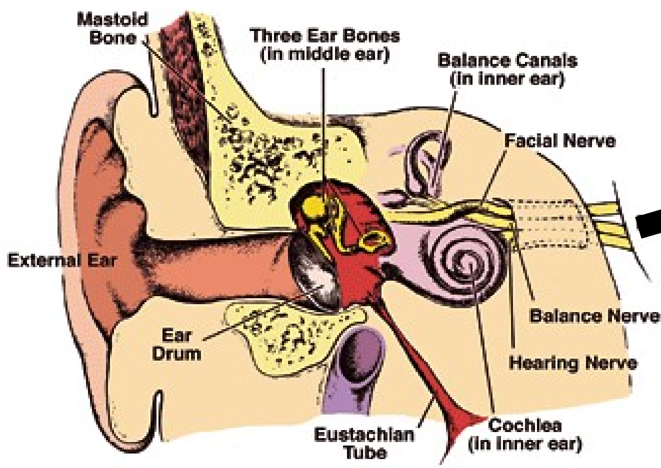
Modeling of Auditory Perception

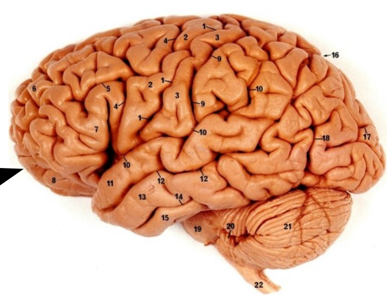
Bernhard Laback and Piotr Majdak
http://www.kfs.oeaw.ac.at
Bernhard.Laback@oeaw.ac.at

Physiological Basics

- Four main stages of the auditory system:
  - Outer ear
  - Middle ear
  - Inner ear (cochlea) and auditory nerve
  - Central processing stages



[Courtesy of the House Ear Institute]



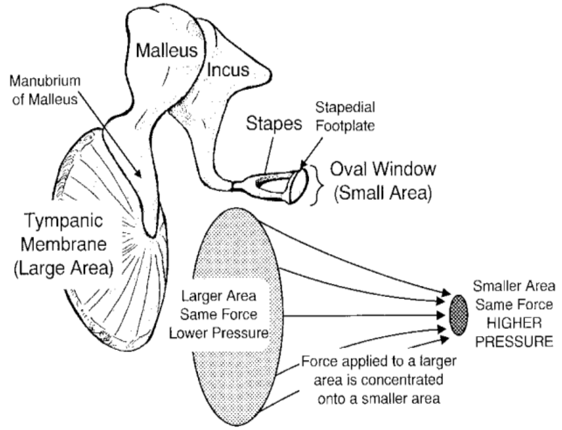
[Ellis et al., 1991]

Outer Ear

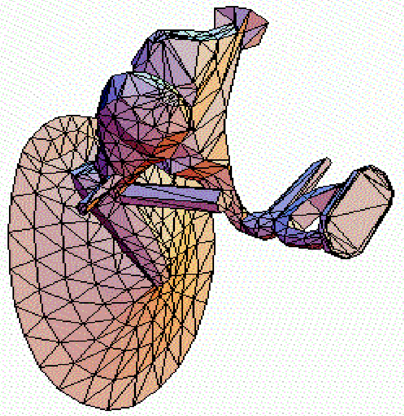
- Pinna:
  - Sound amplification
  - Direction-dependent filtering at high frequencies (> 4 kHz)
    - Important for sound localization in vertical planes (front/back, up/down dimensions)
    - Requirement for “externalization” of sounds
- Auditory canal:
  - Maximum in transfer function for pinna + auditory canal from 1.5 to 5 kHz

Middle Ear

- In tympanic cavity (air-filled)
- Connects ear drum with oval window of cochlea via maleus, incus, and stapes
- Function: adjustment of impedance of waves in air versus liquid
- Best transmission from 0.5 – 4 kHz



