Modeling of Binaural Discrimination of multiple Sound Sources: A Contribution to the Development of a Cocktail-Party-Processor[4]

H.SLATKY
(Lehrstuhl für allgemeine Elektrotechnik und Akustik, Ruhr-Universität Bochum, D-4630 Bochum, Germany)

The human auditory system is able to "focus" on one sound source in the presence of noise, echoes, reverberation and other interfering sources. Such a situation is given, for instance, in a room with more than one speaker ("cocktail-party-effect"). In my study, I intend to find algorithms modeling these binaural phenomena, which can be used for technical purposes.

Lateralisation of multiple sound sources by the auditory system

In order to answer the question how the human auditory system reacts on presenting more than one simultaneous sound source, auditory experiments have been conducted, presenting two sinus or narrow band noise signals simultaneously in an anechoic room.

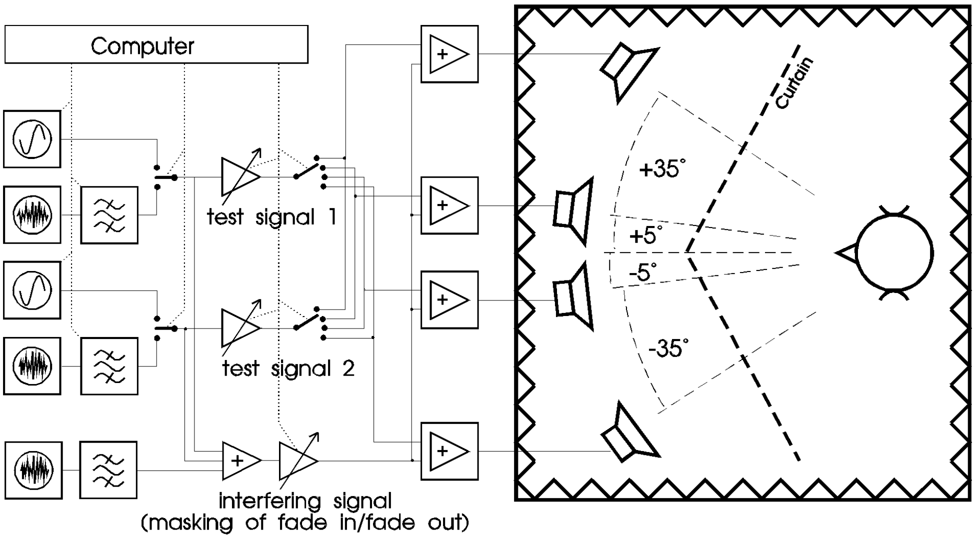


Fig. 1:
Setup for localization experiments of multiple sound sources
Presented signals
1. *Sinus 500 Hz + Sinus (500 Hz + x), x=10.. 160 Hz*
2. *Sinus 2000 Hz + Sinus (2000 Hz + x), x=10..1200 Hz*
3. *Narrow band noise (7% rel. bandwidth): Noise 500 Hz + Noise (500 Hz + x), x=10.. 160 Hz*
4. *Narrow band noise (7% rel. bandwidth): Noise 2000 Hz + Noise (2000 Hz + x), x=10..1200 Hz*

When simultaneously presenting two narrow-band sound sources with spectral differences substantially smaller than the critical bandwidth (i.e. sinusoidal signals 500 Hz + 530 Hz or noise with 7% relative bandwidth 500 Hz + 510 Hz) the auditory system is able to localize these sources correctly and to identify the sound sources by their pitch (high-pitched, low-pitched) [3]

