar to the precedence-effect of the human auditory system, determine input directions even in complex sound fields and provide the desired analysis direction for the processors. Using this kind of processor Cocktail-Party-Processors would be able to orientate themselves autonomously inside closed rooms and search for sound source input directions and would not need an external preset of a "hearing direction". For further model extensions this central control unit could be developed towards an interface, which could also incorporate additional information sources (e.g. optical information.).

2. Psychoacoustical Background

It is the target of psychoacoustical research, besides others, to collect information about the signal processing of the human auditory system. Of major interest, from the perspective of technical applications, are especially the capabilities of the binaural system, to detect of input directions and to process signals of different input directions selectively.

2.1. Recognition of Sound Directions

Encoding of Directions by Head and Ear

Sound from a lateral input direction reaches the facing ear earlier than the distant ear. For natural ear distances of about 20 cm these time differences reach values of up to 600...700 µs. As long as the half wavelength of the sound waves is bigger than the ear distance (this is valid for frequencies below 800 Hz) the interaural phase differences is sufficient, to describe the input direction unambiguously. For higher frequencies the interaural group delay has to used for an unique description of input directions.

Sound waves are diffracted and shadowed by the head. So the sound level at the distant ear becomes lower than at the facing ear. Since diffraction effects are depending on the ratio between the dimensions of the head and the sound wavelength, these interaural level differences are frequency dependent. For low frequencies there are only interaural level differences of some dB (≤ 5 dB at 500 Hz), but for high frequencies they can grow up to more than 40 dB.

There are additional influences on the transfer function "sound field - eardrum" from reflections at the shoulders and at the upper body and from the structure of the pinna. Pinna and the entry of the ear canal act as an acoustical resonator system, whereas its properties are depending on the input direction of the sound. As a result, direction dependent maxima and minima are engraved into the transmitted sound spectra. However, these effects .are only appearing, if the sound wavelength reaches to the dimensions of the pinna or is lower than it.(i.e. for frequencies above 3 kHz) (Shaw/Teranishi [36]).

The influences of head and pinna on the sound signals at the eardrum are described by the free field outer ear transfer functions. It is defined as the quotient between the spectrum of the sound signals at the ear drum and the spectrum of the sound signal, which would appear at the same location without the influence of the head.

Perception of Directions

Under natural hearing conditions in the free field interaural time and level differences as well as the spectral characteristics of the ear signals are involved in the detection of directions and in the direction selective processing of the auditory system. It is a target of psychoacoustical investigations, to appreciate the importance of these parameters. For technical applications (Cocktail-Party-Processor etc.) it is of interest, which parameters are quasi worth to be evaluated from the perspective of the auditory system. An overview about corresponding results of psychoacoustical investigations can be found, for example, at Blauert [5]:

The perception of the azimuth angle of sound sources is mainly influenced by interaural time and level differences, the perception of the elevation angle is mainly based on the evaluation of the ear signal spectra. Inside "directional bands" (Blauert [3]) specific spectral minima and maxima of the sound pressure cause the perception of corresponding elevation angles.

Depending on the frequency range there are different physical influences on the ear signal parameters. This is reflected in the perception of sound signals, too.

For frequencies below 800 Hz there is a unique relationship between the angle of incidence and the interaural phase. Here the interaural phase difference is used for determining interaural time differences. Interaural level differences are rather low in this frequency range, and can only be evaluated with big inaccuracies. For example at a frequency of 500 Hz there is an interaural level difference of maximal 5 dB, with a lateralization blur of about 1 dB (Blauert [5], S.161). The spectrum of the outer ear transfer function is relatively plain, but it shows some slight maxima and minima.. Here a directional band for the front direction is located. (Blauert [3]).

For frequencies between 800 Hz and 1.6 kHz the evaluation of interaural time differences changes over from the evaluation of interaural phase differences to the evaluation of interaural time differences between the signal envelopes. Since the signal periods reach the size of the interaural time differences, a unique evaluation of input directions from the interaural phase differences is no longer possible. The relevance of interaural level differences grows in this frequency range ($\Delta L \leq$ 15 dB). The spectra of the outer ear transfer functions contain a directional band for the rear direction in this frequency range.

Above 1.6 kHz the evaluation of interaural time differences is based on the evaluation of interaural group delays. Also interaural level differences gain in importance. With growing frequency the free field outer ear transfer function becomes more and more jagged and direction dependent, and includes more and more substantial information about the elevation of the sound source, For example, there is a sharp maximum around 8 kHz, which indicates sound from above.

If all direction determining parameters match, the sound signal is perceived from the corresponding direction. If discrepancies appear between the parameters or if the information is incomplete, the localization can get faulty.. If the relationship between interaural time and level difference does not match to natural combinations, two auditory events are perceived or the auditory event becomes diffuse (Gaik [20]). If the received free field outer ear transfer functions do not match to the transfer functions of one's own ears, inside-the-head-location and wrong direction perception can appear (e.g. front-rear-permutation). (Fornefeld [15]).

Dynamic Properties of Direction Detection

Inside rooms, where reflections and reverberation appear, the human auditory system is able, to evaluate the directional cue of the direct sound, which is arriving first, and to determine the direction of the corresponding sound source, but also to ignore directional cues from reflections and reverberation, which arrive later. Psychoacoustically these dynamic effects of directional perception are subsumed under the term "Precedence-Effect".

If different directional cues appear subsequently, the auditory system reacts as follows (compare Blauert [5]): If the time differences between two directional cues are lying within the range of interaural time differences, sum localization can appear for coherent signals and an "averaged direction" is perceived. Under certain conditions for bigger time differences of up to 5..70 ms only the sound source direction is perceived, whose sound arrives first at the listener's ears.(former "law of the first wave front"). This can also apply for incoherent signals in different frequency ranges

(Blauert/Divenyi [6]). But the "law of the first wave front" can also fail for certain signal configurations (Clifton [11], Wolf [47], Blauert/Col [8]). For time differences above the echo threshold the directions of both signals are recognized.

Under certain conditions, according to Franssen [16], the auditory system is at the presence of reflections and reverberation only able to detect the directions of sound sources, if major changes in the signal characteristics appear (e.g. during ascending slopes of the loudness). For stationary signals direction detection might be impossible in this case.

Spectral Properties of Direction Detection

According to Scharf et.al.[35] directional cues are processed within critical bands. The bandwidths of these critical bands for binaural processing correspond to the bandwidths of the critical bands for monaural processing, as described by Zwicker et.al.[52] for the perception of loudness (see appendix B)

2.2. Processing of Signals of certain Directions

Investigations of the *BMLD* (Binaural Masking Level Difference) give information about the enhancements of the signal-to-noise-ratio, which are achieved by the auditory system's direction selective processing. Object of investigation is the capability in detecting a certain signal in the presence of a masker, where signal and masker once have identical interaural parameters and once have different interaural parameters. As interaural parameters oftentimes two different interaural phases are used (0 and π).

For different input directions / different interaural parameters the auditory system can detect within the masker test signals, whose level is up to 10..18 dB lower than in the case of identical. interaural parameters(Blauert [5]). The largest BMLD of 18 dB appears at the frequency range around 300 Hz, for higher frequencies the BMLD decreases to 0..3 dB above 5 kHz. In this frequency range the detection of the test signals seems mostly to happen, when minima appear in the masker's envelop (Langhans/Kohlrausch [24]).

Investigations on masking effects, like BMLD experiments, tell, whether the introduction of a test signal leads to modifications in the perceived sound, independently,. whether the informational contents of the test signals can be recognized. Investigations on the *BILD* (Binaural Intelligibility Level Difference) combine the methods of BMLD investigations with speech intelligibility tests.

If the interaural parameters of test signal and masker differ, so that direction selective processing becomes possible, the level of a speech signal in a disturbing environment can be reduced by up to 9 dB to reach the same speech intelligibility as in the case with at identical interaural parameters.

For hearing under free field conditions the BILD is mainly caused by shadowing interfering sound sources by the head. The *MILD* (Monaural Intelligibility Level Difference) can give information about the improvement of intelligibility, which the auditory system can achieve from shadowing an interfering sound source by the head. The MILD can reach values of up to 6 dB (vom Hövel/Platte [45]).

The term *Cocktail-Party-Effect* denotes the ability of the human auditory system, to process sound information direction specifically and therefore to extract the speech information of one speaker out of a mixture of speakers. The Cocktail-Party-Effect can be described quantitatively by the BILD.

The human binaural system has also the ability to suppress reflections and reverberation. As a result of binaural sound processing modulated signals with a substantial smaller modulation depth can be recognized than in the case of only monaural perception (Danilenko [12]).

The ability of direction selective processing of sound information decreases when the rate of reflections and reverberation is increasing. Therefore the BILD decreases drastically, if the listener resides outside the reverberation radius of a sound source, where the power of the direct sound gets lower than the power of the reflected sound (Plomp [32]). Franssen [16] showed, that inside a reflection dominated sound field a smooth fade over from one loudspeaker to an other one could not be recognized, neither the new sound direction nor the informational contents of the new direction could be recognized.

