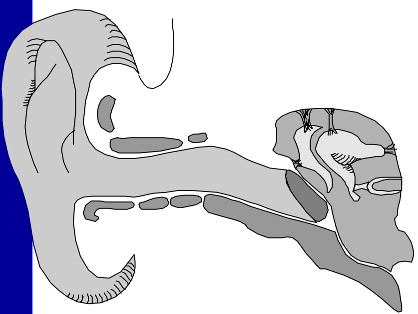
HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

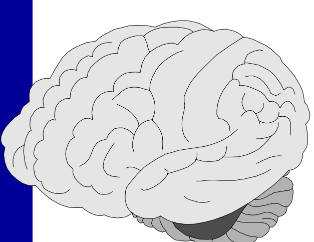
Prof. Dr. rer. nat. Dr. med.

Birger Kollmeier*

Carl von Ossietzky Universität Oldenburg
Medizinische Physik



Auditory principles in speech processing – do computers need silicon ears ?



* with contributions by V. Hohmann, M. Kleinschmidt, T. Brand, J. Nix, R. Beutelmann, and more members of our medical physics group

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

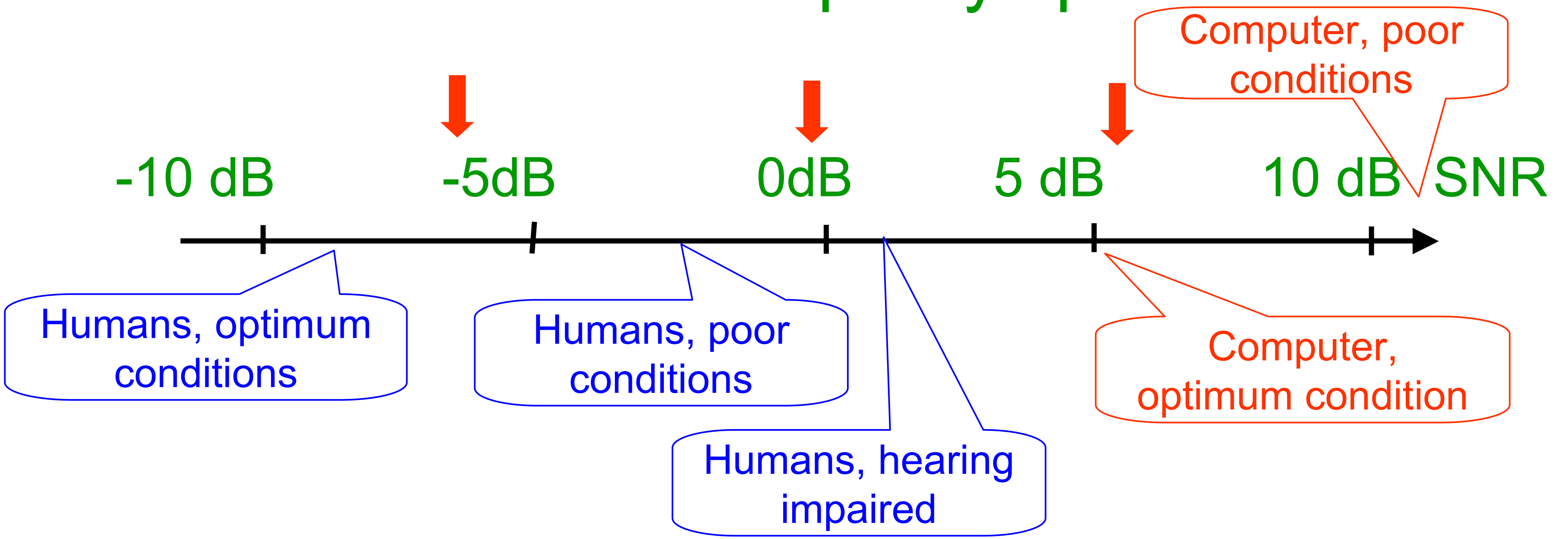
A quick test of your Cocktail-party-processor....

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Can you understand this sentence?

(Three repetitions, increasing signal-to-noise ratio)

Rachel has seven pretty spoons



HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

Outline

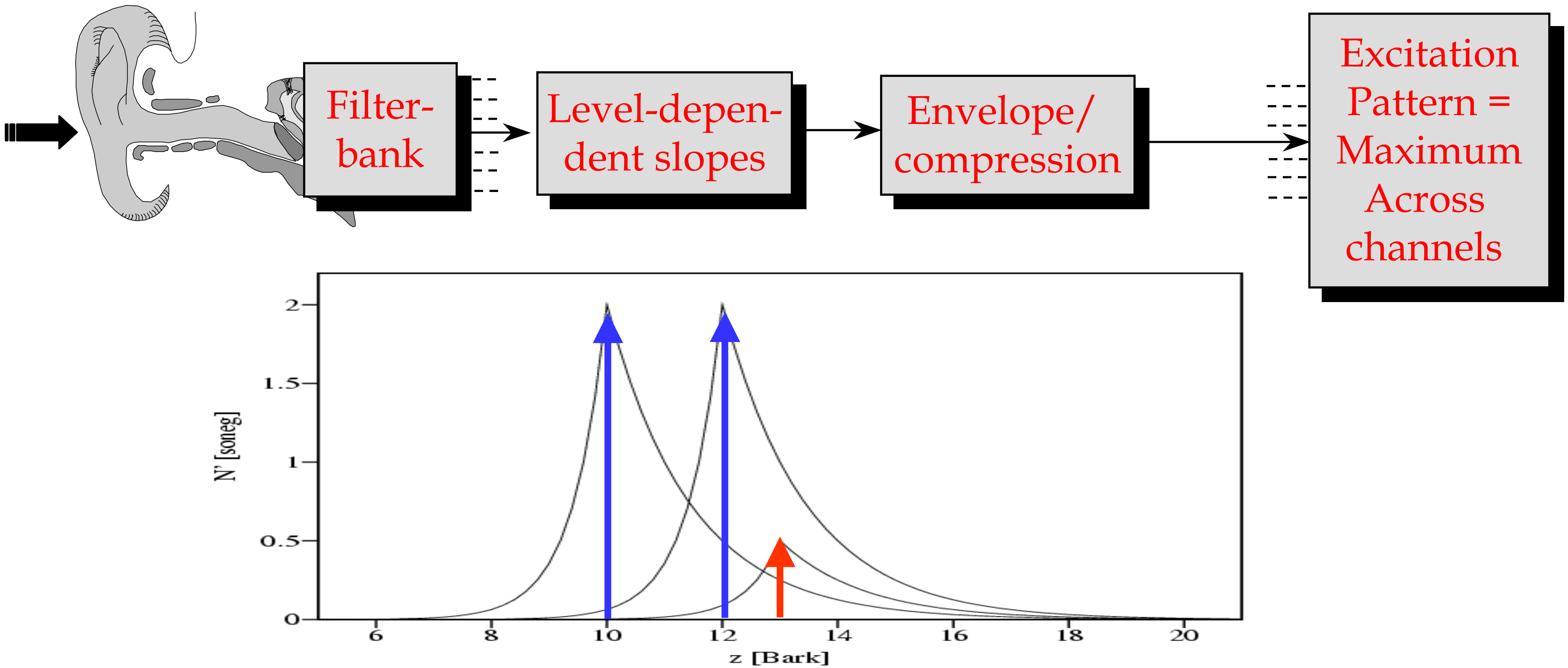
CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

- Auditory principles already „in silico“
- Additional properties not yet exploited
- Auditory models
- Modulation processing
- Binaural information processing
- ...why it matters not only for hearing aids

Auditory properties used in speech processing systems

- Logarithmic/compressive intensity units (dB, loudness, cepstrum)
- Quasi-logarithmic Mel/bark-scale as frequency scale
- Linear predictive coding with quantization noise masked by speech (Schroeder & Atal)
- Masking model for HiFi-coding (MPEG, Brandenburg)
- Speech coding & Audio quality objective assessment (Beerends & Stemerdink)
- RASTA techniques for ASR (Hermansky& Morgan)

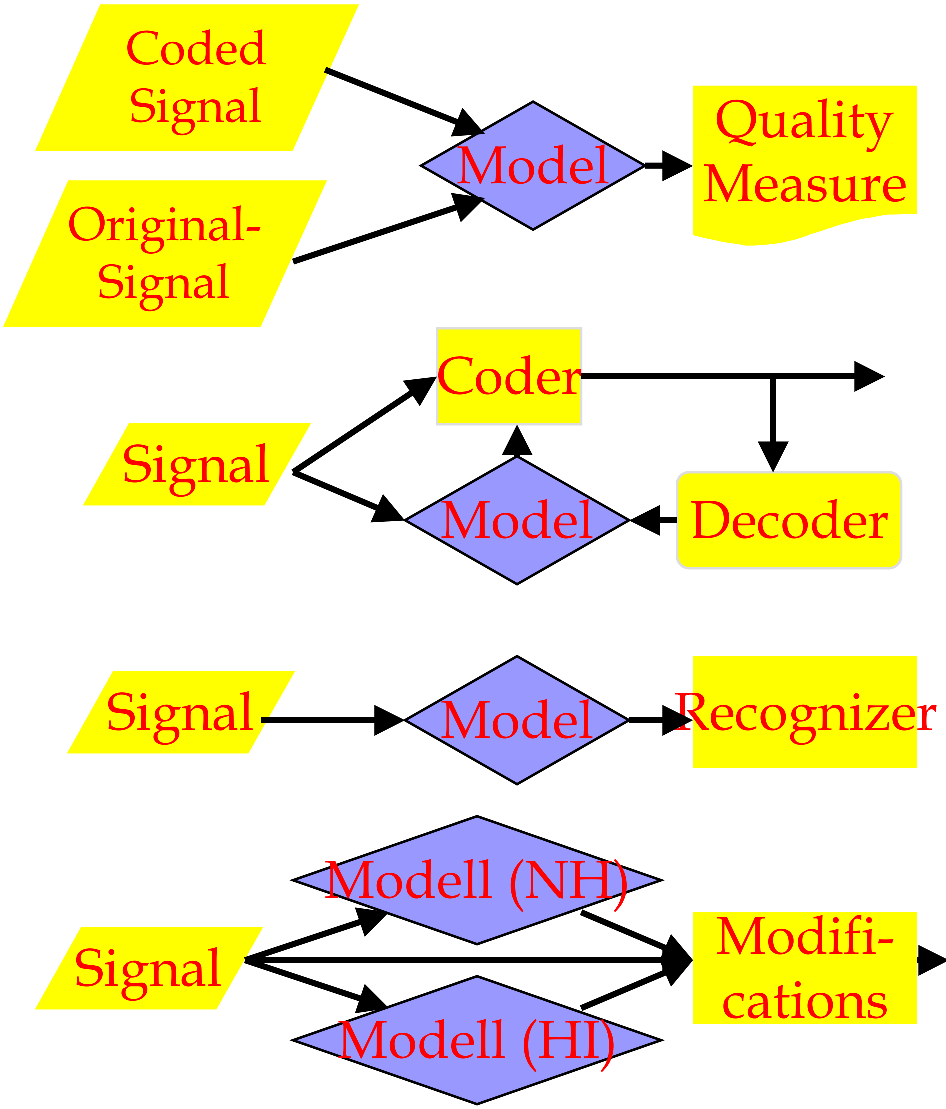
Spectral Masking models



- Classical approach: Loudness model (Fletcher, Stevens, Zwicker & Fastl, Moore)
- Forward & backward masking produce temporal sluggishness