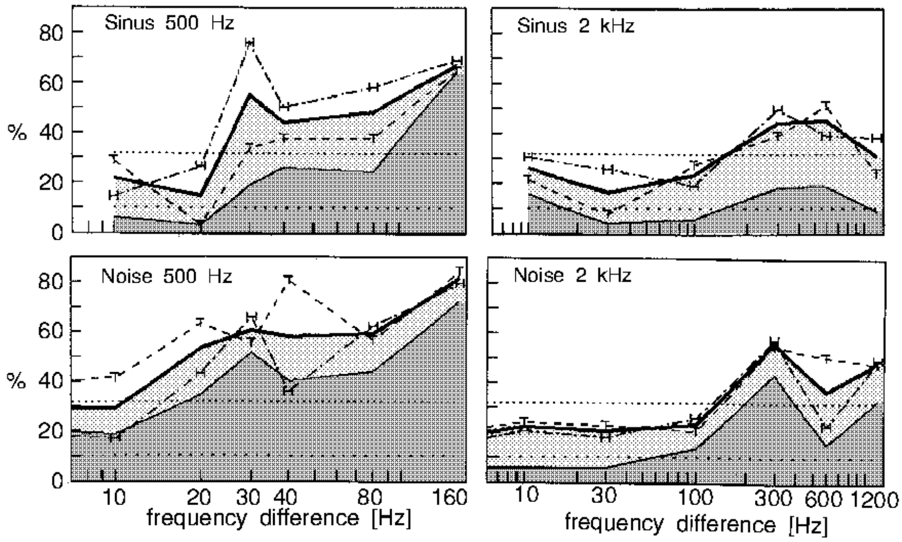


Fig. 2:
Percentage of correctly localized sound sources.
--H-- high-pitched sound source
--T--- low-pitched sound source
one sound source localized correctly
for both sound sources localized correctly
.... guess probability for one sound source
. . . . guess probability for both sound sources
at 500 Hz the guess probability is exceeded for Δf>10 Hz (Noise) or Δf>30 Hz (Sinus)

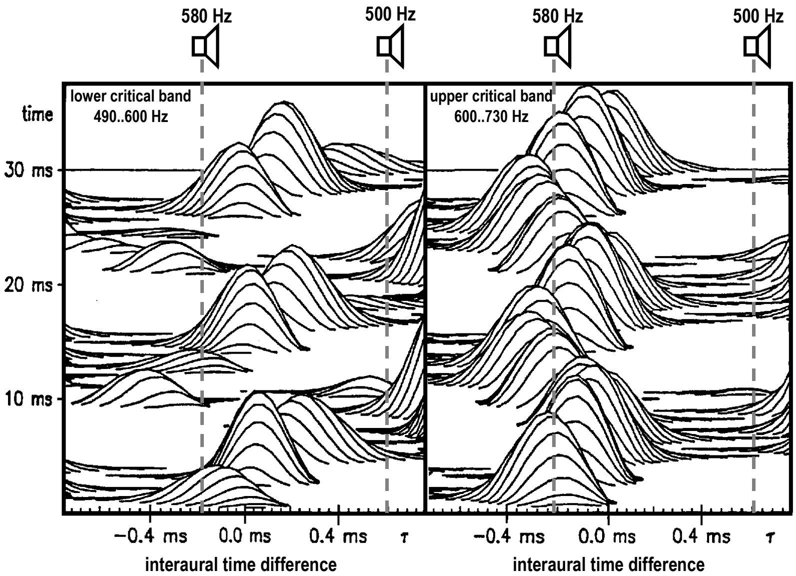


Fig.3:
Interaural cross correlation function[2] of signals of the auditory experiments within concerned critical bands:

Presented signals: :
Sinus 500 Hz τ= 0.6 ms
Sinus 580 Hz τ=-0.2 ms

Dotted lines: interaural time differences of the presented signals

Within the lower critical band there is no correspondence between the positions of the maxims of the cross correlation function and the directions of the sound sources.
Within the upper critical band the positions of the maxims of the cross correlation function corresponds to the direction of the high-pitched sound source.