2.3. Properties and Performance of Binaural Signal Processing

A signal processing model, which is adapted to the processing of the human auditory system should therefore have the following properties:

- Processing of the signals within critical bands.
- Use of interaural time and level differences for determining the input directions and as the basis for direction selective processing as well. Determining elevation information from the analysis of the spectral structure of the ear signals.
- Comparing of found interaural time and level differences with natural combinations for a plausibility check of the auditory events (multiple auditory events).
- Possibility, to extract directions and signal parameters from short time frames, to be able to detect the parameters of the direct sound within reverberant environment.
- A "higher layer" control unit, which defines the desired direction of processing. (adaptation of the control unit to the results of the precedence effect).

The properties of the auditory system indicate, which potentials of signal processing are existing in principle. A technical signal processing model has to cope with these standards:

- Improvement in signal detection of up to 10-18 dB,
- Improvement in speech intelligibility of up to 6-9 dB,
- Reduction of reflections and reverberation.

However, at heavy echoes and reverberations the abilities of the human auditory system for direction selective processing are reduced, yet.

3. Auditory Experiments on the Localization of multiple Sound Sources

3.1. Setup of the Auditory Experiments

Background

If multiple sound sources from different directions are interfering, the human auditory system is able, to extract out of this mixture only the signals of one input direction. Reproduction quality, directional resolution and directional separation capabilities of this "human Cocktail-Party-Processor" are up to now superior to comparable technical systems with two receivers. The human auditory system demonstrates quite impressively, how powerful such systems could be. For constructing a technical system of similar top-level processing quality, one possibility is, to study the properties of the auditory system with the help of auditors experiments and to transfer hereby found signal processing methods into technical systems.

If psychoacoustical investigations are carried out under conditions, where the signal processing of the auditory system is not "perfect" and the auditory events no not match exactly to the signals of the sound sources, conclusions can be drawn to the signal processing methods of the auditory system, for instance by comparing the behavior under error conditions with the errors of technically realizable systems. On the other hand the limitations in the capabilities of the auditory system can give a hint to the needed quality if modeling for technical applications. Theses and models about hearing can be proofed by means of psychoacoustical results, in order to develop technical systems with hearing related properties.

Purpose of the Investigation, used Signals

Subject of the investigations is the ability of the human auditory system, to localize a single sound source within a mixture of sound sources and to extract their signals. The following questions arise in this context:

- How many auditory events are perceived, if several sound sources are emitting sound simultaneously? Under which conditions can the auditory system determine the correct number of sound sources?
- Which directions are perceived? Under which conditions can the auditory system localize the sound sources correctly?
- Which characteristics of the auditory events are perceived? Under which conditions can the auditory system determine characteristics of the sound sources correctly?

It is of interest, how similar signals of different directions may be, in order to be perceived as different sources by the auditory system. Furthermore there is the question, whether the abilities of the auditory system for localization and for signal processing are performed by coupled processes, which means for example, whether all characteristics of this sound sources can be determined correctly as soon as the sound source can be localized, and vice versa.

It is assumed, that the auditory system processes the ear signals for all kinds of analysis within critical bands (Zwicker et al. [52]; Scharf et al. [35]; Kohlrausch [23]).and that sound sources, which are located in different critical bands, can be separated easily (Blauert [4]). In this content it shall be investigated, how the auditory system reacts, if the combined bandwidth of all test signals does not exceed the bandwidth of a critical band. For this reason narrow banded test signals have been chosen (sinus signals, narrow band noise with 7% relative bandwidth).

Furthermore it has to be considered, that the evaluation of interaural parameters in different critical bands is based on different principles. For frequencies below 800 Hz the evaluation of interaural time differences is mainly based on the evaluation of the interaural phase. For frequencies above 1600 Hz it is mainly based on the evaluation of the time differences between the envelopes. As a consequence, the experiments are performed within two different frequency ranges; the range around 500 Hz and the range above 2 kHz. The used signals correspond to the different analysis methods of the auditory system. Signals without envelope patterns (sinus signals) and signals with a structured envelope (narrow band noise).

During the experiments 2 loudspeakers emit different signals simultaneously:

1) Signal 1: sinus 500 Hz; signal 2: sinus 500 Hz+Δf; Δf= 10..160 Hz; 8 test persons.
2) Signal 1: narrow band noise 500 Hz; signal 2: narrow band noise 500 Hz+Δf; Δf= 0..160 Hz; 8 test persons.
3) Signal 1: sinus 2 kHz; signal 2: sinus 2 kHz+Δf; Δf= 10..1200 Hz; 10 test persons.
4) Signal 1: narrow band noise 2 kHz; signal 2: narrow band noise 2 kHz+Δf; Δf= 0..1200 Hz; 10 test persons.

The experiments are performed in an anechoic chamber at sound pressures of 50 dB.

Experimental Setup

The experimental setup has to be designed in such a way, that only the effects under investigation have influence on the results. Therefore the auditory experiments have to meet the following requirements:

1. If only one sound source is active, only one auditory event shall appear. According to Gaik [20] multiple auditory events or diffuse auditory events can appear, if interaural time and level differences appear in unnatural combinations. In order to avoid effects like these, the signals are presented via loudspeakers inside an anechoic chamber.

2. The test persons shall only be able to evaluate acoustical cues. In order to avoid any optical information about the direction of the sound sources, the loudspeakers are placed behind a gauze curtain. (Fig. 3.1).

3. The ear signals shall be reproducible between different test persons, the acoustical parameters shall not change during a test series. Therefore there is a head rest installed at the test person's chair, and the test persons have to lean their head against the head rest. Through this head movements can be reduced, but nevertheless small movements (of some millimeters) cannot be eliminated.

4. Only signals in the stationary state shall be investigated. The test persons shall not be able to get information from transient effects, which would add additional localization cues and have lead to significantly improved localization results in pretests. The test signals are therefore faded in and out smoothly, being overlaid by an interfering signal during this time (pink noise + sum of all test signals, emitted by all installed loudspeakers, level of the interfering signal = 60 dB, 10 dB above the sound level in the stationary state, leading to sum localization for the non-existing direction 0^{o}).

5. The test persons shall not be influenced by prior knowledge about the experimental conditions. Used signals, number and position of the sound sources remain unknown to the test persons. The used speakers change after every test .

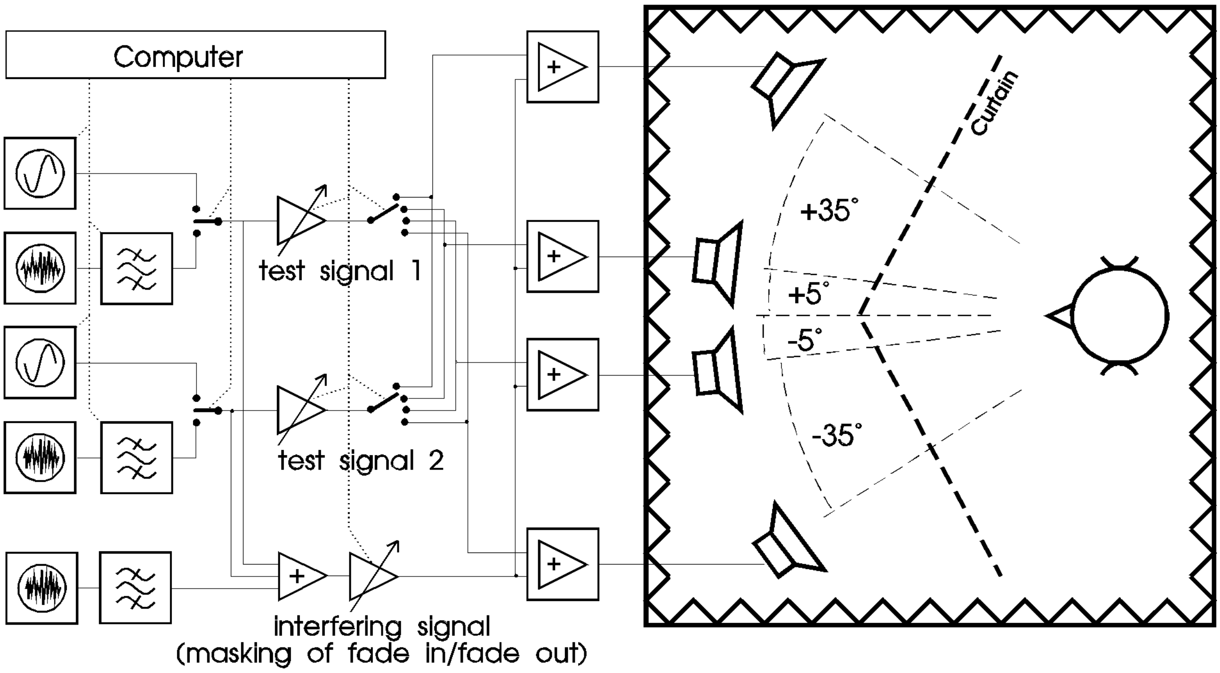

Fig. 3.1: Experimental setup and equipment

6. Localization "by chance" shall be avoided as far as possible. For test signals with small (center) frequency differences many test persons report about auditory events within the angle ranges of 50°..90° and -50°..-90°, independently, whether sound sources are located there or not (Slatky [39]). In order to prevent a corruption of the results by this effect, the loudspeakers are only placed within an input direction range of -35°..35°

7. Since the test persons are not used to the presented signals and hearing situations, there has been an initial training phase for each test person, where the test persons report about their auditory events and they were guided to analyze a complex hearing situation in terms of interfering of multiple sound sources.