Application of auditory models

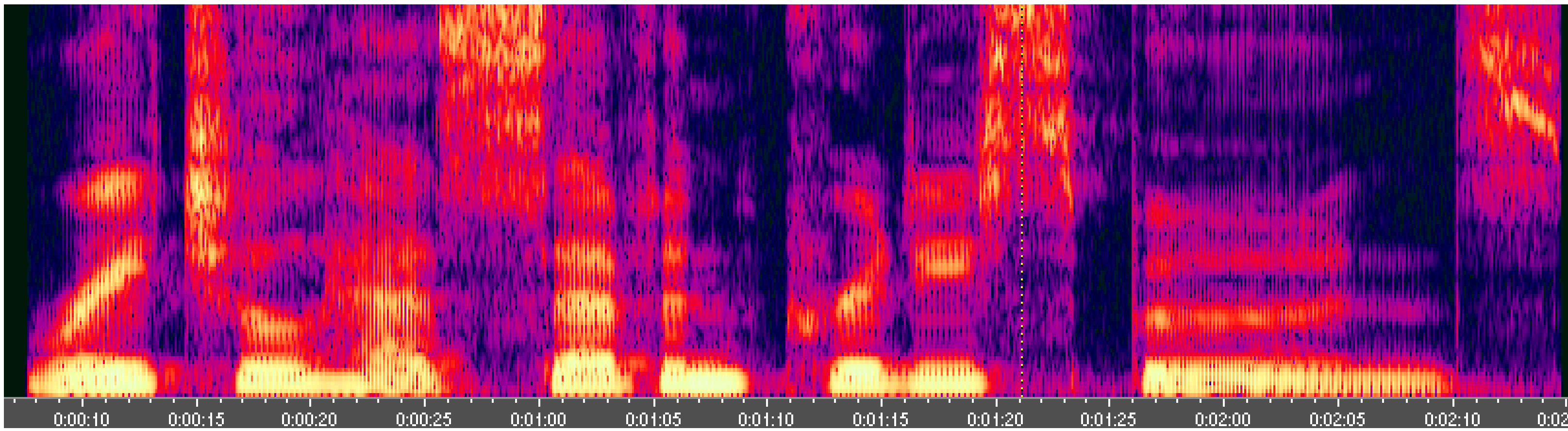
- Assessment of signal quality (Cellular phone networks,....)
- Signal coding (MP3, MiniDisc,..)
- Speech & pattern recognition
- Hearing aids



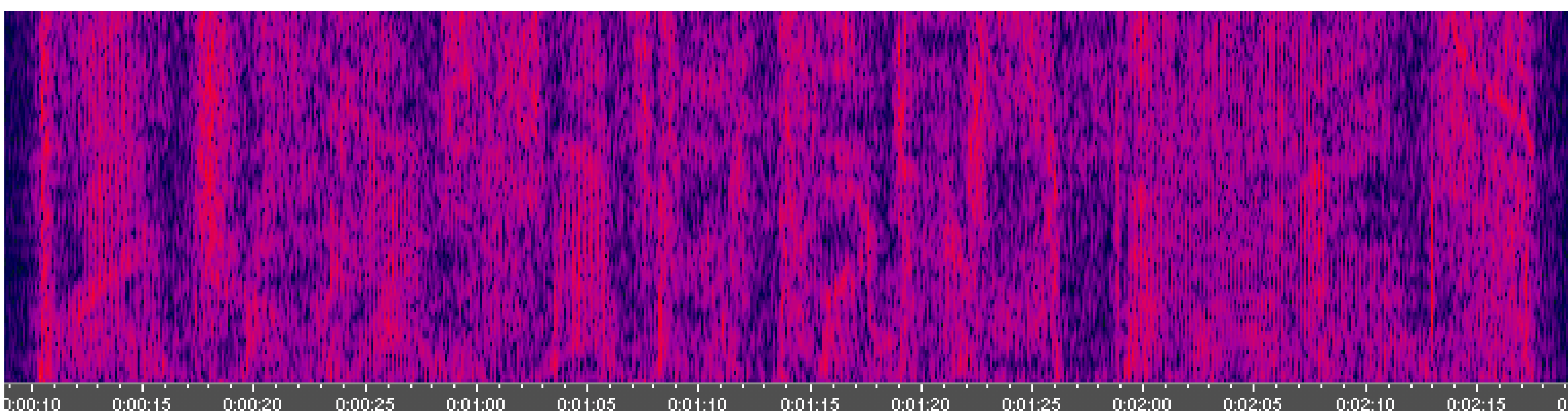
What do we miss when applying classical models?

- Temporal processing
- Modulation processing
- Spectro-temporal modulation processing
- Binaural/spatial processing
- Cognitive effects (interpolation, suppression)

Speech perception without a spectrum?



Flat spectrum (phase only) speech using the Oldenburg sentence test



Ear performs temporal analysis

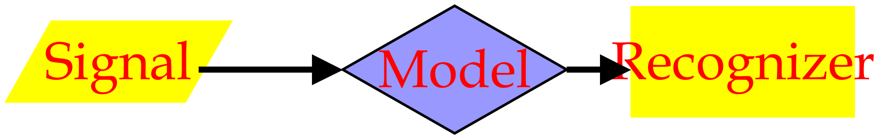
Temporal processing used for robust speech recognition

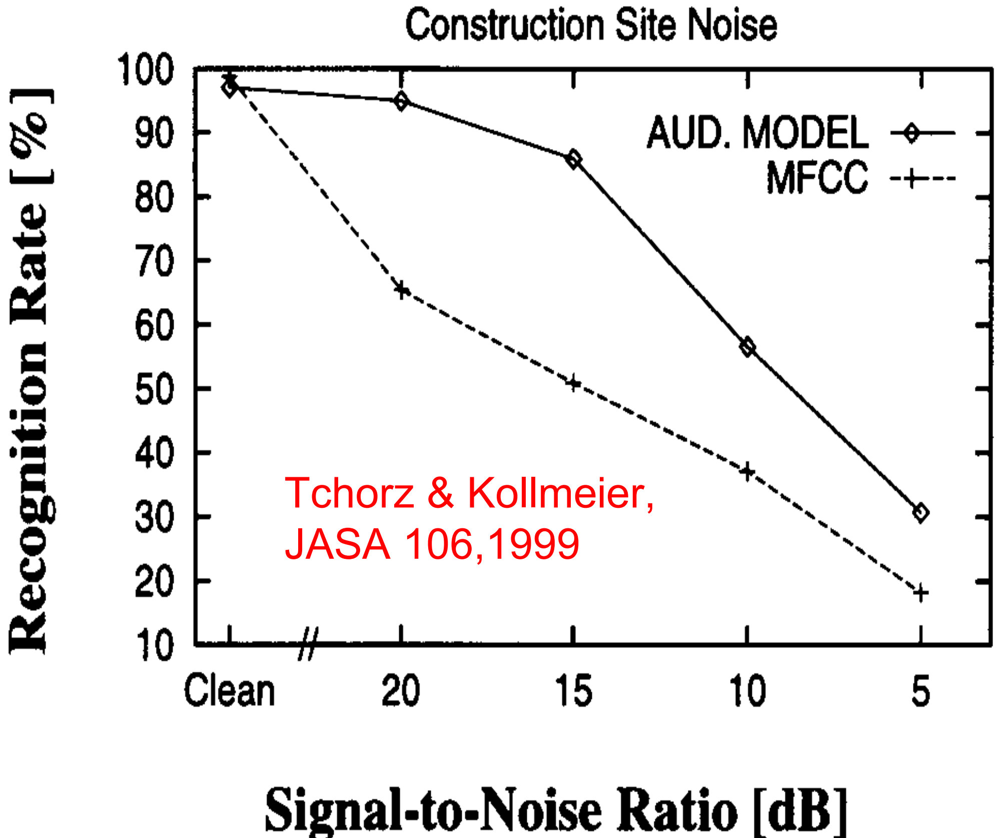
HörTech Kompetenzzentrum für Hörgeräte-Systemtechnik

Carl von Ossietzky Universität Oldenburg

Medizinische Physik

- Hermansky & Sharma: TRAPS
- Perception model

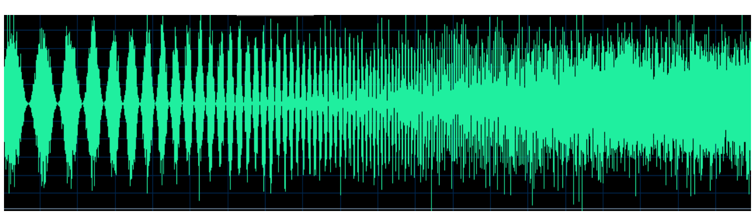




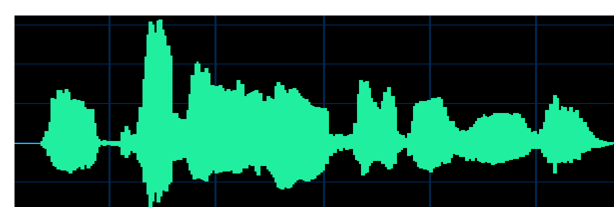
„Auditory front end for speech recognizer": Robustness against noise

What do we miss when applying classical models?

- Temporal processing
- ➡ **Modulation processing**
- Spectro-temporal modulation processing
- Binaural/spatial processing
- Cognitive effects (interpolation, suppression)



Sinusoidally Amplitude modulated noise



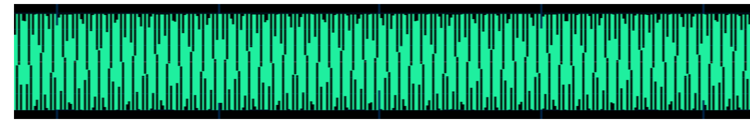
Spoken sentence

Asymmetry in masking: Which is the „better“ masker?

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

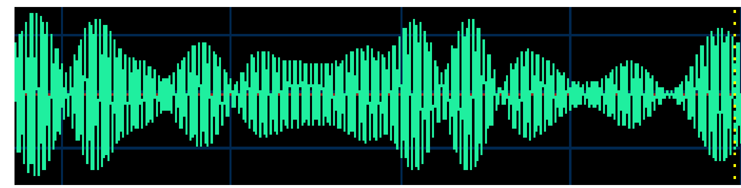
CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Tone 2 kHz



Narrow-band noise

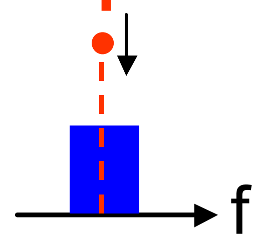
2 kHz, 256 Hz bandwith



Count the audible steps!

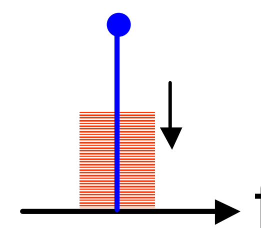
Tone in steps masked

by continuous noise



Noise in steps masked

by continuous tone



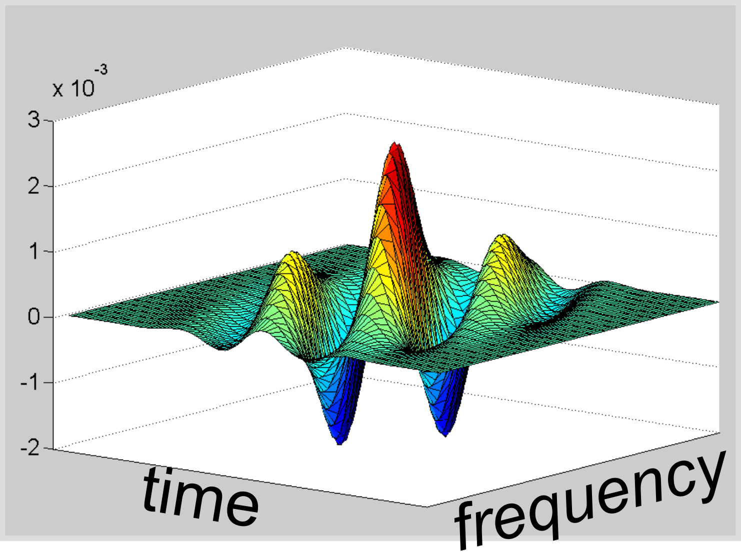
Ear performs detailed envelope analysis (modulation spectrum)

What do we miss when applying classical models?