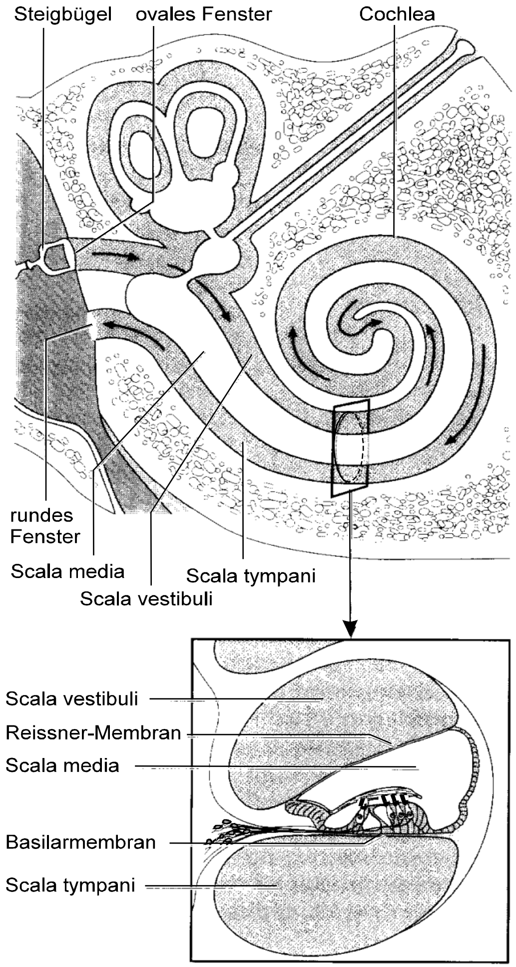
[Gelfand, 1997]



Inner Ear (Cochlea)

- Tubelike, involuted structure, 35-mm length
- Three compartments: scala vestibuli (SV), tympani (ST), and media (SM)
  - Basilar membrane (BM) with Cortical Organ: sensory structures
- Oval window connects stapes with SV at “base”
- ST and SV connected at “apex”
- Electric potential difference (-40 mV) between SV/ST and SM

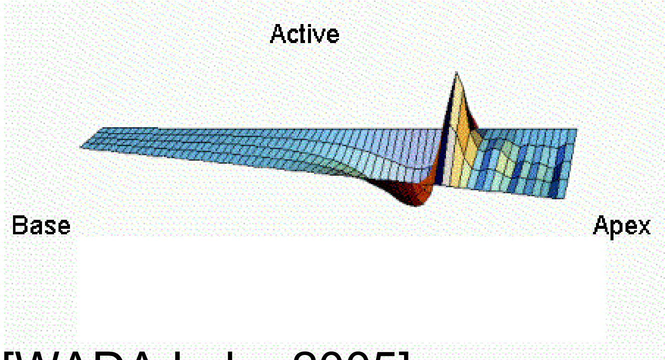
$\rightarrow$ at BM



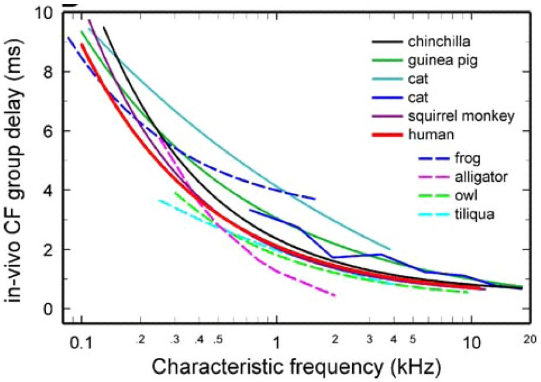
[Kießling,et al. 1997]

Vibration of Basilar Membrane (BM)

- Periodic pressure wave from oval window causes pressure difference between SV and ST
  - Traveling wave (van Békésy, Nobel Prize in 1961)
    - For high frequencies maximum displacement towards base, for low frequencies towards apex
    - Determines characteristic frequency (CF) for each place
    - Involves frequency-dependent delay (cochlear delay)
- Vibration shape essential for spectral coding