8. The test persons shall be able to specify the directions of their auditory events as precise as possible. For this purpose numbers from -10 to +10 are placed on the curtain, characterizing the corresponding sound input direction.

9. The evaluation of the experiments shall base on psychoacoustical relevant parameters. Therefore the incidence angle of the sound sources Θ and the responses of the test persons are mapped onto the lateralization scale L_t (range of values : -10(left)..+10(right)) or onto the normalized interaural time difference τ_L.

$$L_t = 10 \sin(\Theta) \qquad \tau_L = 625\ \mu s\ \sin(\Theta) \qquad (3.1/1)$$

10. Single sound signals shall be localizable for test persons. During pretests nearly all test persons could determine the direction of a 500 Hz sinus signal with an accuracy of 10°. For signals of about 2500 Hz the hit rate decreased to 60% (signal from 20° to the right, permitted inaccuracy: ±15°).

Questions to the Test Persons

The test persons are asked to describe their auditory events as comprehensive as possible. Therefore they are asked, to describe each directionally distinct auditory event on a questionnaire in terms of direction, expansion, relative pitch, relative loudness (compared to other auditory events) and sound characteristics (single tone, mixture of tones, "harsh"; only for tests with sinus signals). From the analysis of these results it is expected, to get information about the signal processing methods, which are used by the human auditory system. This analysis is based upon the following assumptions concerning the signal processing mechanisms of the human auditory system.

- Identifying the number of sound sources is considered as a first step of directional analysis. The perceived number of auditory events can give information about it.
- Identifying the input directions of the sound sources is considered as the next step of directional analysis. The correct estimation of the sound direction is seen as a precondition for the direction selective processing of sound signals. The perceived directions and the width of the auditory events can give information about it.
- Identifying the characteristics of sound sources qualitatively is considered as the first step of direction selective processing. The percentage of correct assignments of the relative pitches to directions can give information about it.
- Identifying the attributes of the sound sources quantitatively is considered as a further step of direction selective signal processing, The assignment of the relative loudness of an auditory event of a certain direction can give information about it.
- The direction selective separation of sound signals is the result of direction selective signal processing, The perceived sound of the auditory events can give information, to which extend the characteristics of the sound signals can be evaluated by the auditory system direction selectively.

In this context not only the ability of the auditory system is of interest to detect several directions and signal attributes, but it is also of interest to get information about the interactions of the parameters and sequence of detection. This might give insight into possible signal processing steps of the auditory system, which can be used for modeling.

3.2. Results of the Auditory Experiments

Descriptions from the Test Persons

The test persons, even experienced ones, classified the auditory experiments as very difficult. Some test persons needed several minutes to analyze the presented sound situations. Some test persons reported, that position and characteristics of the auditory events changed after listening a longer time. Only the persons, who conducted the experiments, had less problems to analyze the sound situation. Since it could not be clarified, whether the reason was a bigger hearing experience or the knowledge about the configuration of the experiments (number of sound sources, possible positions), their results have been ignored.

All test persons perceived at sinus signals with very low frequency difference (f_1=500 Hz; f_2=501 Hz) auditory events moving across the room, similar to "binaural beats" at head phone experiments (Blauert [5]). The speed of the movement and the loudness depended on the perceived position of the auditory event. The characteristics of these auditory events was nearly independent from the real sound input directions. Up to frequency differences of 5-10 Hz the course of the auditory event could be tracked - albeit with some problems.

At low frequency differences (Δf=5-10 Hz for sinus signals of about 500 Hz) the auditory event was oftentimes described as diffuse and "inflating the whole space", whereby local loudness maximums and minimums could be identified. Loudness maximums have been perceived particularly at the borders of the evaluation range (±90°). This didn't change, if the used loudspeakers changed.

At medium frequency differences ($\Delta f \leq 30$ Hz for sinus signals of about 500 Hz) the auditory events have been described less often as diffuse. Clearly contoured auditory event positions appeared to some extend after longer listening. Hereby the characteristics of the auditory events (sound, loudness) changed. After that the characteristics of the auditory events remained constant.

At big frequency differences above a critical band width (f_1=500 Hz; f_2=660 Hz), narrow contoured positions could be assigned to the auditory events.

There have been subjective differences in describing the auditory events. Out of the answers to the questionnaires at least 4 different types of perception or of description can be derived, respectively:

- Reporting of large auditory event widths: The described auditory events spreads over a relatively large range of angels (e.g. -20°..+90°), but there is no indication about centers of gravity in it.
- Reporting of the centers of auditory events: The described auditory events are oftentimes only composed of single points on the lateralization scale, but there is no indication about the width of the auditory events.
- Reporting of the boundaries of auditory events. The described auditory events contain some distinctive points (maximums or boundary points), but there is no indication about the areas between them.
- Reporting of unique auditory events. Only one auditory event is described, but there is no indication about multiple auditory events.

In principle, it cannot be determined, whether such different descriptions of the same sound situation are based on different types of perception or are based on different methods, to describe the same perceptions.

Evaluation Method

Since the reported direction of the auditory events is quantified (i.e. in steps of one on the lateralization scale from -10 up to +10), an inaccuracy in the reported values of the half step width is inhering, in Fig, 3.2 named as *localization blur*. The width of the auditory events, which has been reported by the test persons, is combined with this inaccuracy and forms the so called *effective auditory event width* (Fig. 3.2).

Because of uncertainties in the localized directions, which have been reported by the test persons, and because of probably big differences in reception or in describing a reception, a so called *lock-in-range F* is defined around the sound source direction, and all portions of an auditory event, which correspond to this lock-in-range, are rated as correctly localized. The lock-in-range has to be constant for all experiments. The lock-in-ranges of two active sound sources shall not overlap. (Fig. 3.2).

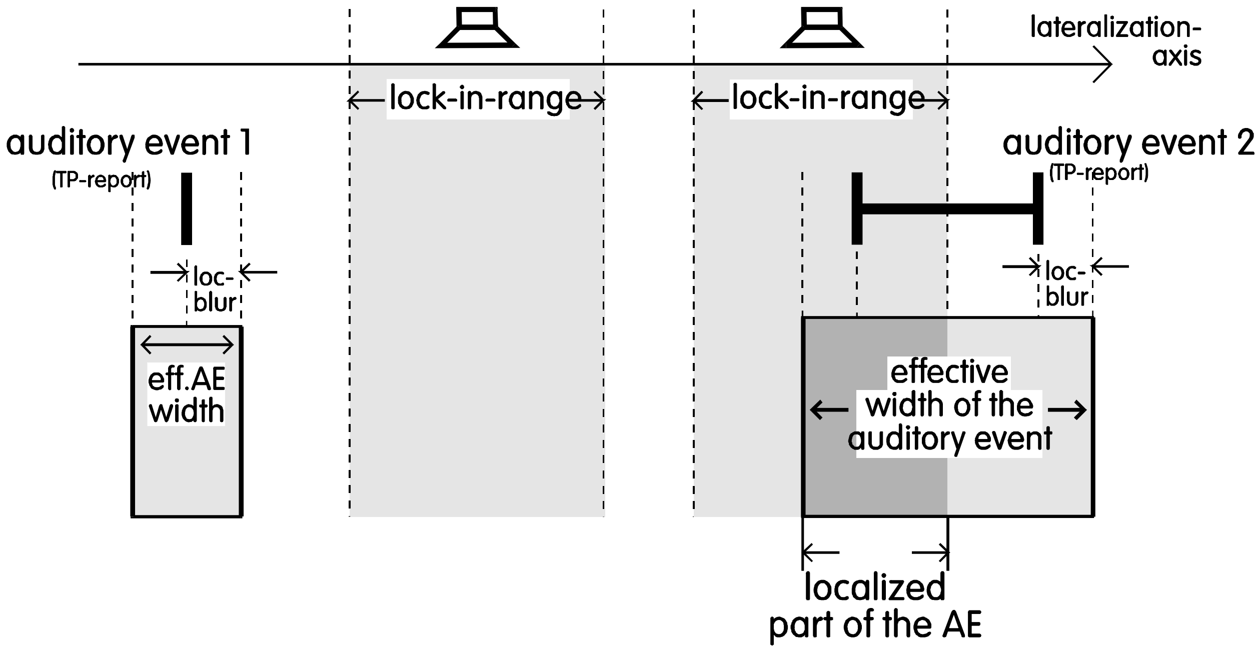

Fig. 3.2: Lock-in-range, effective auditory-event-width (AE=auditory event).