Medizinische Physik

- Temporal processing
- Modulation processing
- Spectro-temporal modulation processing
- Binaural/spatial processing
- Cognitive effects

Gabor spectro-temporal feature

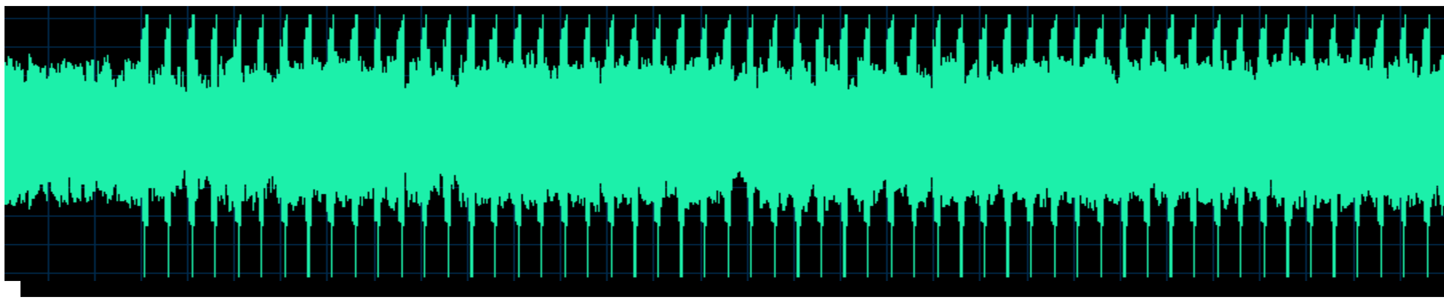


Continuity illusion: Can we trust our ears?

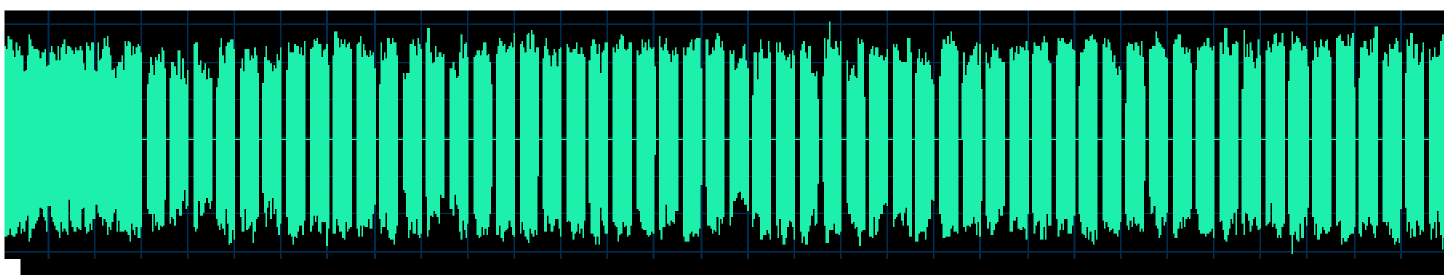
HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

Carl von Ossietzky Universität Oldenburg
Medizinische Physik

Music + pauses + noise (+6dB)



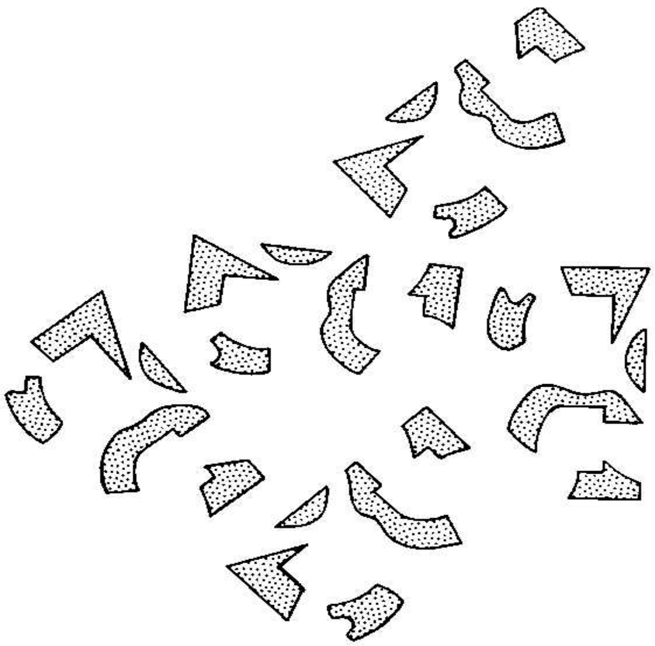
Do you hear ongoing music?

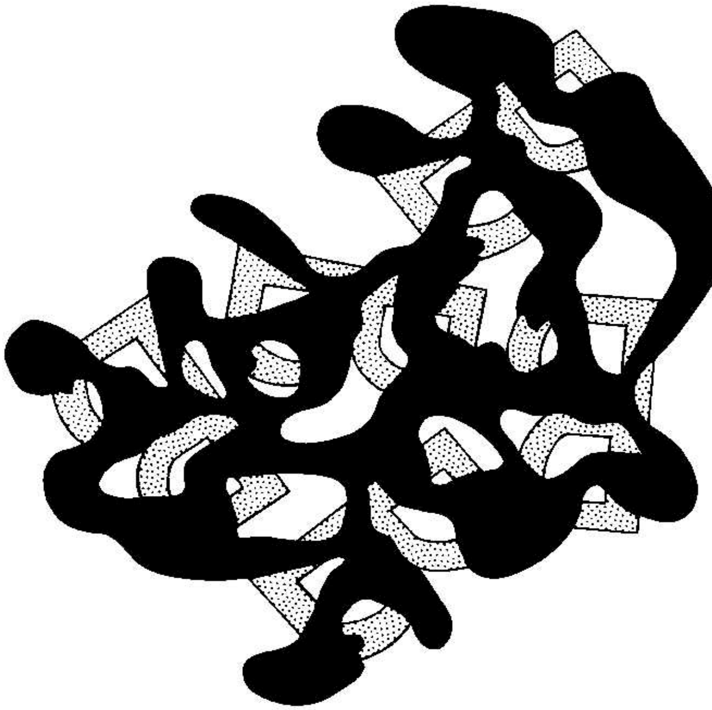


Music & pauses (500/125 ms)

➡ top-down processing by our brain

Visual analogy





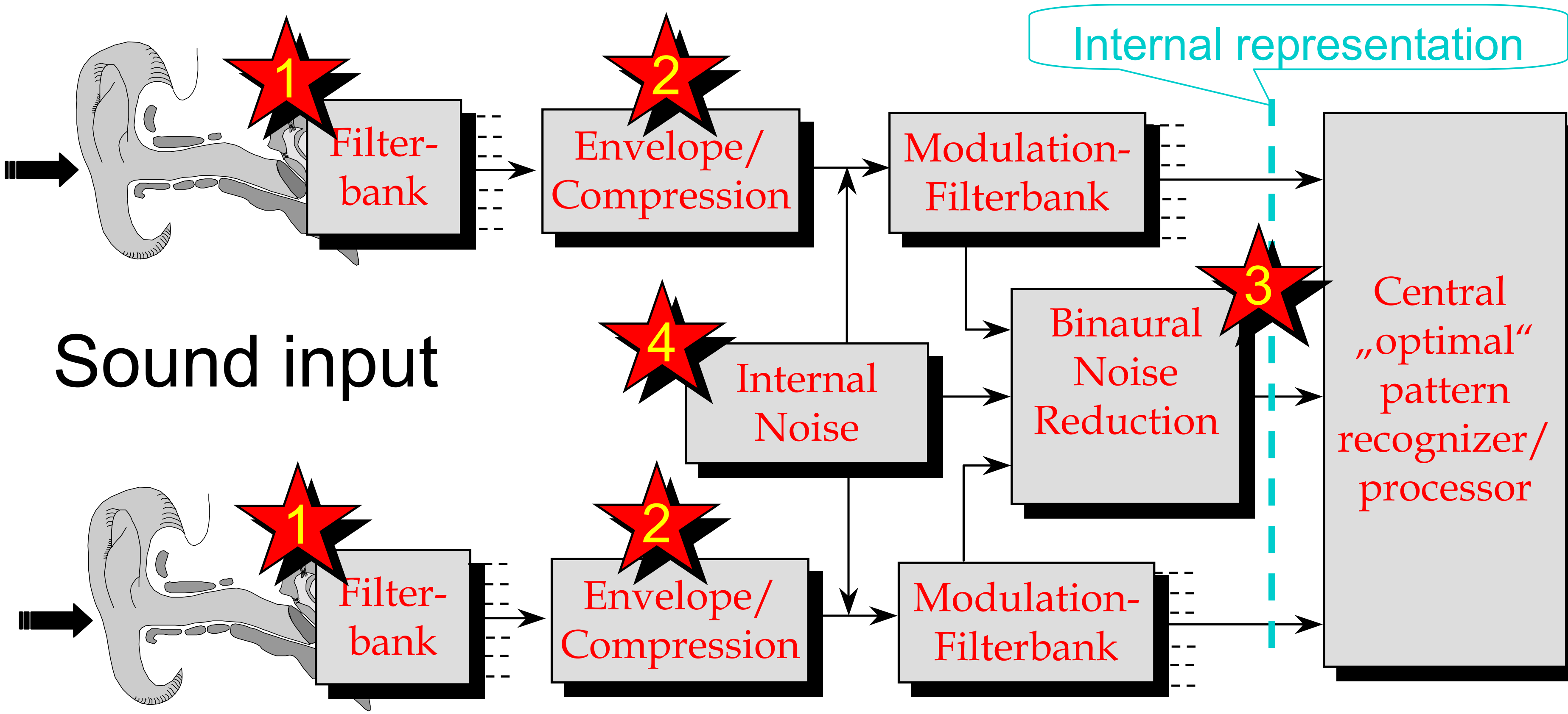
Bregman´s Bs

What do we miss when applying classical models?

- Temporal processing
- Modulation processing
- Spectro-temporal modulation processing
- Binaural/spatial processing
- Cognitive effects (interpolation, suppression)

How can we quantify these effects and put them to work in speech processing?