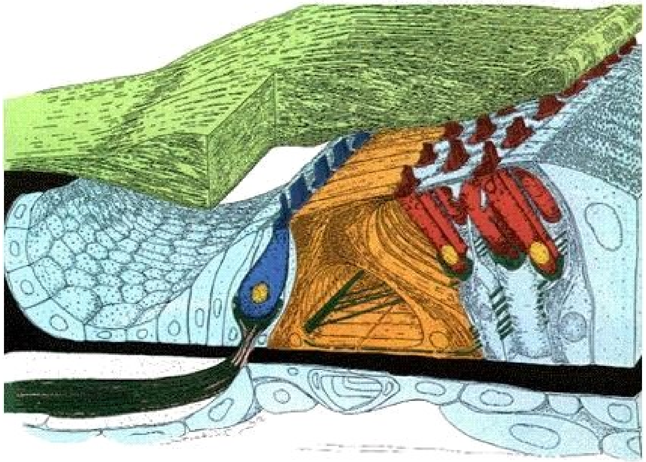
[WADA Lab., 2005]

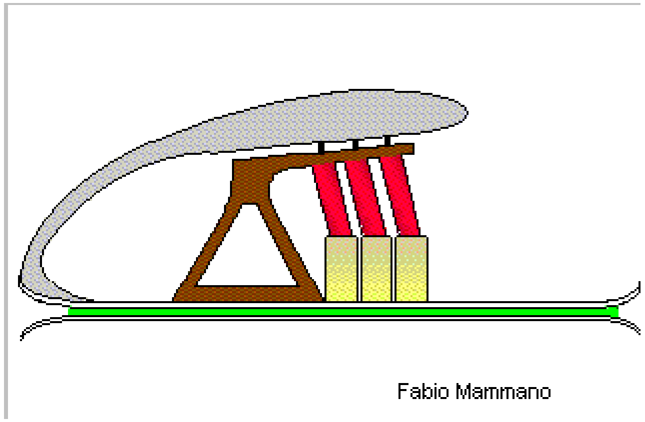


[Ruggero & Temchin, 2007]

Micromechanics of Cochlea (Cortical Organ)

- BM is covered with Tectorial memrane (TM)
- 1 row inner hair cells, IHC (total: 3500)
- 3-4 rows outer hair cells, OHC (total: 25000)
- Thin hairs on top of hair cells (sterocilia) connected to TM
- Shear forces cause deflection
  - Depolarization of hair cells
    - ➔ Neural action potential (spike)

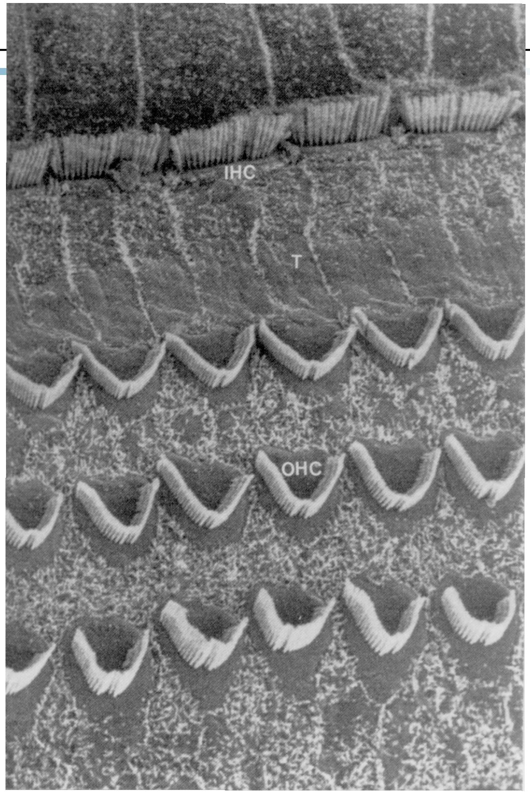




[Mammano 2004]