The term *correct localization* is used subsequently to denote the matching of auditory event localization and sound source direction (incl. lock-in-range).

The term *localization rate LG* denotes subsequently the part of the effective auditory event width, which matches to the lock-in-range of a sound source. The localization rate can be interpreted as the probability, to which an auditory event is in relationship to a certain sound event. If all sound sources are localized correctly, the sum of the localization rates of all auditory events meets the number of sound sources. If the number of auditory events N_{HE} exceeds the number of sound sources N_{SQ}, the localization rate has to be normalized by the number of sound sources:

$$LG = \frac{(\text{lock-in-range}) \cap (\text{eff.AE-width})}{(\text{eff.AE-width})} \; \frac{N_{SQ}}{\mathrm{Max}(N_{SQ}, N_{HE})} \qquad (3.2/1)$$

A sound source is considered to be localized correctly, if the averaged localization results are better than random answers would be. The localization rate of random answers is denoted as *guess probability* w_r. If the possible *answer range AntB* is given (e.g..±10) and a *lock-in-range F* is defined for all sound sources (e.g.±1.5) and multiple auditory events are counted without devaluation until the number of sound sources N_{SQ} is reached, the *guess probability* w_r results in:

$$w_r = \frac{N_{SQ} \cdot F}{AntB} \qquad (3.2/2)$$

For 2 sound sources the values above result into a guess probability of 30%.

For evaluation the auditory events of all test persons are valuated (e.g. by the localization rate) and averaged over all experiments and test persons. If useful, the results can also be related to a reference situation. The evaluation method is described in Appendix A.

The following items are evaluated: number of auditory events, localization descriptions, pitch of the auditory events, sound and loudness. These terms are typically depicted as a function of the (mean) frequency difference between the sound sources.

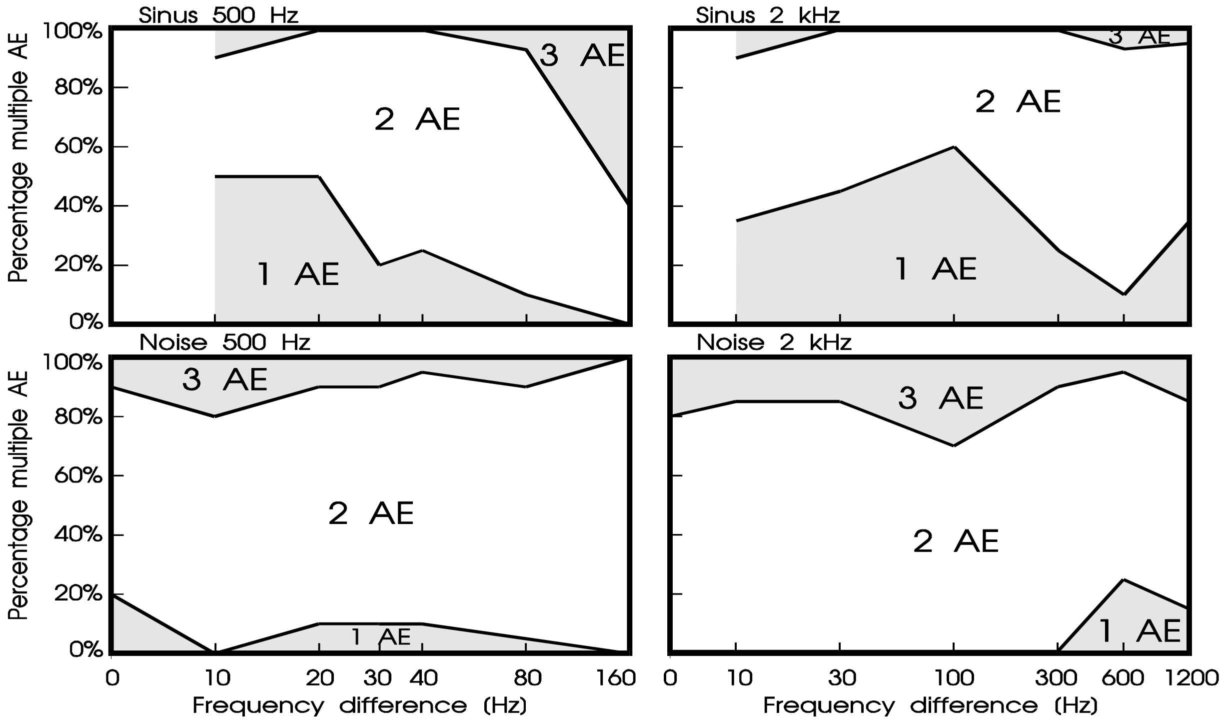

Fig. 3.3: Number of auditory events (AE) at auditory experiments with 2 sound sources as a function of the (mean) frequency difference between the test signals

Number of auditory events

The correct estimation of the number of sound sources is considered as a first step for a correct analysis of the sound situation. The number of perceived auditory events is depicted in Fig. 3.3. These results of the auditory experiments mean:

At experiments with noise signals the number of auditory events mostly corresponds to the number of sound sources. At experiments with sinus signals and low frequency differences oftentimes only a single, diffuse auditory event appears (for $\Delta f<30$ Hz at $f\approx500$ Hz, for $\Delta f<300$ Hz at $f\approx2$ kHz). Presenting sinus signals of 500 Hz and 660 Hz oftentimes 3 auditory events appear, one of them is characterized as a low frequency event, with a pitch in the range of the difference frequency of the signals.

From this follolws, if two sound sources are present and their frequency distance is smaller than a critical bandwidth (≈110 Hz around 500 Hz, ≈340 Hz around 2 kHz), the auditory system is able to perceive two distinct auditory events and to estimate the number of sound sources correctly. Hence the preconditions for an enhanced signal processing of the sound signals are fulfilled.

At 500 Hz-sinus experiments with frequency differences above a critical band width indeed a third auditory event appears from directions of >±70°. The sound corresponds to a sinus tone, the pitch corresponds to the frequency difference between the test signals. The characteristics of this auditory event matches to the parameters of the envelopes of the ear signals. This could indicate non-linear effects (difference tone generation) or the evaluation of envelope information by the auditory system. The phenomenon is not analyzed in detail within this scope.

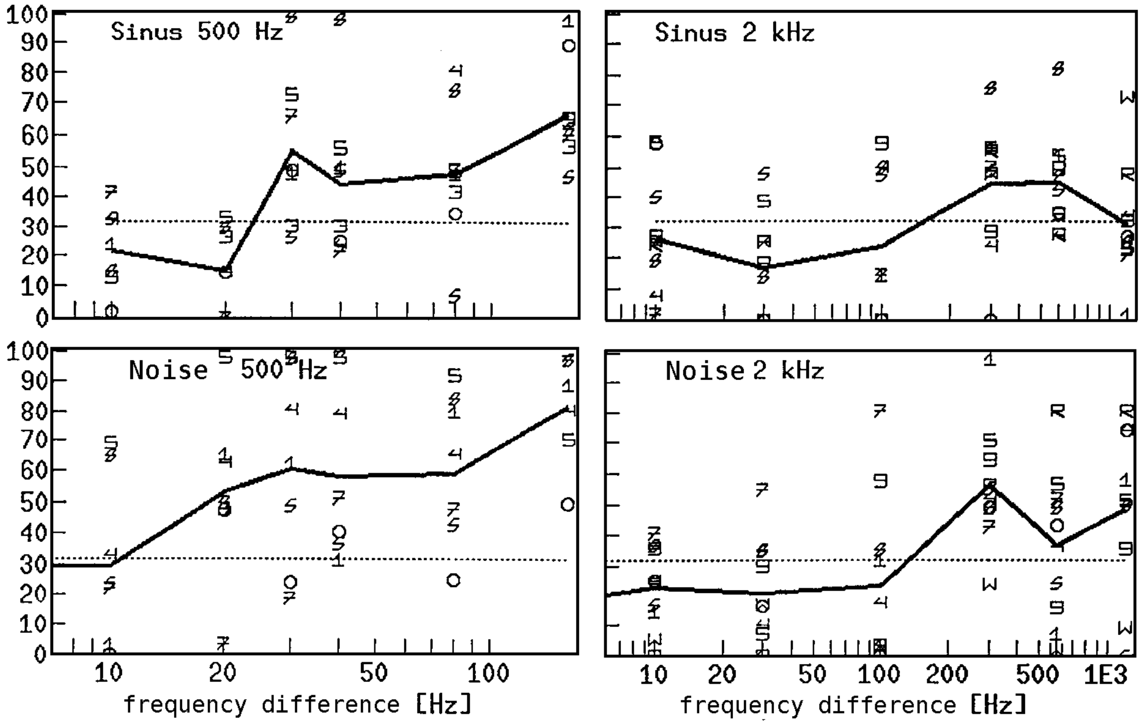

Fig. 3.4: Matching of sound source and auditory event direction.