Binaural models

Presenting these signals to binaural models, which are based on cross correlation functions within critical bands and which determine the direction of incidence directly from the positions of the maxims of the correlation function (i.e. LINKDEMANN[2], GAIK[1]), only one incidence direction can be determined correctly, because maxim positions of only one (from two) concerned critical bands stay constant in time. The cross correlation pattern at the other critical band varies quickly with time. A direct evaluation of directions of incidence is not possible.[3]

Assuming, that the auditory system analyzes the incidence directions within critical bands and that the localization process of the auditory system can be described by cross correlation functions, a method must exist, to extract relevant information on sound directions out of these patterns. ("recomputation mechanism"[3])

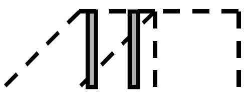
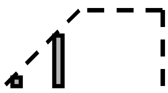
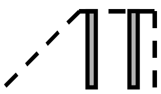
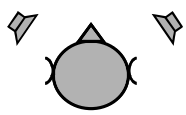
		Localisation	**Sound**	**Loudness**
Signals within critical bands				
Signals within upper critical band		Localisation of high-pitched signal exspected	Sound of high-pitched signal exspected	High-pitched signal with reduced loudness
Signals within lower critical band		no localisation exspected	Mixture of both signals exspected	Sum of both signals exspected
Result of auditory experiments		Both signals localised correctly	Original sound for high-pitched signal Mixture of signals at direction of low-pitched signal	High-pitched signal 140% of loudness of low pitched signal
Consequence for binaural modeling		Extension of auditory models necessary	Model and experiments match	Extension of auditory models necessary

Fig.4:
Comparison between cross correlation models and the results of the auditory experiments

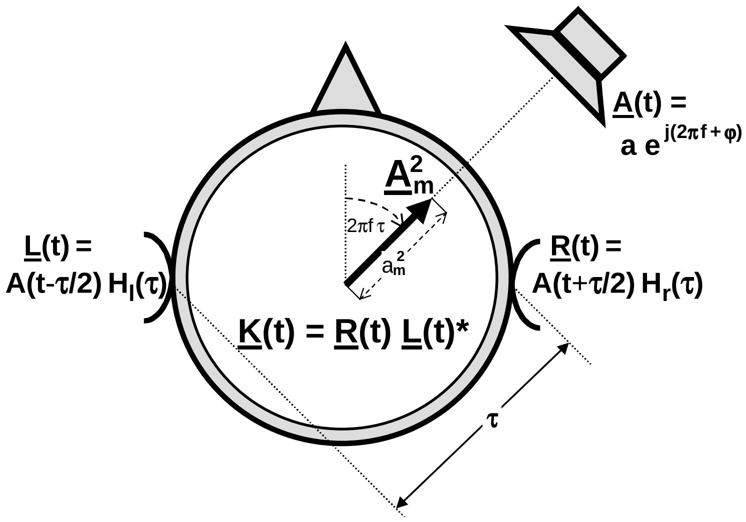


Fig.5:
Interaural cross product k(t) for one sound source witch constant amplitude

Searching for a suitable mathematical description

Another method of describing binaural interactions within critical bands is the complex cross product of the analytic time functions of the ear signals. The features are:

- Using analytic time functions within critical bands, ear signals may be processed with reduced data rate, so processing becomes faster.
- The dependence of binaural interaction patterns on ear signals can be evaluated in mathematical exact form.
- In the presence of stationary signals from only 1 or 2 directions, the binaural interaction pattern results in a simple geometric form (see below).

Within critical bands arbitrary signals can be described as amplitude and frequency modulated sinus signals. Their analytic time function $\underline{A}(t)$ is: (f(t)=frequency, a(t)=magnitude, φ(t)=phase)

$$\underline{A}(t) = a(t)\, e^{+j2\pi\, f(t)t \,+\, j\,\varphi(t)}$$

The corresponding ear signals are: (τ=interaural time difference, $H_l(\tau)$, $H_r(\tau)$ outer ear transfer functions)

$$\underline{L}(t) = \underline{A}(t-\tau/2)\, H_l(\tau) \qquad\qquad \underline{R}(t) = \underline{A}(t+\tau/2)\, H_r(\tau)$$

The cross product $\underline{K}(t)$ of left and right ear signals results to:

$$\underline{K}(t) = \underline{R}(t)\, \underline{L}(t)^* = a_m(t)^2\, e^{+j2\pi\, f(t)\tau} \qquad\qquad a_m(t)^2 = a(t)^2\, H_l(\tau)\, H_r(\tau)$$