Model framework



Model of the „effective“ processing in the auditory system *& Impairments* 1 4

Dau, Kollmeier& Kohlrausch, 1997, Zerbs, 2000. Kollmeier, 2000, Derleth et al., 2001

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

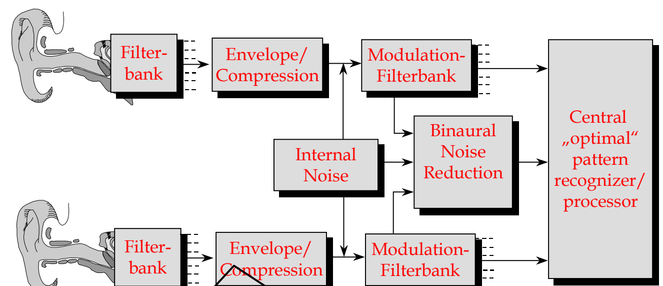
Approach: Analysis by model & simulation system

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Acoustical scene

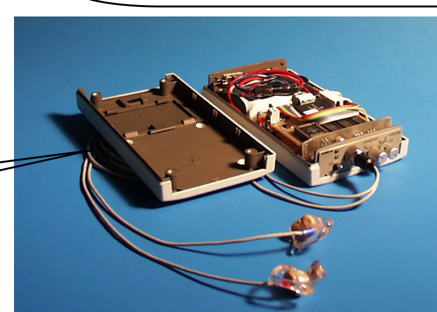
Human listener

Test result



Predicted Test result

Auditory model



Hearing Aid/ computer

System perfor- mance

Working hypothesis

Auditory System exploits all available acoustical cues, employing

- Lossy Front End (quantified by auditory model with information compression):

  acoustical input ➡ „internal representation“

- Perfect Back End: central pattern recognition (limited only by „internal noise“)

➡ Technical „copy“ of auditory front end yields near optimum performance of technical system

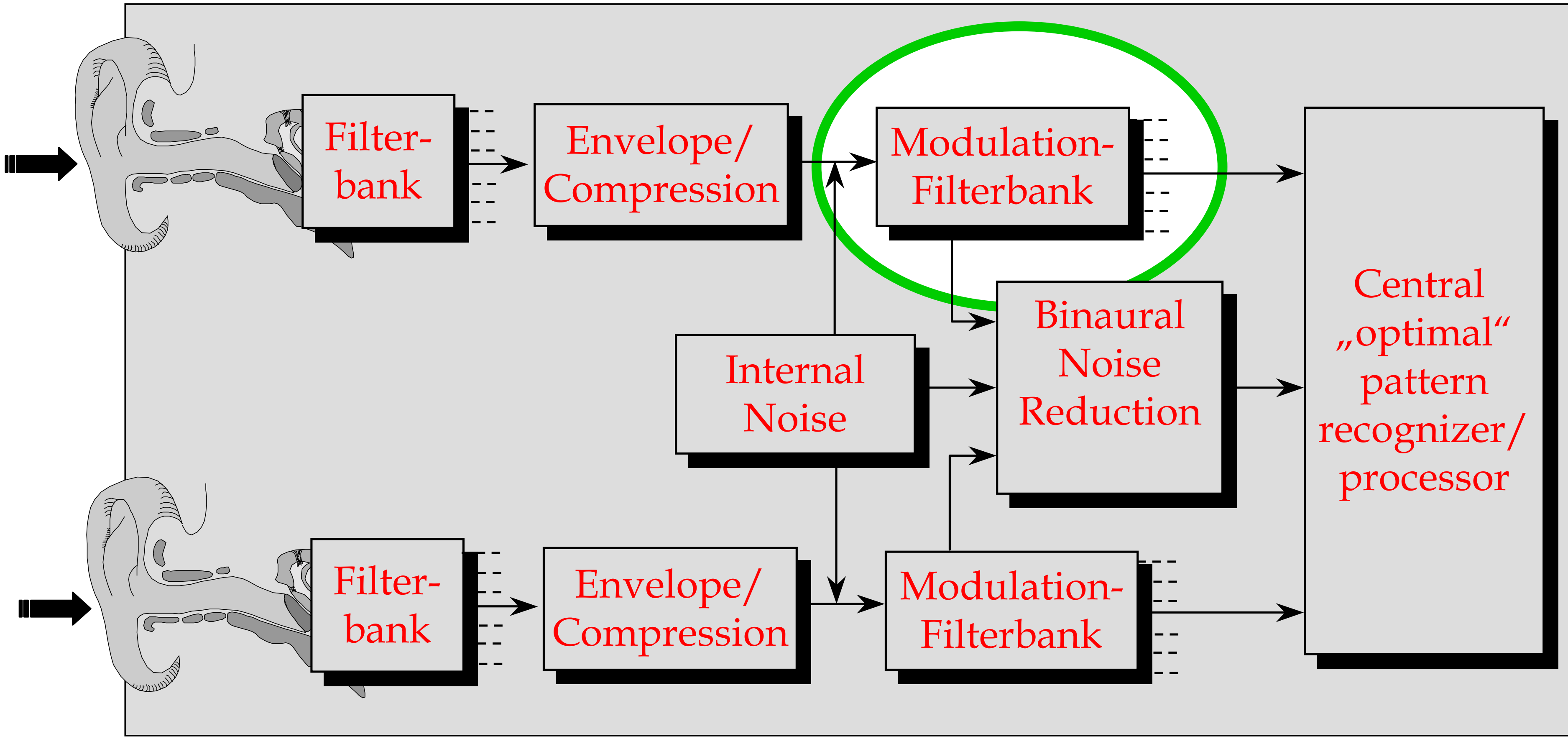
Outline

- Auditory principles already „in silico“
- Additional properties not yet exploited
- Auditory models
- Modulation processing
- Binaural information processing
- ...why it matters not only for hearing aids

Model framework: Modulation frequency analysis

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik



Model of the „effective“ processing in the auditory system

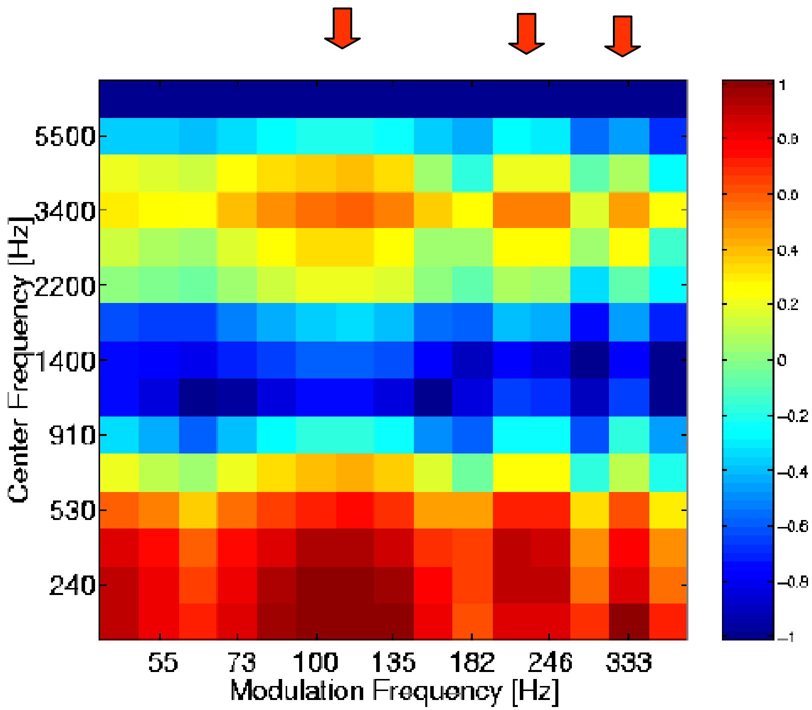
Dau, Kollmeier& Kohlrausch, 1997, Zerbs, 2000. Kollmeier, 2000, Derleth et al., 2001

Modulation frequency selectivity

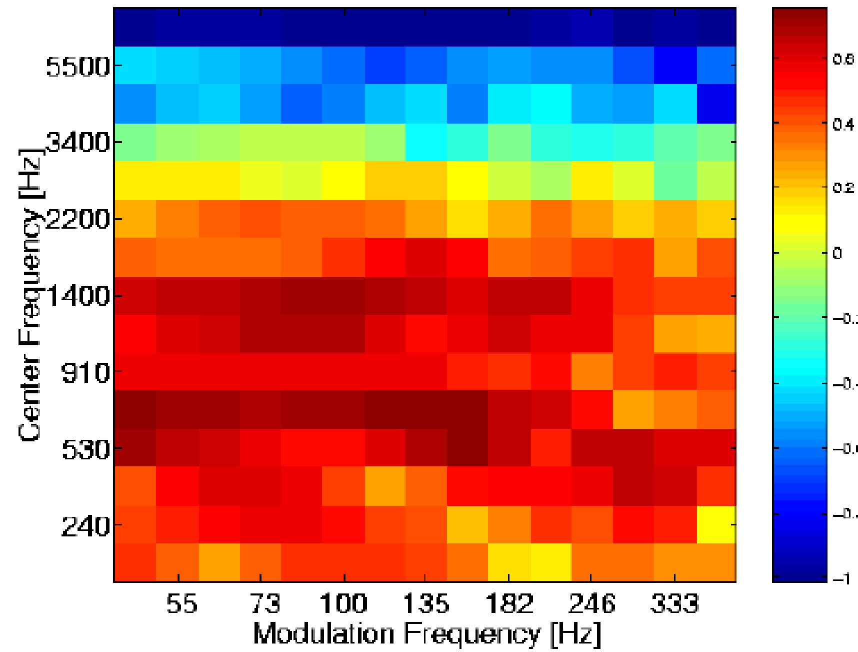
- Modulation-map in the auditory system (Langer & Schreiner)
- Psychoacoustics: Modulationfilterbank (Dissertation Dau, Dau et al., JASA 1997, Dissertations Verhey, Derleth, Ewert)
- Signal processing, noise reduction (Kollmeier & Koch, JASA 1994)
- Advantage: Separation of different auditory objects covering the same frequency region

Examples of Amplitude Modulation Spectrogram (AMS)