“Activity” of Outer Hair Cells

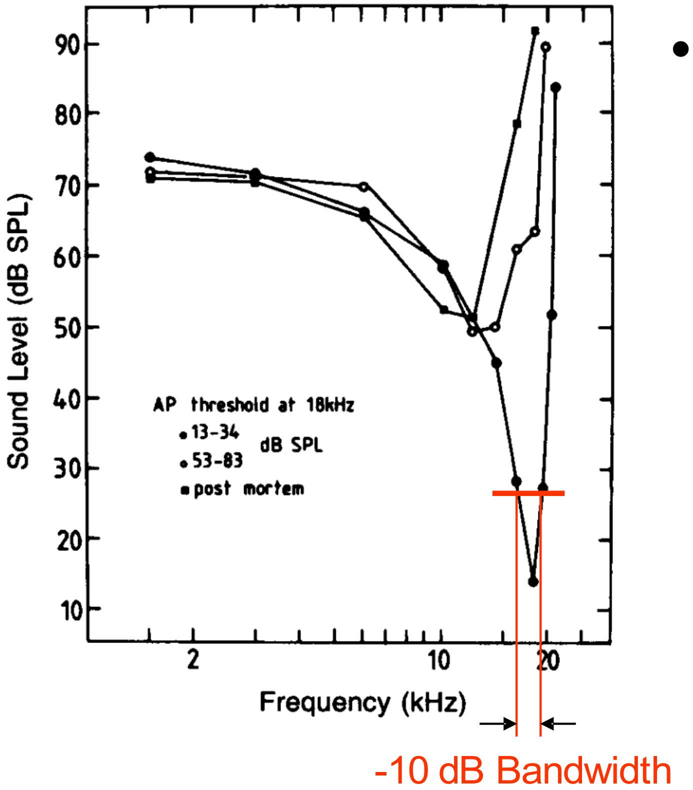
[WADA Lab., 2005]



[Dallos et al, 1986]

- Displacement of stereocilia causes contraction (length variation) of OHCs
  - In phase with signal
  - Active amplification of cochlear vibration
  - Only at low levels (saturation at 50 dB SPL)

Tuning Curve of Basilar Membrane (BM)



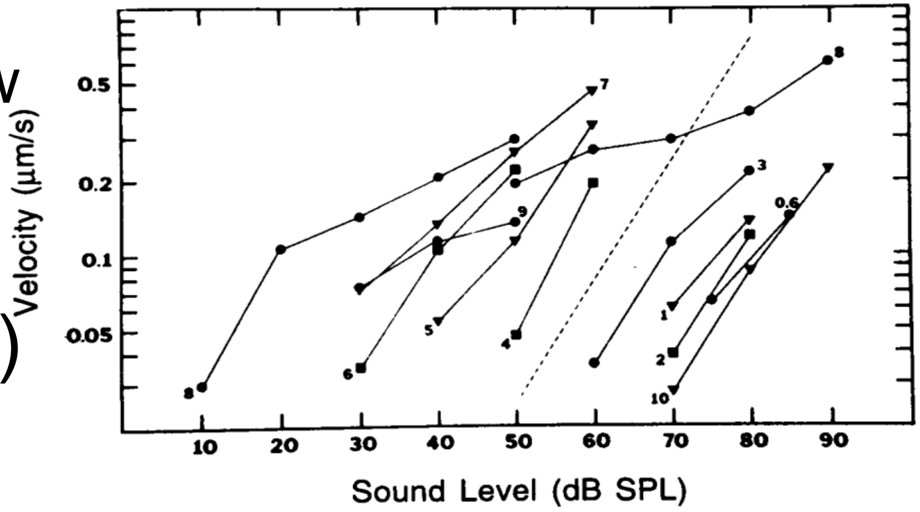
CF: 18 kHz [Sellick et al, 1982]

- Tuning Curve:
  - Tone level required to cause constant BM displacement at fixed CF as a function of tone frequency
    - Filled circles: living animal
    - Empty circle: animal in bad constitution (elevated absolute threshold)
      - OHC damage
    - Filled squares: post mortem