For sinusoidal signals (a(t), f(t), φ(t)=const.) the locus curve of the interaural cross product $\underline{K}(t)$ is represented by a single point in the complex plane. The magnitude is proportional to the

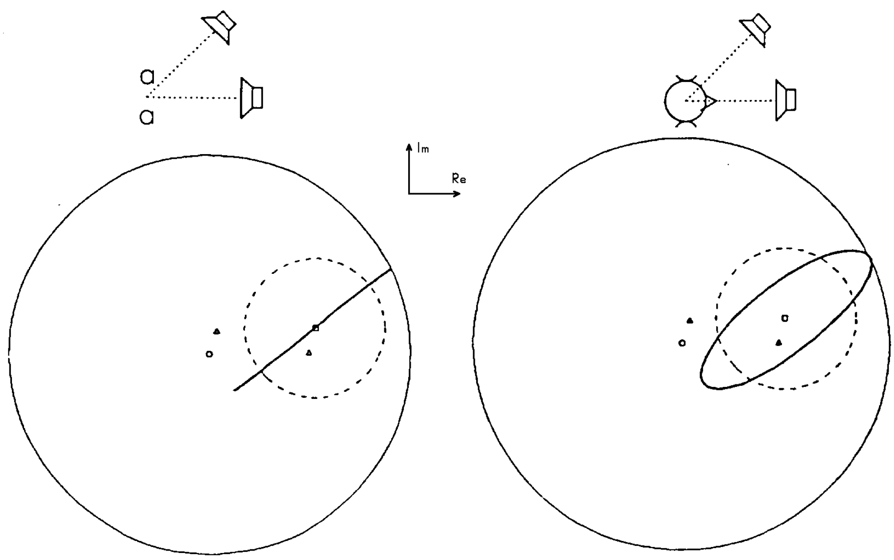


Fig. 6:
locus curve of the cross product, presenting 2 sound sources:
a) sine 500Hz, a=1 , τ_a=0μs
b) sine 560Hz, b=0.5, τ_b=400μs
left figure: interaural level difference 0dB
right figure: interaural level difference 6dB
Δ locus curve of each sound source alone
□ complex mean value
---circle around mean value, radius =standard deviation

medium energy of the ear signals, the phase correlates to the interaural phase. This corresponds to the results of cross correlation models depicting the maxims in polar coordinates.

Presenting 2 signals $\underline{A}(t)$, $\underline{B}(t)$ from different directions, the corresponding ear signals are added and binaural beats arise. The locus curve of the cross product varies quickly with time. When presenting stationary signals, the locus curve has the form of a straight line or of an ellipsoid, depending on the interaural level differences. Introducing the complex mean value $\underline{\mu}$ and the complex standard deviation $\underline{\sigma}$ of this time dependent locus curve, a system of complex equations can be obtained. Interaural phases $2\alpha=2\pi f_a(t)\tau_a$, $2\beta=2\pi f_b(t)\tau_b$, and the mean amplitudes $a_m(t)$, $b_m(t)$ of the sound sources can be estimated from this equation system.

$$\underline{\mu}(t) = 1/2T \int_{t-T}^{t+T} \underline{K}(t')\, dt' \qquad \underline{\sigma}^2(t) = 1/2T \int_{t-T}^{t+T} [\, \underline{K}(t') - \underline{\mu}\,]^2\, dt'$$

$$\underline{\mu}(t) = a_m(t)^2 e^{j2\alpha} + b_m(t)^2 e^{j2\beta} \qquad \underline{\sigma}^2(t) = 2\, a_m(t)^2\, b_m(t)^2\, e^{j2(\alpha+\beta)}$$

Properties of the presented algorithm

The accuracy of this method depends on the integration time and the variation rate of sound source attributes. Stationary signals (sine, harmonic signals) and a long integration time result into a sufficiently accurate estimation (error < 1dB) up to differences in the sound source levels of 100 dB. Using signals with varying amplitudes (noise, speech) the integration time must be short (10-20 ms). Thus, the range of accurate estimations of sound source magnitudes and directions is limited to sound level differences of -20 dB between desired signal and interfering signal.