(voiced) speech

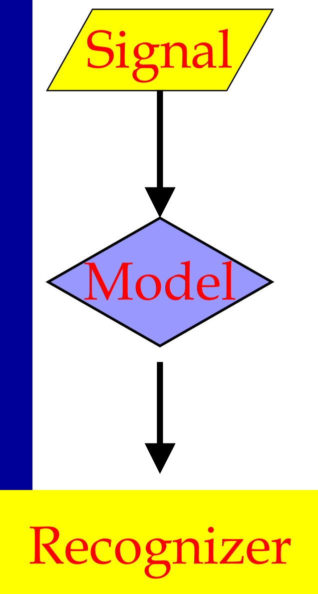


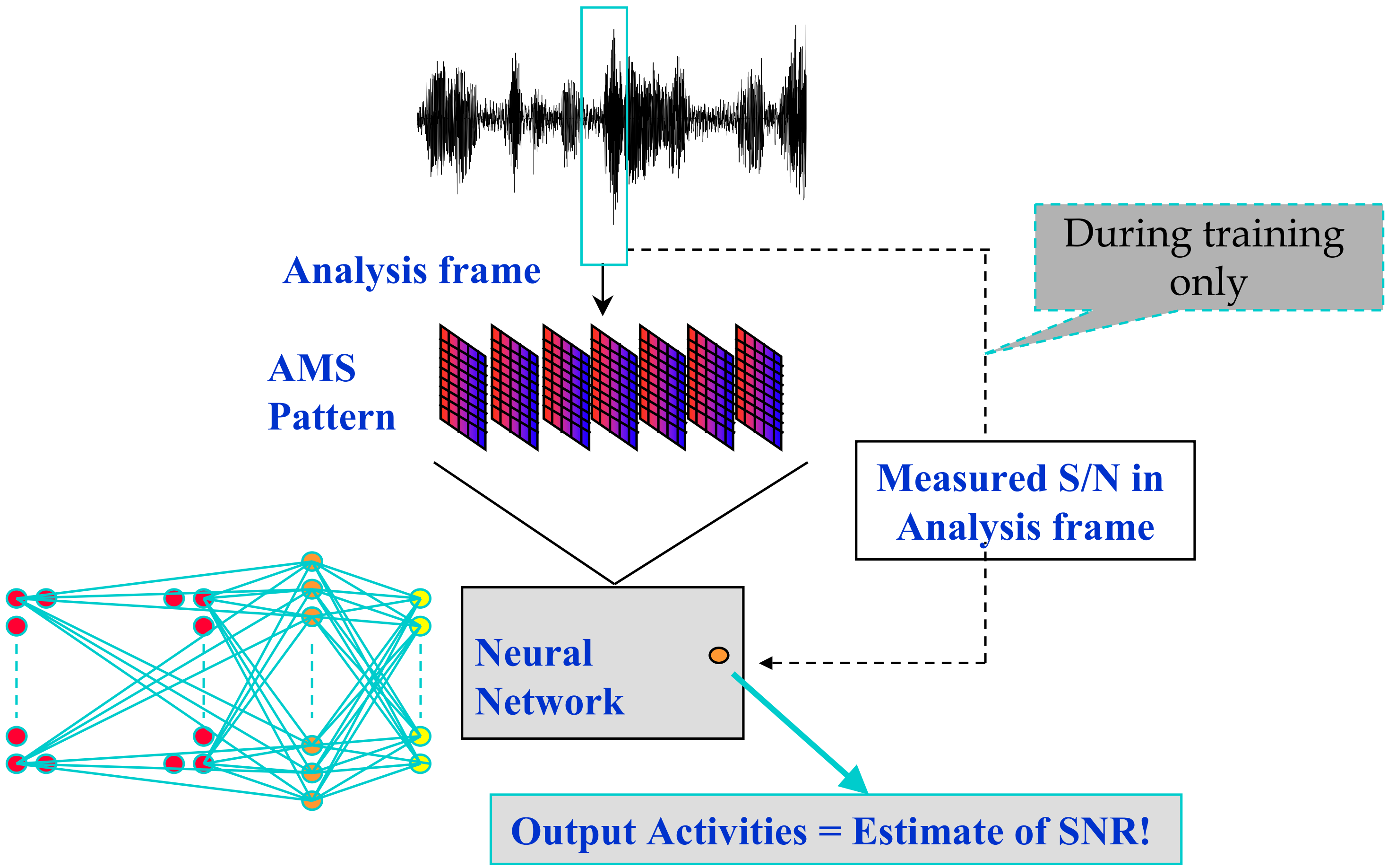
speech simulation noise



➡ Speech shows joint distribution in frequency/modulation frequency domain

SNR Estimation from Modulation spectrograms





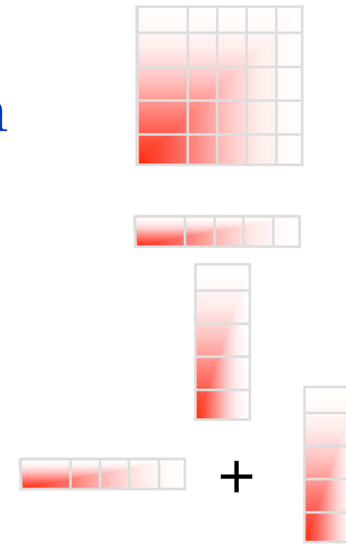
Speech-to-Noise Ratio estimate either broadband or multiple narrowband

Importance of two-dimensional features in modulation spectrogram

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

Carl von Ossietzky Universität Oldenburg
Medizinische Physik

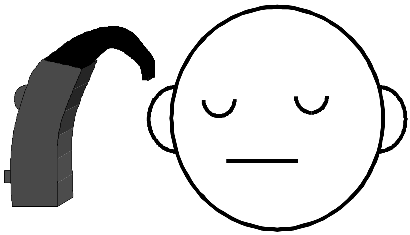
- Estimation error based on
  - Full 2-dim distribution — 5.2 dB
  - modulation spectrum — 6.6 dB
  - bark spectrum — 7.6 dB
  - combination of both — 5.8 dB



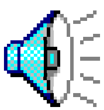
➡ Joint distribution of modulations and spectrum required!

Monaural Noise reduction using Amplitude Modulation cues

Suppression of a fluctuating background noise using AMS



Industrial noise with speech

unprocessed  processed

Improves speech recognizer in noise (Tchorz & Kollmeier, 2002)

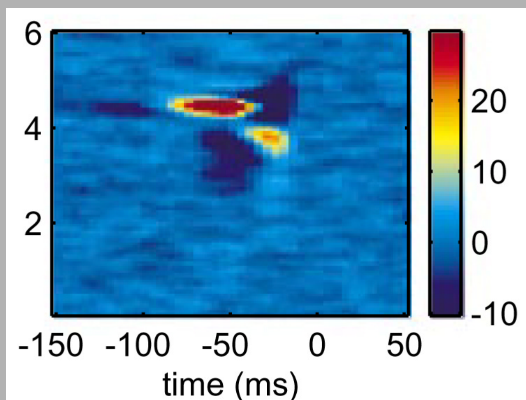
➡ Modulation frequency analysis is important for speech perception & promising for speech processing

Spectrotemporal features from neurophysiology

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

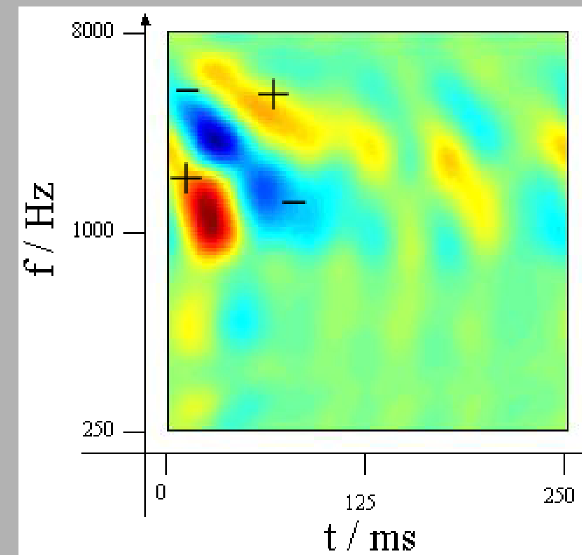
CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Receptive fields of cortical neurons



DeCharms et al. (1998)

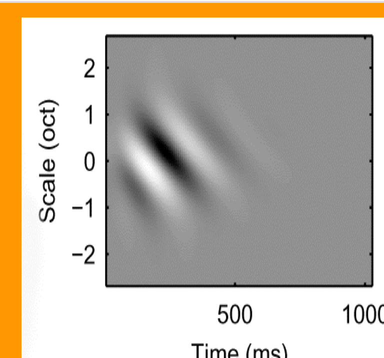
Depireux et al. (2000)



Indications for spectro-temporal feature extraction

Model of modulation perception

Chi et al. (1999)