1345789RSWO	*individual results of the test persons*
———————	*mean localization rate*
....	*guess probability*

Localization of the sound sources

The main questions behind the auditory experiments are: Under which conditions is the human auditory system able to determine the input direction of multiple sound sources correctly? Is an analysis within critical bands sufficient in order to describe these abilities? Or is a finer frequency resolution necessary or even the introduction of additional signal processing methods?

For the evaluation of the localization results a lock-in-range around the sound sources of ±100 µs (±10° for the front direction) is defined and an uncertainty in the localization results of ±50 µs (±5° for the front direction). Values given in µs correspond to the normalized interaural time difference, which are based on a maximum of 625 µs (formula 3.1/1). These premises result into a guess probability of 32% (dotted line). Fig. 3.4 depicts the mean localization rate as a function of the frequency difference between the sound sources. In addition the individual results of the test persons are displayed, too. The following consequences result from this evaluation:

At the experiments above 2 kHz the mean localization rate decreases to a certain extend for higher (center) frequency differences (Sinus 3200/2000 Hz, Noise 2600/2000 Hz).

At experiments with sinus signals above 500 Hz the higher pitched sound source (curve H) can be localized much better than the lower pitched sound source (curve T), while for 500 Hz noise signals there is a trend towards a better localization of the lower pitched sound source. At experiments above 2 kHz clear preference for one sound source type can not be observed.

The mean localization rate is significantly (i.e. at least 50%) above the guess probability for frequency differences Δf_L bigger than

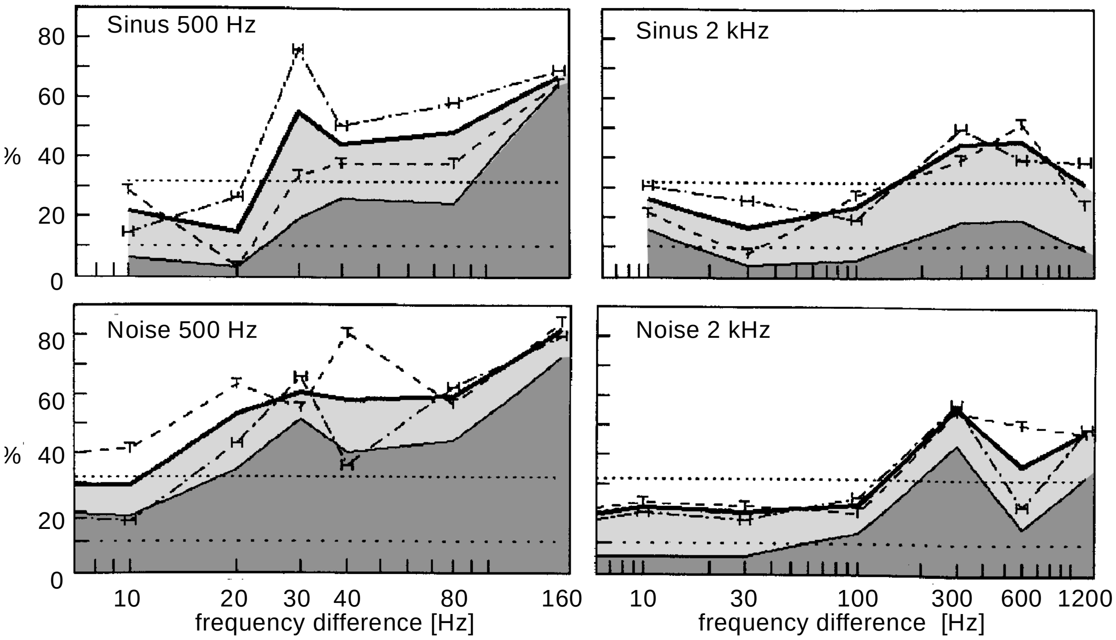

Fig. 3.5: Localization rate for individual loudspeakers (LS) or loudspeaker combinations.

———————	*(light gray background) mean localization rate*
—·—H—·—	*localization rate of the high pitched loudspeaker*
— —T— —	*localization rate of the low pitched loudspeaker*
················	*guess probability*
———————	*(dark gray background) localization rate for localizing both loudspeakers*
· · · · · · · ·	*guess probability for localizing both loudspeakers*

Sinus	500 Hz	$\Delta f_L =$ 30 Hz
Noise	500 Hz	$\Delta f_L =$ 20 Hz
Sinus	2000 Hz	$\Delta f_L =$ 300 Hz
Noise	2000 Hz	$\Delta f_L =$ 300 Hz

Fig. 3.5 additionally breaks down the results into loudspeaker results for higher and lower pitch.

When evaluating the localization rate for the cases, where both sound sources can be localized simultaneously and correctly (thin solid line), the corresponding lower guess probability of 10% is exceeded significantly (by at least 50%) for frequency differences Δf_{L2} above

Sinus	500 Hz	$\Delta f_{L2} =$ 30 Hz
Noise	500 Hz	$\Delta f_{L2} =$ 0 Hz
Sinus	2000 Hz	$\Delta f_{L2} =$ 300 Hz
Noise	2000 Hz	$\Delta f_{L2} =$ 300 Hz (100 Hz)

For frequencies near 500 Hz two sinus sound sources can be localized simultaneously and correctly for frequency differences of at least 30 Hz. For 7% narrow band noise (bandwidth 35 Hz) this is already possible for independent sources with the same center frequency. The spectral distance between both signals is here much smaller than the critical band with (about 110 Hz at 500 Hz).

For signals above 2 kHz a correct localization is possible at frequency distances of at least 300 Hz, this is nearly the critical band width. For some signal configurations the localization rate decreases again for frequencies above.

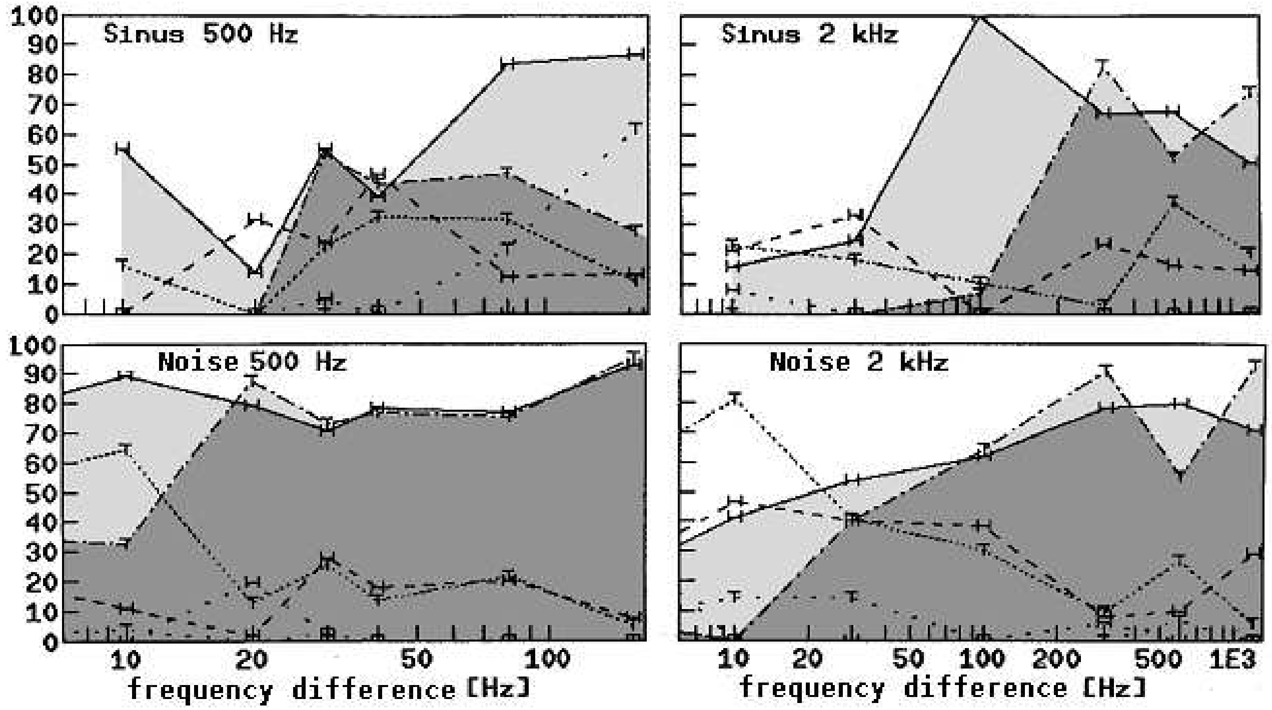

Fig. 3.6: Relative pitch of auditory events (AE) and loudspeakers (LS) for the case that both sound sources are localized correctly
same pitch: ---H-high pitched LS, high AE,- •T- • • -low pitched LS, low AE
different pitch:- -H - high pitched LS, low AE , • •T • • • low pitched LS high AE
• •T • •low pitched LS, medium AE
light gray background: Correspondence of relative pitch for 1 auditory event
dark gray background: Correspondence of relative pitch for 2 auditory events

As a consequence, binaural models, which try to reproduce these effects, must be able (at least for the lower frequency range) to detect and process inside one critical band signals from two input directions simultaneously (see also chapter 3.3 and 3.4).