Compared to other methods of directional selection (beam microphone, linear microphone array technique) the algorithm leads to rather sharp directional beams for receiver distances, which are substantially shorter than the wave length (ear distance). In the low frequency range directional beams of +/-150 µs (+/-15° related to the front direction) can be obtained.

Presenting more than two sound sources within the frequency band of one critical band, the attributes of the two most intense sound sources can be estimated by using the locus curve of the cross product. For a given direction it is possible to estimate the probability that estimators of the algorithm correspond to this direction (evaluation of the error of estimation). In this way the probable amplitude of a signal coming from a desired direction can be estimated.

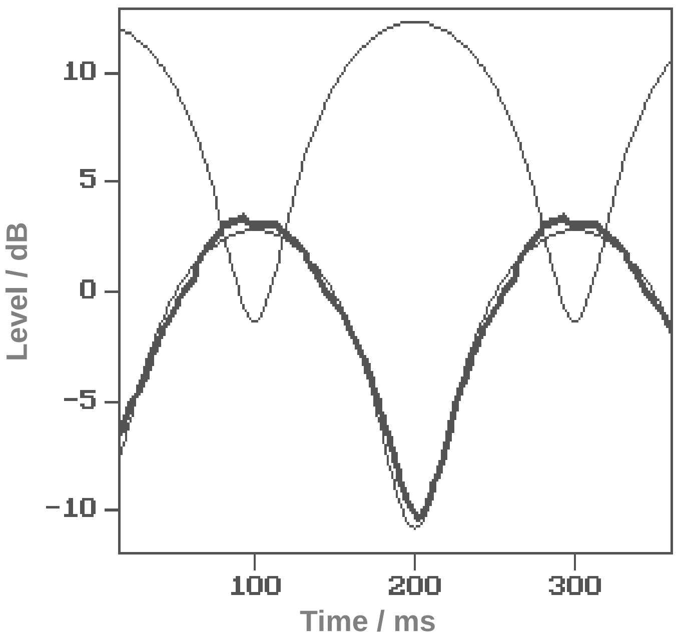


Fig. 7:
Directional filtering of amplitude modulated signals.
Desired signal: level = 0dB
sine 560Hz, f_{mod}=5 Hz, τ=400µs

Interfering signal level=10dB
sine 500 Hz, f_{mod}=5 Hz, τ=0µs

—— Signal envelopes of desired and interfering signal
▬▬ Estimator for the envelope of the desired signal