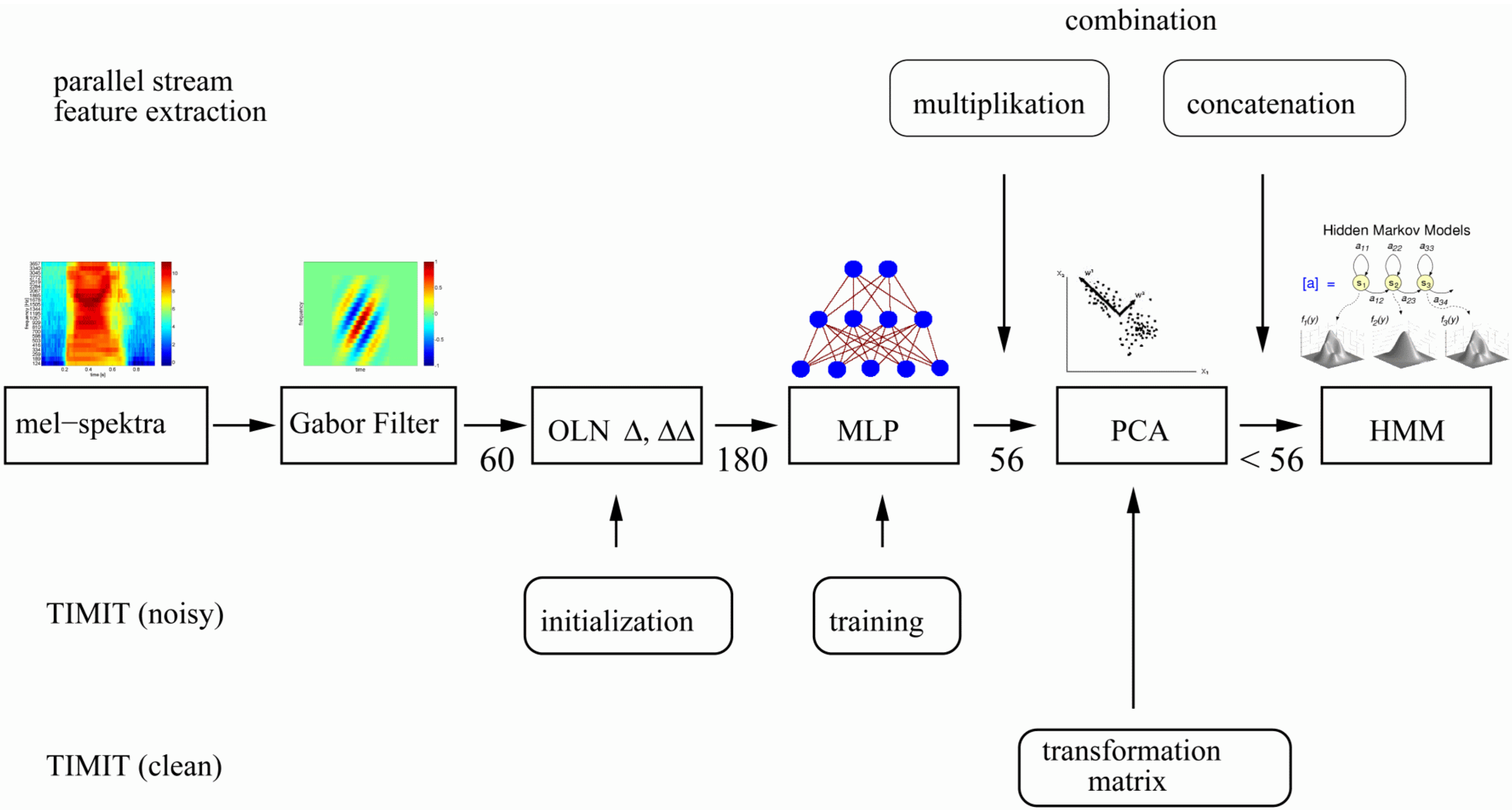


Gabor Tandem recognizer

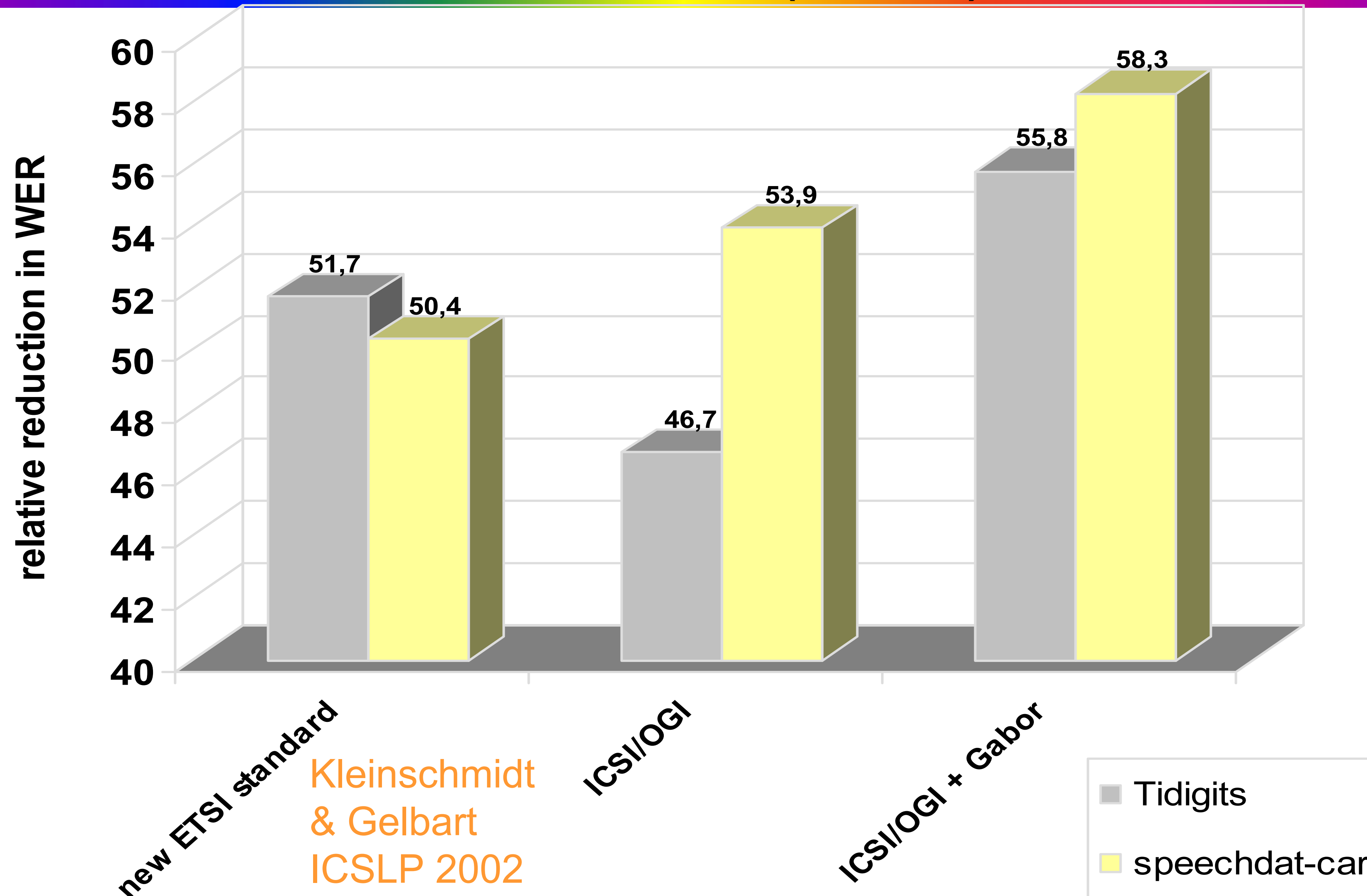
Medizinische Physik

More details: M. Kleinschmidt: 'Localized Spectro-temporal Features for ASR'

Eurospeech Session SThBb - Time is of the Essence, Thursday 10am, Room 2



Aurora: relative reduction in word error rate (WER)



Kleinschmidt & Gelbart ICSLP 2002

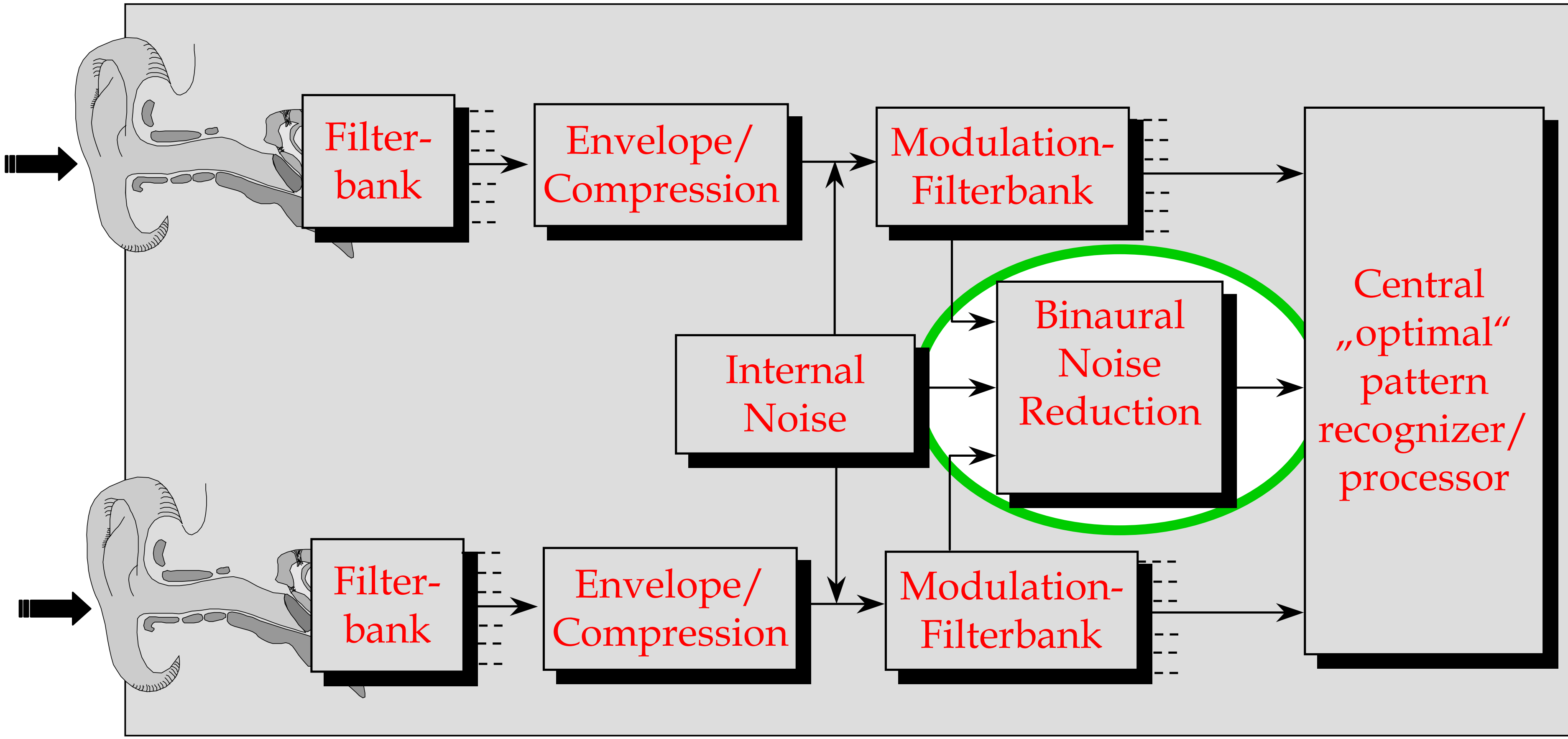
Outline

- Auditory principles already „in silico“
- Additional properties not yet exploited
- Auditory models
- Modulation processing
- Binaural information processing
- ...why it matters not only for hearing aids

Model framework: Binaural noise reduction

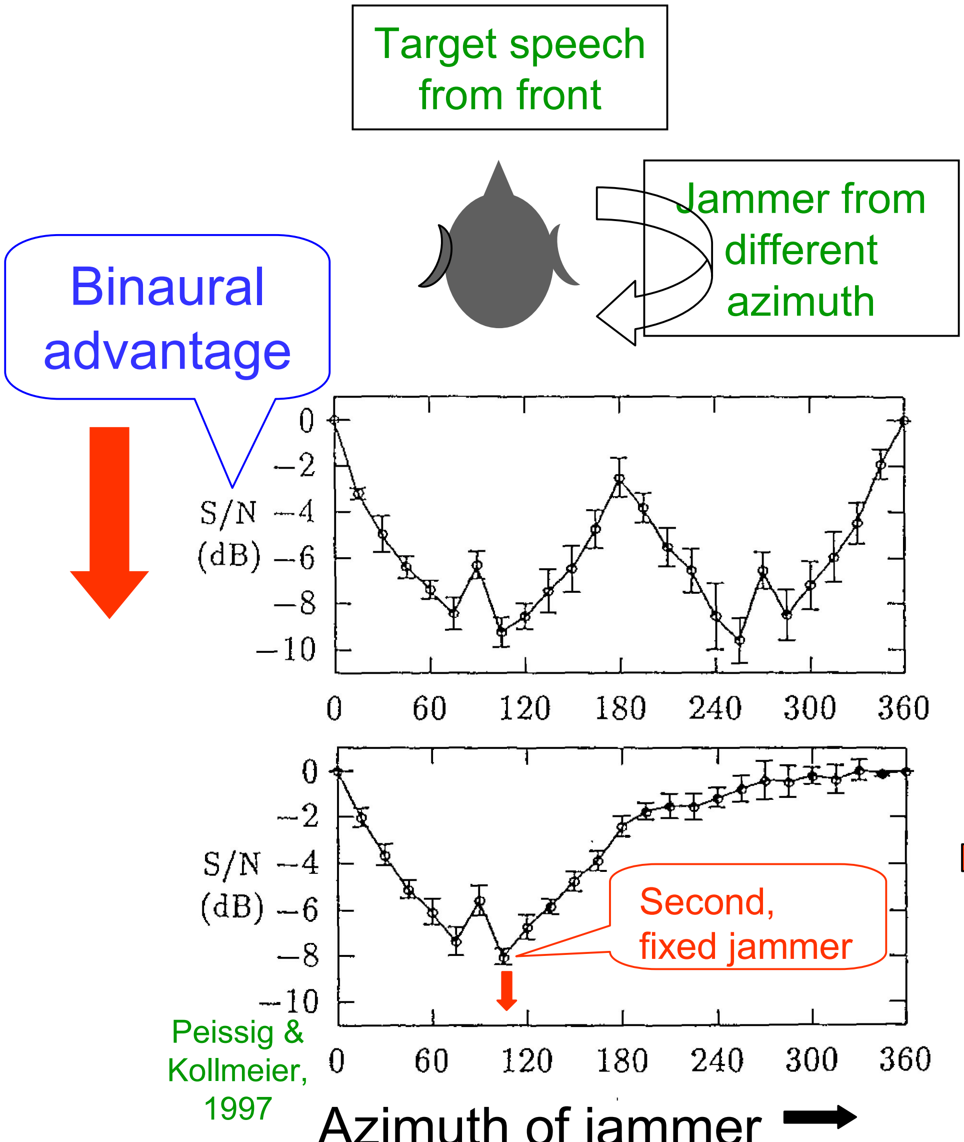
HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik



Model of the „effective“ processing in the auditory system

Speech Reception Threshold for different spatial arrangements



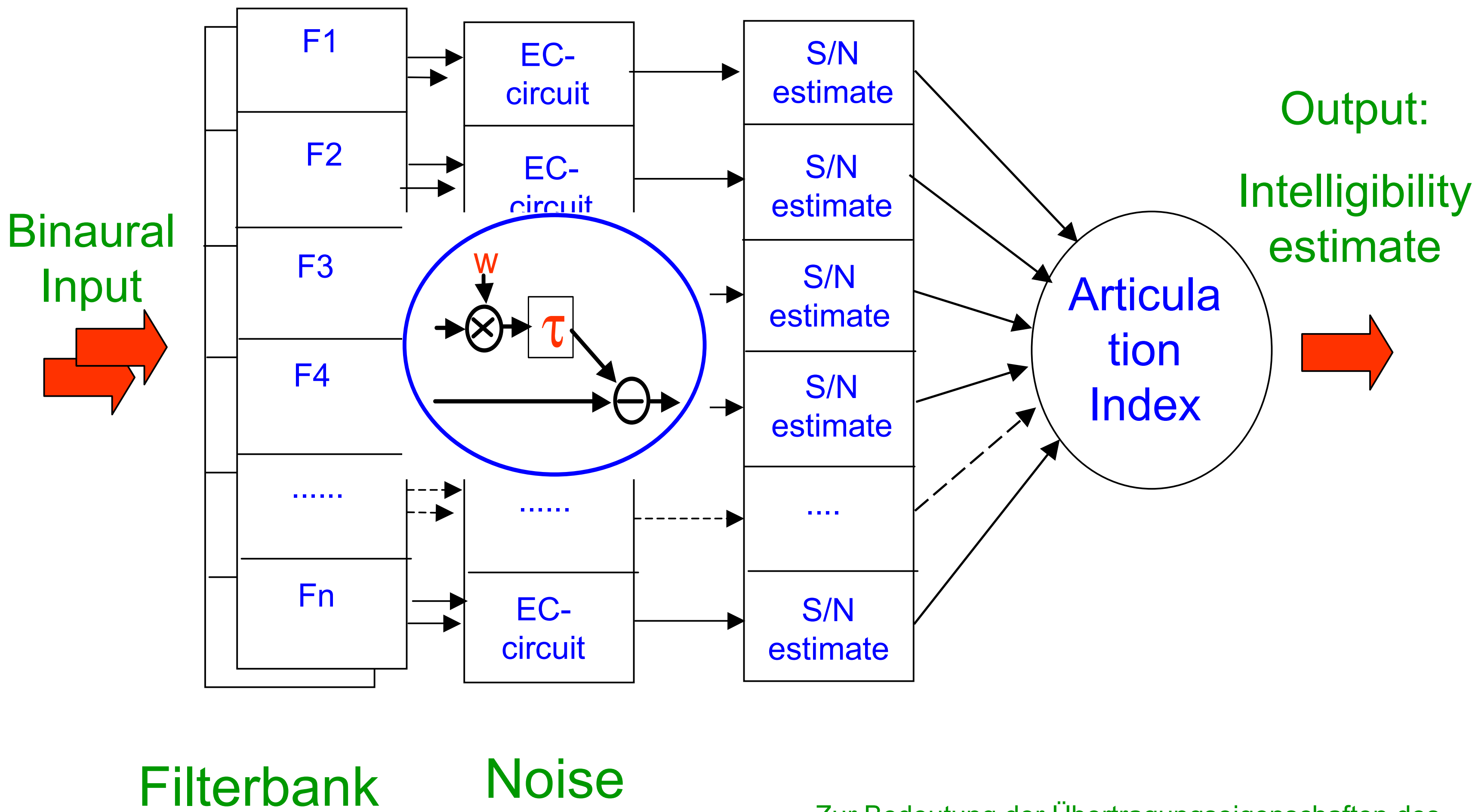
- One continuous jammer provides maximum effect
- Second, opposite jammer can not be cancelled simultaneously