Relative Pitch Perception

The relative pitch of correctly localized sound sources provides information, whether the sound signals from a specific direction can be identified. A correct assignment of pitches to directions is considered as a first step of a direction selective signal processing.

Fig. 3.6 displays for auditory events, where the directions of auditory events and sound sources match, how often the perceived relative pitches (higher/lower than the other auditory event) correspond to the relative pitches of the sound sources, too. Such correspondences can be observed from the following frequency differences on:

30/80 Hz	at the experiments with	Sinus	500 Hz
20 Hz	at the experiments with	Noise	500 Hz
300 Hz	at the experiments with	Sinus	2000 Hz
100 Hz	at the experiments with	Noise	2000 Hz

At sinus experiments around 500 Hz and high frequency differences oftentimes a third low-frequency auditory event appears (difference tone). The relative pitch of the low frequency loudspeaker is then perceived as medium pitch (between difference tone and high signal frequency).

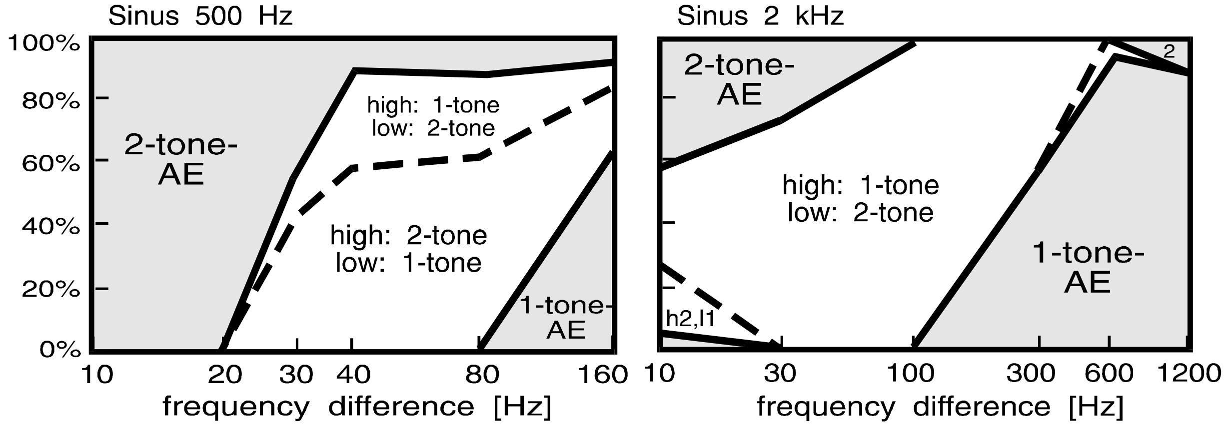

Fig. 3.7: Sound of the auditory events (AE) at sinus experiments for the case, that both sources are localized correctly
1-tone-AE : tonal (sinus like) sound for both auditory events
2- tone-AE: sound of a tone mixture (modulated, rough chord like sound) for both auditory events
1-tone/2-tone: one AE with tonal, one AE with mixture like sound.
low /high : sound of the AE from the direction of the low /high source

Especially at low frequency differences the pitch of the higher loudspeaker is more often perceived correctly than the pitch of the lower one.

Only for frequency differences, where both sound sources can be localized correctly, the pitch of both sources can be identified correctly, too.

This indicates, that the abilities for localization and for pitch identification are coupled. The few cases of correct localization at low frequency differences seem to be more likely cases of random localization, at least the pitch of the corresponding sources couldn't be determined correctly there.

Sound of localized sound sources

For sinus experiments the sound characteristics of the auditory events are evaluated, in terms of being more similar to to a sinus tone or to a mixture of sinus tones (modulated signal, rough sound, mixture of tones, chord). These auditory event characteristics, subsequently denoted as "sound", can give information, whether the human directional processing combines different signals spectrally or processes them separately. Therefore a mixed sound (modulated signal, rough sound, mixture of tones, chord) indicates a common spectral processing of different sound sources, a single tone perception indicates an individual processing. For sinus signals these differences can be noticed easily, but at noise signals different processing strategies only lead to a modification of the bandwidth, and it is very difficult to detect these differences by their sound.

Fig. 3.7 illustrates the sound of the auditory events when both sound sources are localized correctly. Displayed is, which combination of auditory event sounds appears. (single tone/mixture). Additionally these auditory events are related to the pitch of the sound sources.

If for correctly localized sound sources the frequency difference between both sources exceeds a critical band width, the sounds of both auditory events match to sounds of the sound sources (single sinus tones). For low frequency differences only the sound of one auditory event corresponds to the source signal, at the 500 Hz experiments and for frequency differences between 30 Hz and 80 Hz this is mostly the auditory event from the direction of the low pitched source, at the 2 kHz

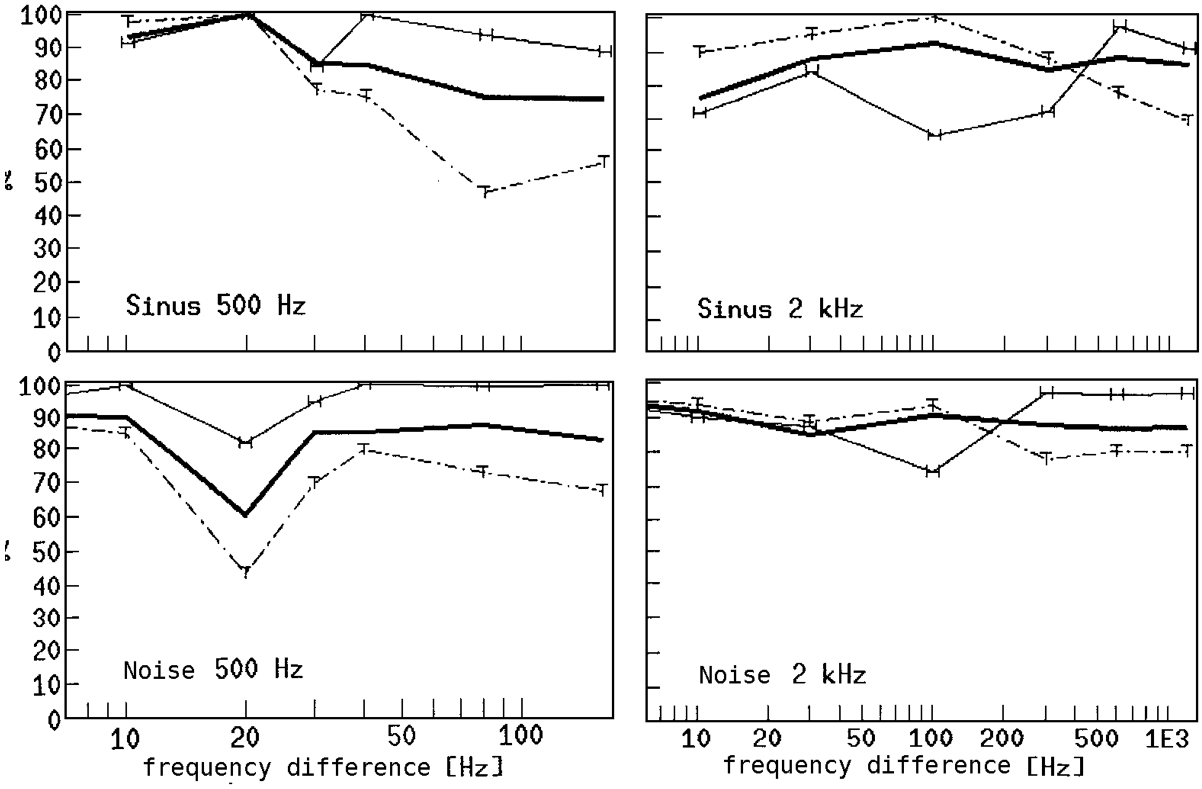

Fig. 3.8: relative loudness of the auditory events (AE) for the case, that all sound sources are localized correctly.
---H-- relative loudness of the AE from the direction of the higher pitched source
•- •-T- •- relative loudness of the AE from the direction of the lower pitched source
------------ mean relative loudness of the localized sources

experiments this is the auditory event from the direction of the high pitched source. The sounds of the other auditory events correspond to a mixture of both sound signals. For frequency differences below 30 Hz both sound sources can only be localized correctly in a few cases, in theses cases the sound of both auditory events corresponds to a mixture of the sound signals.

Therefore a direction selective processing of both sound signals is only possible, if the signals are located in different critical bands. If the spectra of the sound signals overlap inside one critical band, the directions of both sound sources can be localized correctly, but the sound signals, which belong to one direction, can no longer be reconstructed completely. From this follows, that the ability for localization is not combined with the ability for a direction selective extraction of the signal spectra.