x-axis: time in ms
y-axis: level in dB,
relative to mean desired signal

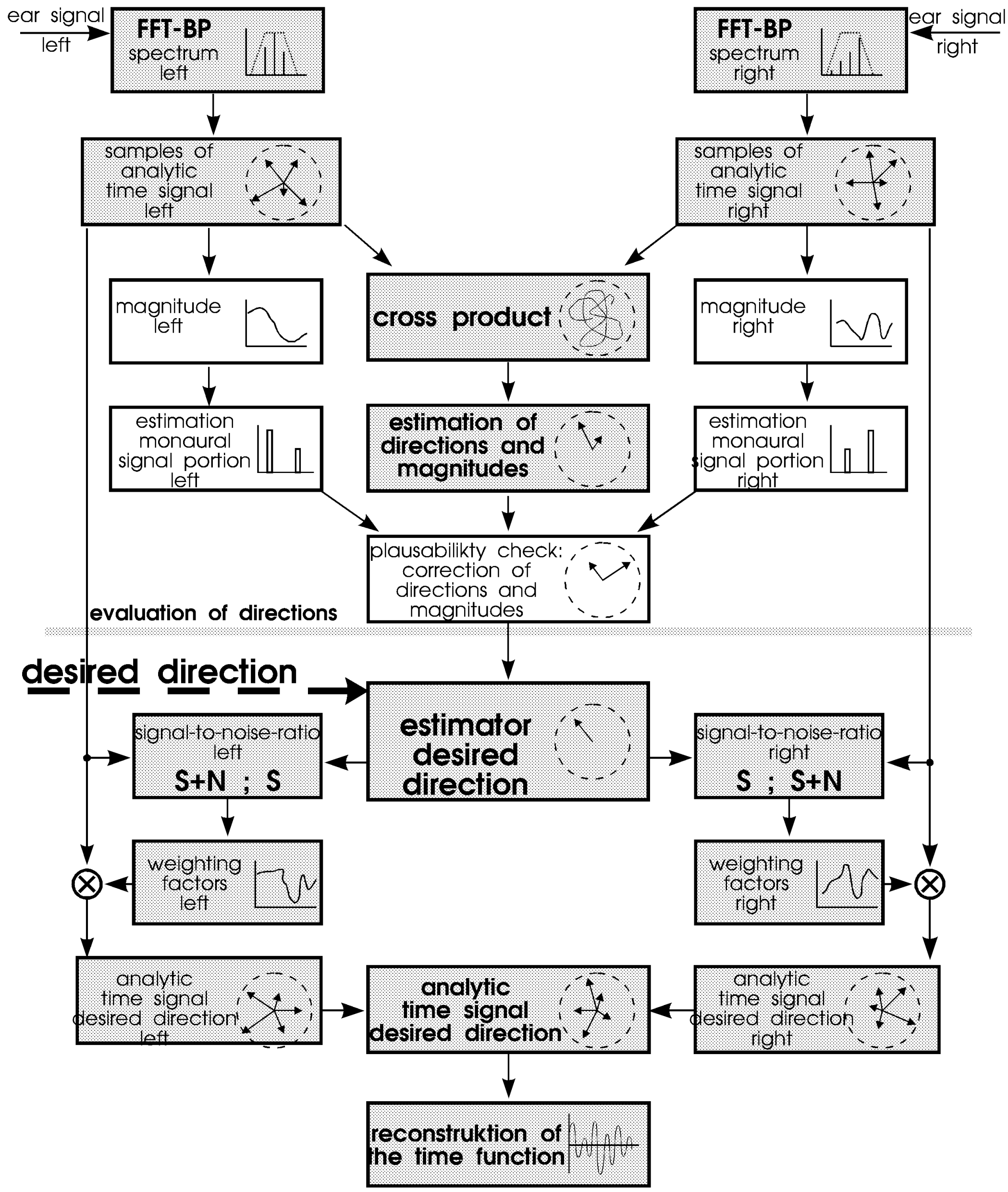


Fig: 8:
The binaural signal processing model (inside one critical band)

Construction of a binaural signal processing model

A binaural signal processing model based on this algorithm must include the following units:

- Preprocessing: critical band filtering of the ear signals and evaluation of the analytic time signal.
- Evaluation of the cross product and its complex mean value and standard deviation.
- Estimation of directions and amplitudes of sound sources from the statistical parameters of the cross product, estimation of the error and validity range of the estimation.
- Choice of the desired direction.
- Estimation of the probable magnitude of the desired signal by considering estimated values and errors of estimation.
- Evaluation of the signal-to-noise-ratio in each ear signal by comparing the estimated desired signal with the ear signal magnitudes => weighting factors for the ear signals.
- Generation of the processed broadband signal out of these weighted critical band signals.

Using this process, an enhancement of a desired speaker's signal of up to 20 dB can be obtained, presenting 2 speakers under free field conditions with original signal-to-noise-ratios of up to -30 dB. Intelligibility of the desired speaker grows considerably.