➡ Binaural hearing operates like 2-sensor adaptive beamformer

v.Hövel-model (´84) model

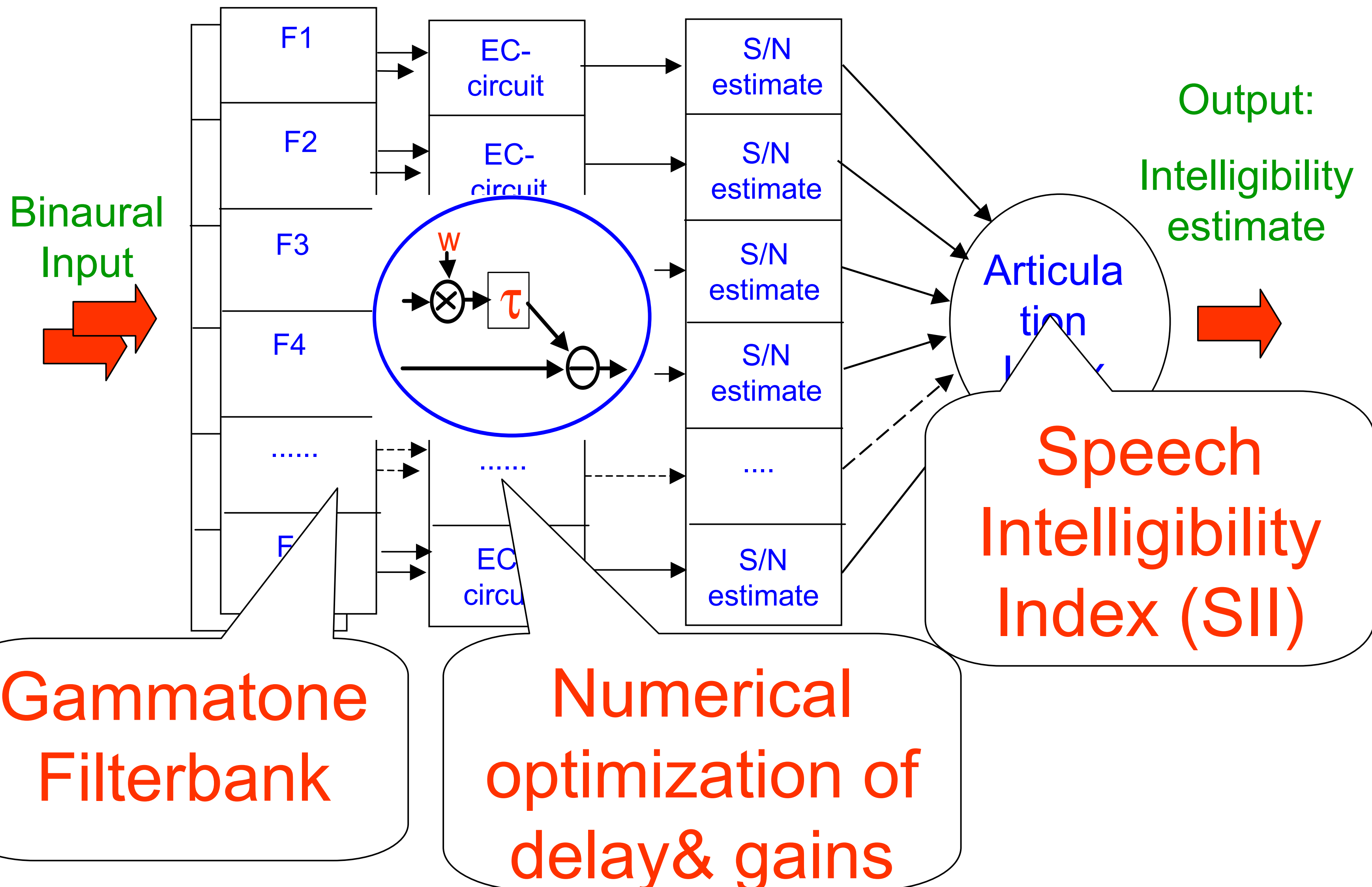
HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik



Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung, Dissertation, RWTH Aachen

v.Hövel-model (´84) model: modifications

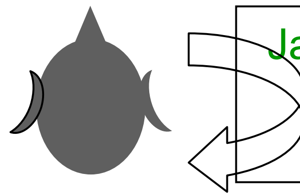


Speech Reception Threshold for different spatial arrangements

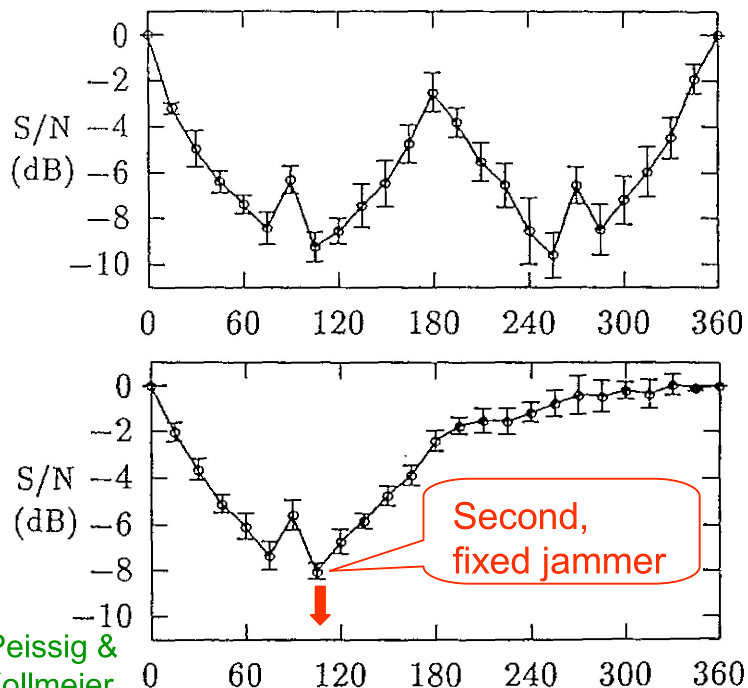
HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

Carl von Ossietzky Universität Oldenburg
Medizinische Physik

Target speech from front



Jammer from different azimuth

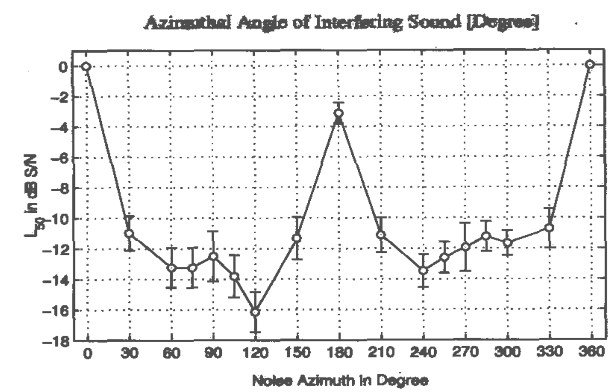


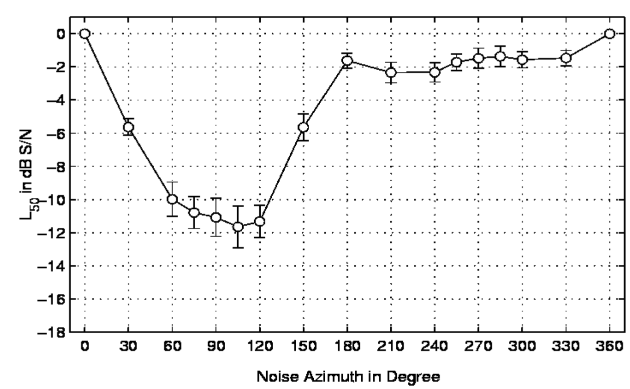
Peissig & Kollmeier, 1997

Azimuth of jammer

Predictions by binaural model

(R. Beutelmann, T.Brand)