Loudness of the localized sound sources

In Fig. 3.8 illustrates the perceived relative loudness (loudest auditory event = 100%), for the case, that both sound sources are localized simultaneously and correctly.

For frequency differences, where most of the test persons can localize both sound sources correctly, the higher pitched sound source is mostly perceived louder than the low pitched one. For lower frequency differences either there is no difference in loudness (Sinus 500 Hz, Noise 2 kHz) or the loudness of the low pitched source prevails. (Sinus 2 kHz)

Results of the experiments in the frequency range around 500 Hz

If two sound signals from different input directions are presented at the frequency range around 500 Hz, sinus signals cannot be localized correctly below a frequency difference of 30 Hz, then the localization rate lies below the guess probability. Here mostly diffuse auditory events appear. If a coincidence between the directions of sound source and auditory event happens, the relative pitch of the sound sources cannot be determined correctly. This suggests, that the correct localization is only caused by a random accordance of the directions. The sound corresponds always to a mixture of both source signals. Accordingly the loudness of both auditory events is equal.

For sinus signals with a frequency differences of at least 30 Hz and for independent noise signals with equal center frequency the test persons are able to determine the number of sound sources correctly. (Exception: sinus signals above one critical band width) and to localize the sound sources correctly. The frequency distance, which is necessary for localizing, is significantly lower than the critical band width of 110 Hz. The correct assignment of relative pitches to the auditory event direction is not always possible here. If both sound sources are localized correctly, mostly the sound of one source can be extracted correctly from the interfering sound signals and is perceived as a single tone, whereas the perceived sound from the other direction corresponds to a mixture of the signals.

For frequency differences of at least 80 Hz (which is already below the critical band width of 100 Hz) the directions and the relative pitches of both signals can be determined correctly. At sinus experiments one signal is perceived in its original sound, the sound of the other one corresponds to a mixture of both signals. The higher pitched sound source is mostly perceived as the loudest one. This corresponds to the perceived sound, which itself corresponds to the sum of both signals.

If the (center) frequency difference exceeds the critical band width, the majority of the test persons can identify the directions of both sound sources correctly. But at sinus experiments there are in 60% of all cases 3 auditory events. The pitch of one auditory event matches to the difference of both signal frequencies, which corresponds to the period of the envelop of the ear signals. Apart of this effect the relative pitches of the sound sources can be determined correctly and the original sound of both signals can be extracted. The loudness of the higher pitched sound source remains significantly louder than the lower pitched sound source. From the curves of equal loudness an increased loudness of the higher pitched source would be expected, but significantly larger loudness differences could be observed in these experiments.

Results of the experiments in the frequency range around 2 kHz

At the experiments with sinus signals around 2000 Hz test persons could only realize the sound situation correctly and perceive 2 auditory events, if the frequency difference between the signals exceeded 300 Hz, this is a critical band width. For noise signals there was no problem in perceiving two separate auditory events..

For spectral differences of at least 300 Hz both sound sources could be localized correctly (for $\Delta f \geq 300$ Hz for sinus signals and for $\Delta f \geq 100$ Hz for noise signals, which results in a total bandwidth of $\Delta f \geq 240$ Hz). In these cases the relative pitch of the sound sources was also identified correctly.

For sinus signals with frequency differences below 300 Hz the sound of the higher pitched auditory event is mostly described as a single tone, whereas the sound of the lower pitched auditory event is mostly described as a mixture of both signals (2 tones, rough or modulated sound). Above a

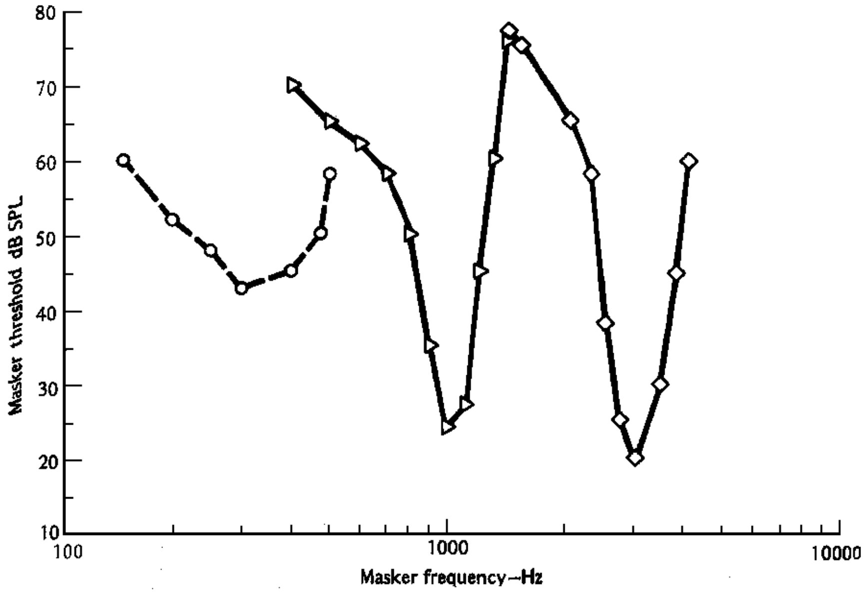

Fig. 3.9: psychoacoustical measured tuning-curves: masking threshold of test tones of 300 Hz, 1000 Hz and 3000 Hz against the frequency of a sinus masker (Wightman/McGee/ Kramer, [46])

critical band width the sounds of both localized sound sources are perceived correctly as single tones.

The perceived loudness and perceived sound are in accordance with each other. Auditory events, whose sound correspond to the sum of both signals, appear louder than single tone auditory events.

If an additional critical band width is exceeded (at frequency differences of about 700 Hz (noise) or 1100 Hz (sinus)) the results do not differ from the situation before. For sinus signals the localization rate decreases for big frequency differences. For 1200 Hz frequency difference the localization rate decreases drastically and reaches the guess probability,

3.3. Theses about the Signal Processing of the Auditory Systems

Concerning the signal processing of the auditory system the following thesis are postulated:

- The binaural analysis of the ear signals is carried out inside critical band wide frequency ranges.
- Critical bands are established by combining corresponding hair cell regions of the inner ear. The slew rate of the critical band filters corresponds to the so called tuning curves: 30..100 dB/Oct for the low frequency slope, up to 300 dB/Oct for the high frequency slope. Fig. 3.9 shows psychoacoustically measured tuning curves for 3 frequency ranges. The tuning curves indicate, how the audibility threshold of a sinus test signal changes, when a signal of another frequency is present. The tuning curves indicate, whether and to what extend signals of different frequencies are processed together by the auditory system (They characterize quasi the discrimination power of the auditory system).

  Identical critical bands are formed for both ears and for all analysis objectives of the auditory system (direction, loudness, sound, assignment of loudness, pitch and sound to directions). The critical band widths of Zwicker et.al.[52] (appendix B) should therefore be transferable to models of binaural signal processing (about 100 Hz for frequencies below 500 Hz, a third octave for frequencies above).

When presenting a single tone with a level of 50 dB, a range of at least one octave (3 critical bands) is stimulated above threshold. The same applies for two signals with a low spectral difference.

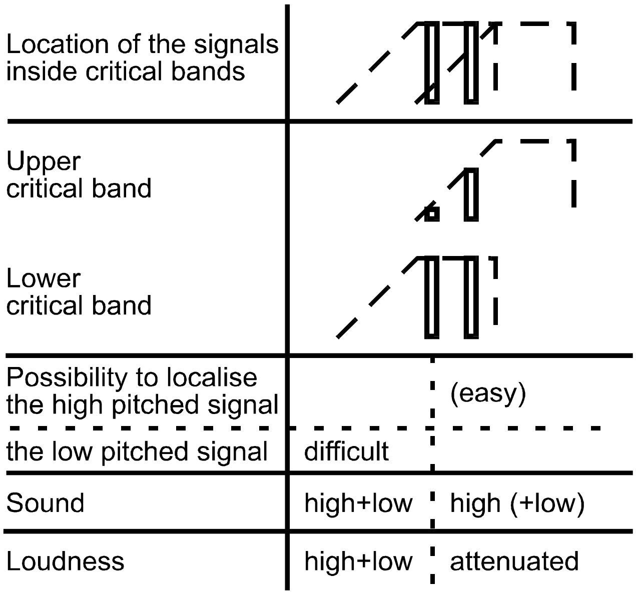

Fig. 3.10: Expected results of the auditory experiments in case of an independent analysis of the signals in different critical bands .

It is assumed, that the main critical band is formed by combining the hair cell regions with the biggest stimulation. The upper and lower neighbor critical bands capture those parts of the stimulation, which are not covered by the main critical band. This thesis is supported by investigations on loudness mechanisms (i.e. Zwicker et.al.[52]), where signal frequencies are combined for evaluation, until a critical band width is reached.. Similar results arise from investigations on binaural critical bands (Scharf et.al.[35]; Kohlrausch [23]), and from the evaluation of binaural auditory experiments (Slatky [39]).