By processing complex analytic time functions instead of real signals, data rate and computation time can be reduced significantly. Since the magnitudes of spectral components in the range

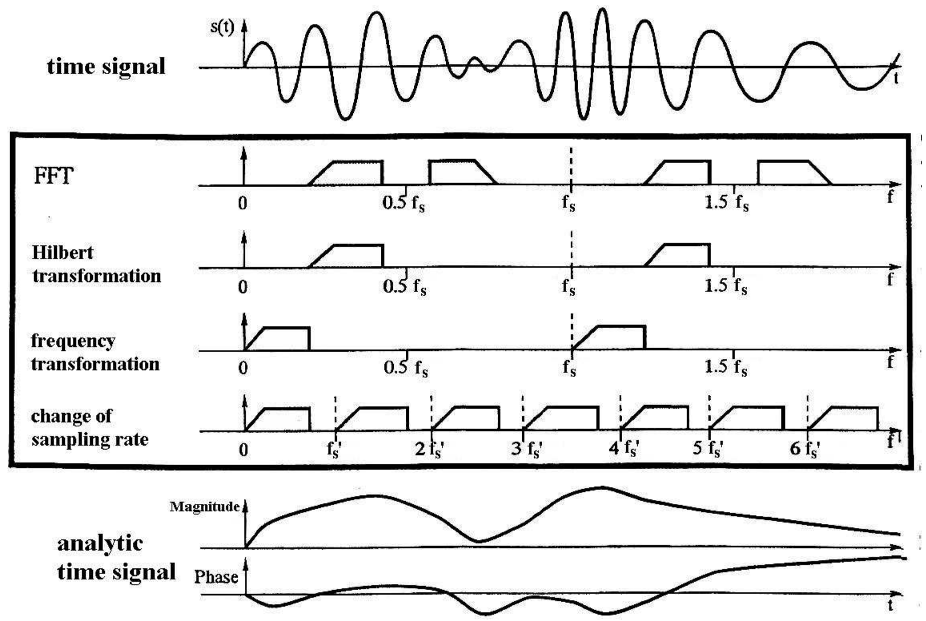


Fig. 9: Preprocessing unit: Generation of the analytic time signal combined with the reduction of the sampling rate

$f_S/2..f_S$ are zero (f_S=sampling rate), critical band filtered signals can be transformed to the low frequency range and be processed with a sampling rate corresponding to the bandwidth of the critical band filter. Using 24 critical bands, the data rate can be reduced to 10-20%, compared to a digital filter bank without down-sampling

Discussion

The presented algorithm is based on the evaluation of the interaural phase. For the high frequency range (f>800 Hz) the relationship between the direction of incidence and the interaural phase gets ambiguous. When interaural phases of desired and interfering directions meet, there is no effect in directional filtering. This problem could be solved by an additional directional filter mechanism based on interaural level differences.

In Psychoacoustics this model can be used for the interpretation of multiple sound source effects and especially the precedence effect. For this purpose a "directional processor" should be added to the model, which selects the desired directions out of the estimators and marks signals from other directions (i.e. echoes) as interfering signals, which should be suppressed. Exceptions of the precedence effect can be explained as the taking of a new desired direction.

Multiple images, which arise when interaural time and intensity differences do not match (GAIK[1]), can be interpreted by the model as differences in the directional estimations out of phase and level differences.

Technical applications of a directional filter can be directional selective hearing aids, directional selective front ends for speech processing systems (speech recognizer, hands-free-telephones) or a low frequency supplement to beam microphones and microphone arrays.

---

1 GAIK(1990); Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale; Fortschritts-Berichte VDI, Reihe 17: Biotechnik, Nr.63; VDI-Verlag, Düsseldorf

2 LINDEMANN(1986): Extensions of a binaural cross-correlation model by contralateral inhibition; JASA 80; p.1608

3 SLATKY (1990); Lokalisation simultan abstrahlender Schallquellen: Konsequenzen für den Aufbau binauraler Modelle; Fortschritte der Akustik DAGA'90, Wien; DPG-Verlag, Bad Honnef, Germany, p.751

4 Based on::SLATKY(1991); Ein binaurales Modell zur Lokalisation und Signalverarbeitung bei Darbietung mehrerer Schallquellen; Fortschritte der Akustik DAGA'91, Bochum; DPG-Verlag, Bad Honnef, Germany