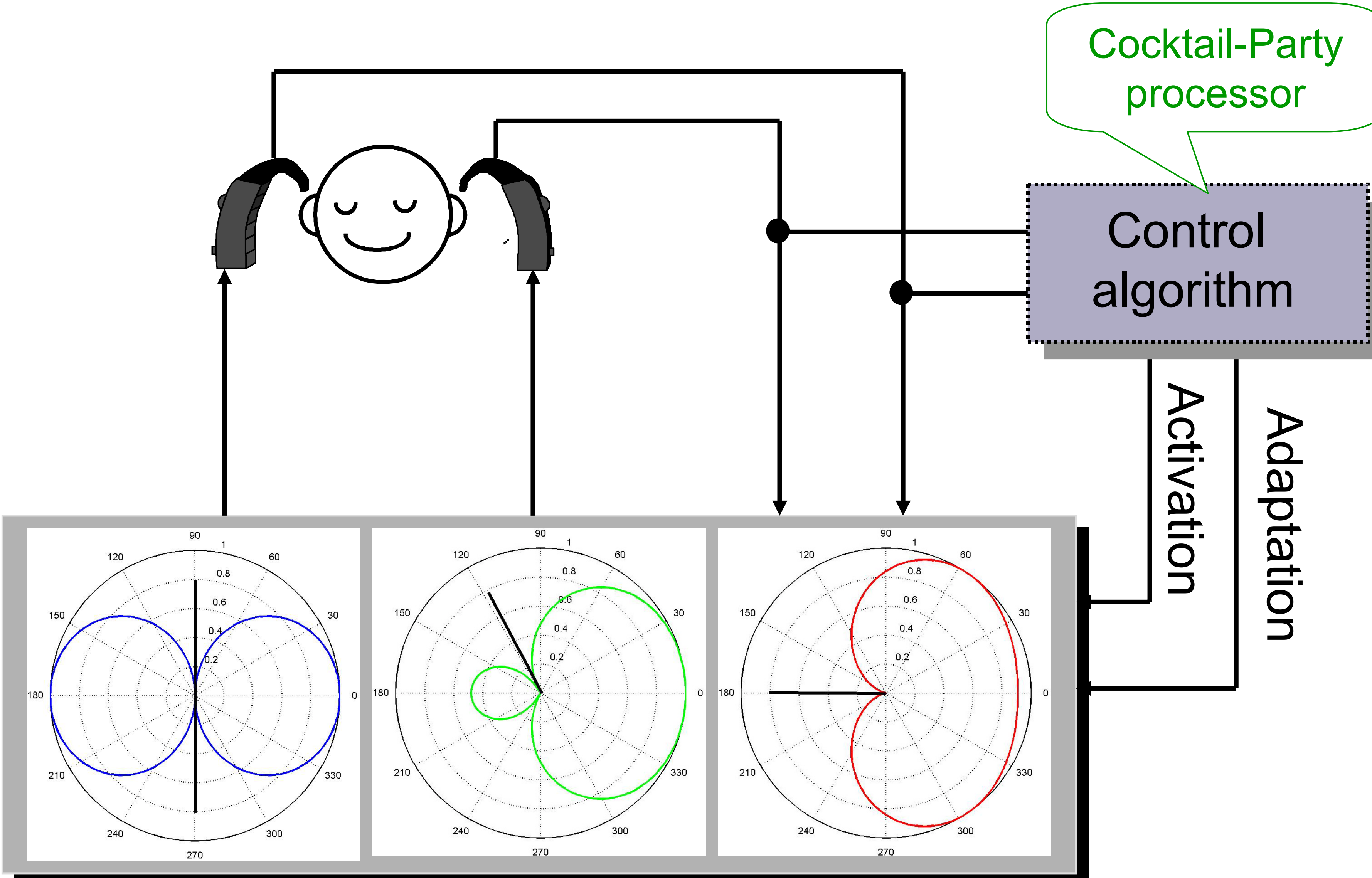




Corresponding binaural Beamformer hearing aid

HörTech Kompetenzzentrum für Hörgeräte-Systemtechnik

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik



HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

Performance of two-input „Cocktail party processors“

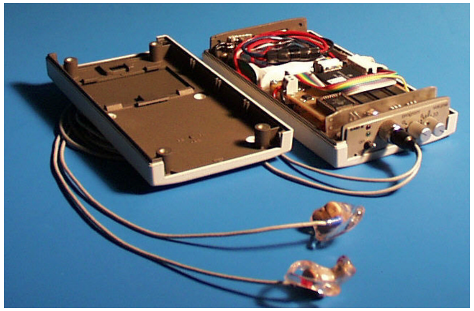
CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Blind Sound Source Separation (Anemüller&Kollmeier, 2002)

- Mixture of two sources in a room
- Separation of first source second source

Binaural situation-adaptive directional filter (Wittkop, 2000)

- One speaker in stationary noise
- One speaker from the front
  - + 3 interfering speakers
  - + Algorithmus

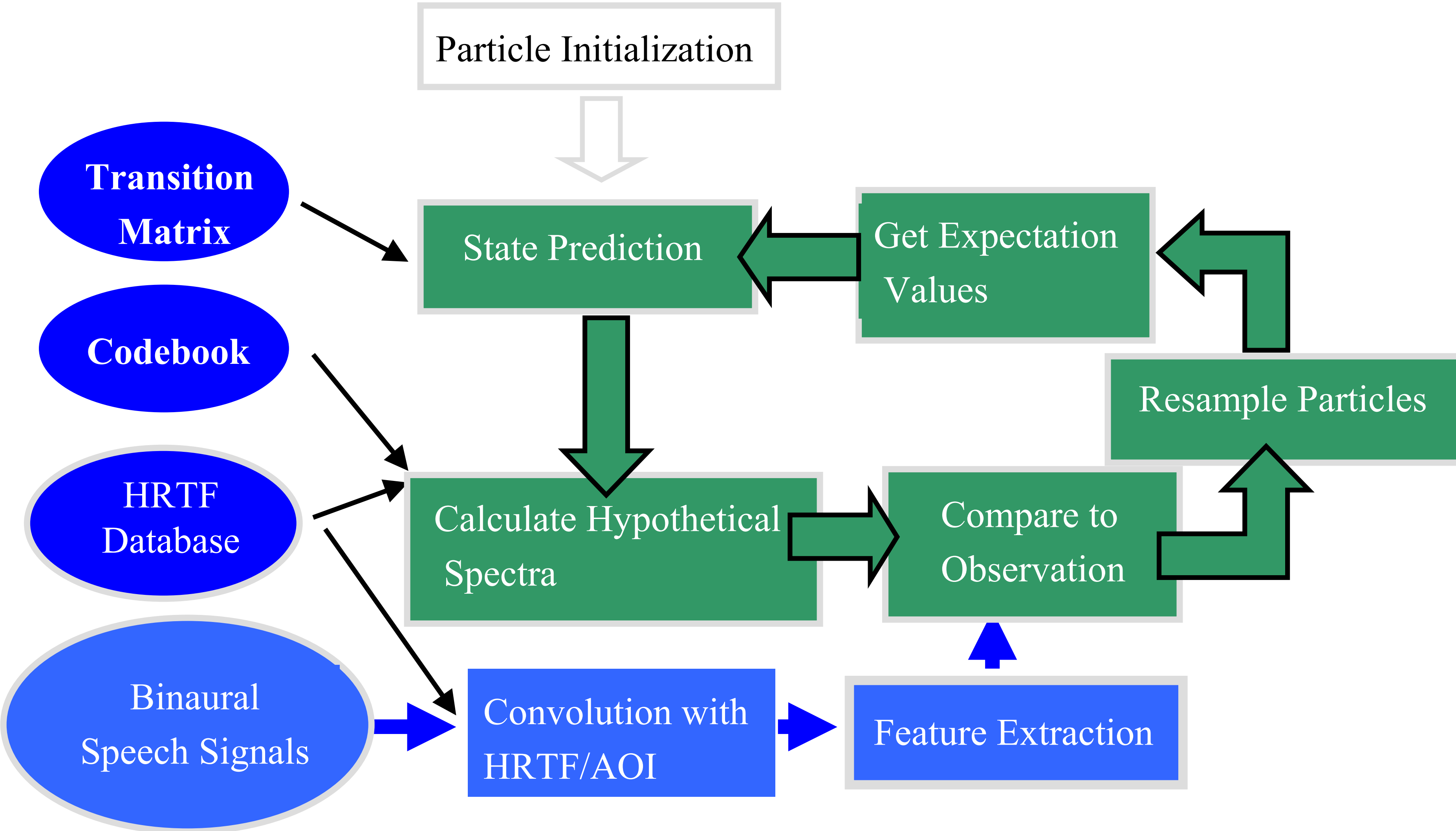


Localization model-driven beamformer (Nix & Hohmann, 2002)

- 2 Sources, unprocessed
- 2 Sources, processed, first direction // second direction
- 3 Sources, unprocessed//processed

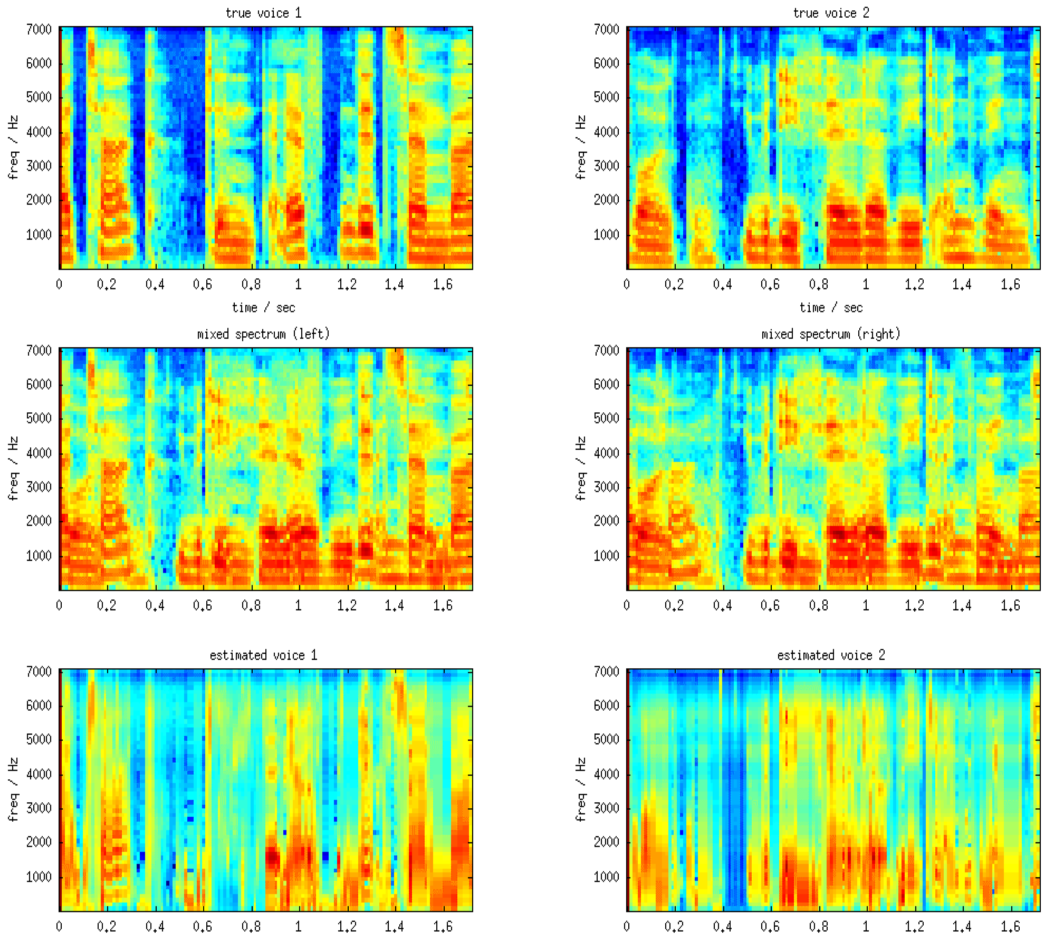
➡ No convincing separation of more than two sources

Particle filter to estimate best beamformer online



Details: Johannes Nix, M. Kleinschmidt, V. Hohmann: 'CASA by Using Statistics of High-Dimensional Speech Dynamics and Sound Source Direction, Eurospeech 2003 Session PTuDe - Speech Enhancement II, Tuesday 4pm, Main Hall, Level-1

Results: On-line Recovery of Spectral Envelopes



original voices

left/right ear signal

estimated voices (left side)

(see: Nix, Kleinschmidt, Hohmann, Tuesday 4pm)

Separation from multiple sources with much computation & „cognitive“ complexity!

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

Conclusions

CARL VON OSSIETZKY universität OLDENBURG
Medizinische Physik

Test result

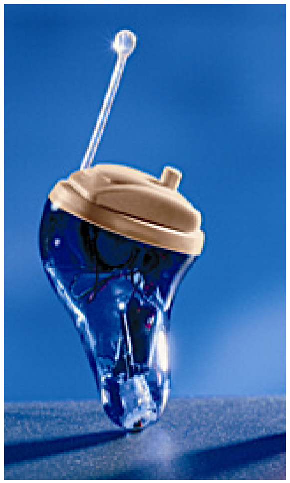
Predicted Test result

System perfor-mance

- Auditory principles for speech processing look promising
- Interaction experiment-model-application
- Amplitude Modulation Spectrogram !
- optimally switched two-sensor beamformer to mimic binaural system
- Top-down vs. bottom-up processing yet to be explored

Hearing aid or personal communication device?

Medizinische Physik



In-the-ear hearing aid

HörTech Prototype

Behind-the-ear hearing aid

- Technology for hearing aids and mobile phones converge
- ➡ Knowledge from hearing aid design is required for modern speech communication systems!

....thank you!

HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

Carl von Ossietzky Universität Oldenburg
Medizinische Physik

A binaural hearing aid to sit in –

The ultimate way of achieving a good performance in cocktail parties!

Auditory throne in front of the new„House of hearing“ (Oldenburg)