Appendix B gives an overview about further models describing the position of critical bands..

3.4. Comparison between Auditory Experiments and Binaural Models

The question is, whether the results of the auditory experiments can be explained by an independent analysis of interaural parameters in different critical bands without any further processing steps, or whether additional signal processing steps have to be postulated. Fig. 3.10 shows the signal portions in different critical bands and the expected results, if signal attributes are directly determined from the contents of isolated critical bands,

Fig. 3.11 depicts the results of a binaural cross-correlation-model according to Lindemann [25] when analyzing the signals of the auditory experiments in different critical bands, using critical band filters as described above. The correlation pattern represent a measure for the signal power in dependency of the interaural time difference and the time.

With the assumed critical band filters at least 2 critical bands would be stimulated by the signals of the auditory experiments, the main critical band, which contains the signals of both sound sources with their original amplitude, and the upper neighbor critical band, where the lower pitched signal is attenuated more significantly by the slope of the critical band filter than the higher pitched signal. Inside the lower neighbor critical band also some (relatively poor) signal portions of the lower pitched signal could be expected. .

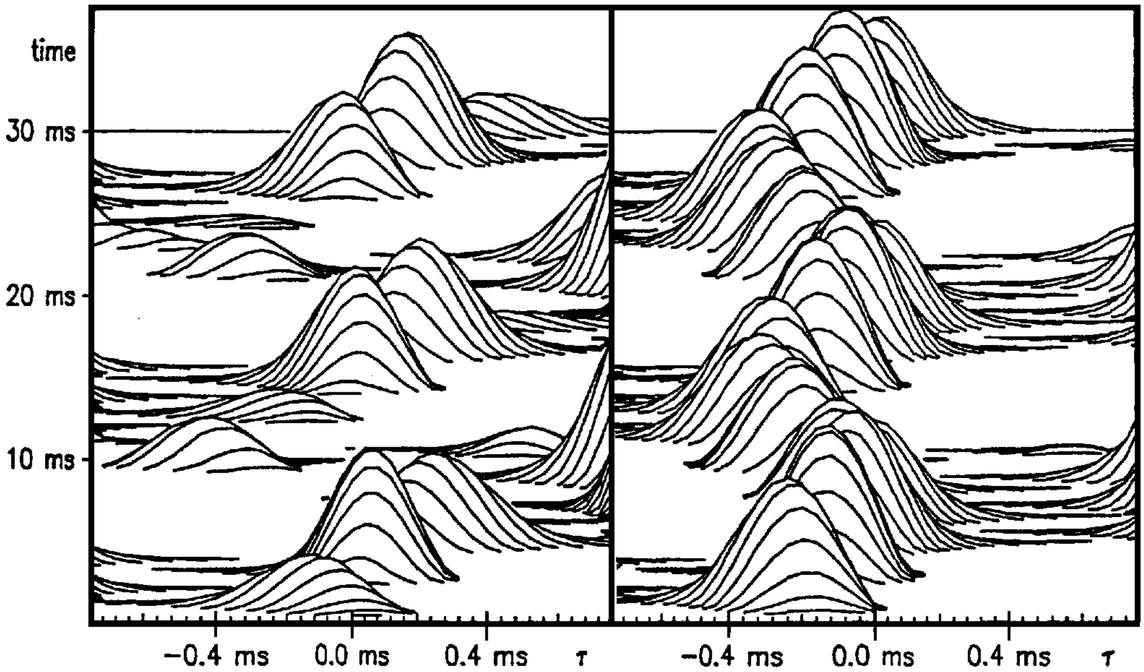

Fig. 3.11: Normalized correlation pattern of the binaural model according to Lindemann [25] for 2 sound sources of equal amplitude and different interaural time delay τ.
Signals: sinus 580 Hz, τ_a=-200 µs + sinus 500 Hz, τ_b=600 µs
left: main critical band ; right: upper critical band

The interfering of signals of different frequencies results into interaural phase and level differences, which fluctuate time dependently. Inside the main critical band, which contains both signals with the same amplitude, the positions of the maxima of the cross correlation function (Fig. 3.11) fluctuate heavily in time. The center of gravity of the pattern corresponds rather to the mean interaural difference than to one of the existing input directions. A direct evaluation of the input direction and the amplitude of one sound source based on these patterns seems not to be possible. (A more accurate examination of interaural parameters for two interfering sound sources can be found in chapter 4.2).

In the upper critical band both signals are attenuated by nearly the same extend by the slope of the critical band filter, as long as the frequency differences are small. The correlation patterns are similar to the correlation patterns inside the main critical band. If the frequency difference grows, the signals are attenuated differently by the slope of the critical band filter. Direction and parameters of one signal can be determined directly.

For signals of 500 Hz and 530 Hz, where both sound sources can be localized correctly at the auditory experiments, there would be a level difference of 3-8 dB between both signals inside the upper neighbor critical band (critical band filter slope 30-100 dB/octave). This would not be sufficient for a stable directional estimation based on these this critical band signals. Inside the lower neighbor critical band level differences between both signals of 8-25 dB would appear (critical band filter slope 100-300 dB/octave).

In this case the correct *localization* of both signals cannot be explained by an independent evaluation of signals inside critical bands. The interaural parameters of one signal can indeed be determined from one neighbor critical band. But information about the input direction of the other sound source would only be achievable, if the variant cross correlation patterns inside the main

Table 3.1:
Comparison Auditory Experiments vs.
Results of a Cross Correlation Model (CCF-Model)
CB= Critical Band

Results of the Auditory experiments	Results of the CCF-model: Main-CB	Results of the CCF-model: Up.neighbor CB	Conclusion
Sound			
for signal frequencies below a critical band width only 1 signal with original sound	mixture of both signals	higher pitched signal prevails sound determinable	evaluation is carried out inside of critical bands
- - - - - - - -	- - - - - - - -	- - - - - - - -	
above a CB-width the sound of both signals identifiable	lower pitched signal sound determinable	higher pitched signal sound determinable	
Localization			
Experiments around 500 Hz: above 20...30 Hz frequency difference both sources localizable	variant CCF: source direction not directly determinable	invariant CCF above Δf=20...30 Hz: direction of higher pitched signal determinable	"re-computation - mechanism": determining of the direction of the lower pitched signal from variant CCF
- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -
experiments around 2 kHz: above 300 Hz frequency difference both sources localizable	invariant CCF	invariant CCF	source direction directly from CCFs determinable (Δf > CB-width)
Loudness			
At correct localization higher pitched source louder than lover one (loudness of lower signal 70% of higher signal)	amplitude corresponds to sum of both signals	amplitude low (attenuated via slope of the critical band filter.)	"re-computation-mechanism": Assignment of loudness to source directions

critical band could be interpreted. In order to reproduce the results of the auditory experiments, signal processing algorithms are necessary, which are able to determine the directions of sound sources from variant cross correlation functions (Table 3.1).

For determining the relative *pitch* of both signals an independent evaluation of the signals of different critical bands is sufficient. If the direction of one sound source can be determined from one neighbor critical band, information about the relative pitch of the signal can be achieved simply from the location of this critical band. The relative pitch of the signal mixture inside the main critical band results from the relative location to that neighbor critical band. As a consequence, the determination of the relative pitches would only be possible, if at least one sound signal can be localized.

Also the characteristics of the perceived *sound* of the auditory events corresponds to an independent evaluation of the critical band signals. For frequency differences below a critical band width only the sound of one localized sound source can be evaluated correctly, while the perceived signal from the direction of the other sound source sounds like a mixture of both sound source signals. This corresponds to the signal content of the critical bands: A single tone in one neighbor critical band, a mixture of the signals inside the main critical band. Here too, a determination of the sound is only possible, if at least one sound source can be localized correctly.

The *loudness* of the auditory events does not correspond to the expected results for an independent evaluation inside critical bands. For correct localized sound sources at the 500-Hz-experiments, for example, the higher pitched sound source was perceived louder than the lower pitched sound source. But in the binaural model the upper neighbor critical band, which might be used for the evaluation of the direction of the higher pitched sound source, contains much lass portions of the signal power than the main critical band. This would suggest a contrary loudness impression. The direct assignment of the power inside critical bands to the loudness of a direction leads to wrong results. As a consequence, an additional mechanism for auditory models is needed, which assigns loudness to input directions.

To reproduce the results of the auditory experiments by a binaural model, a mechanism is required, which is able to evaluate directional information from critical bands with variant interaural parameters or from variant cross correlation functions, and which assigns loudness to directions. The integration of a direction specific processing of signal phases or signal frequencies into this mechanism is not mandatory, according to the results of the auditory experiments. This mechanism should at least offer the possibility to include additional directional information sources, for example information from other critical bands or from optical information (look at the sound source), from reminder etc.

In the following chapters possible algorithms for these purposes shall be investigated. The focus of these investigations shall not only be, to identify possibilities for describing the results of the auditory experiments by a model, but above all to point out possible applications of such a model for signal processing purposes.