Input-Output Function of Cochlea

- On-frequency (at CF)
  - Compressive response at medium levels
  - Linear response at low and high levels
- Off-frequency (below and above CF)
  - Linear response over the entire dynamic range

[Robles et al., 1986]

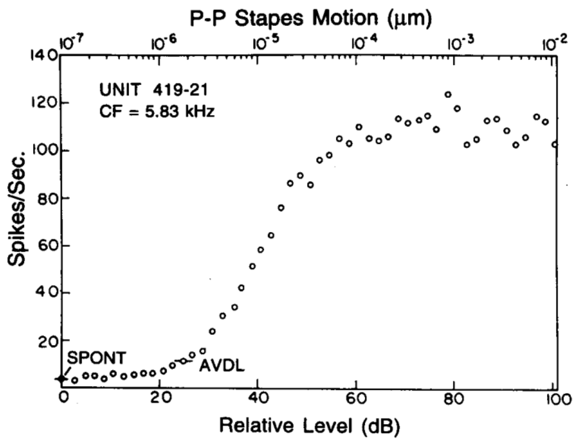


BM response at fixed place (CF = 8kHz) for different tone frequencies

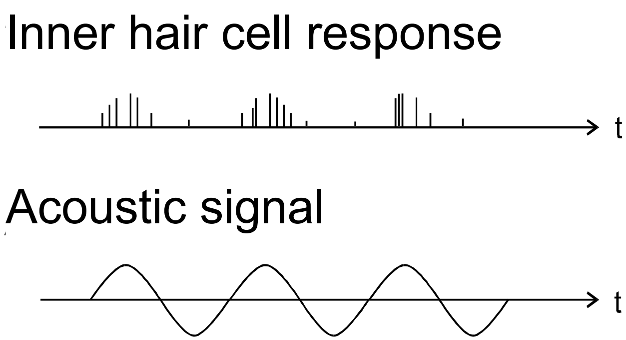
Neural Response

- Spontaneous firing: without stimulus
- Three classes of neurons
  - High spont. rate (18-250 spikes/s)
  - Medium spont. rate (0.5-18 spikes/s)
  - Low spont. rate (< 0.5 spikes/s)
- Neurons with high spont. rates have low thresholds
- Thresholds differ by up to 80 dB between classes

Dynamic Range and Timing



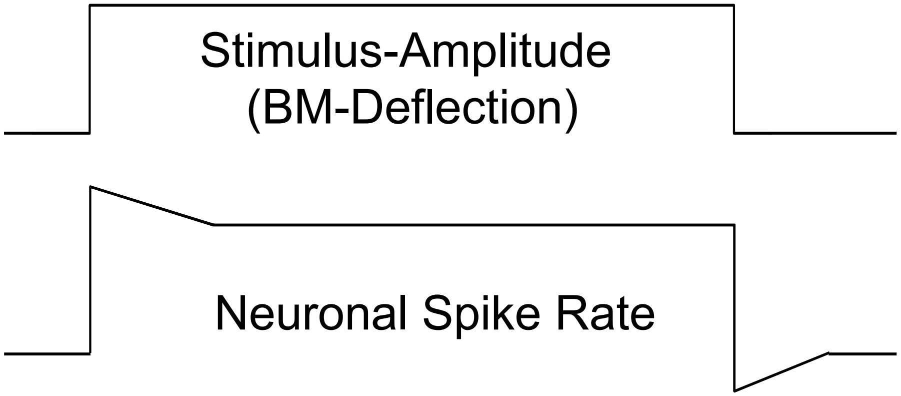
[Kiang, 1968]



- Dynamic range of neurons: 20 – 60 dB
- Phase Locking:
  - Synchronization of spike pattern with signal waveform
- Half-wave rectification
- For signals > 1 kHz decreasing phase locking
  - Due to neural refractory time (approx. 1 ms)

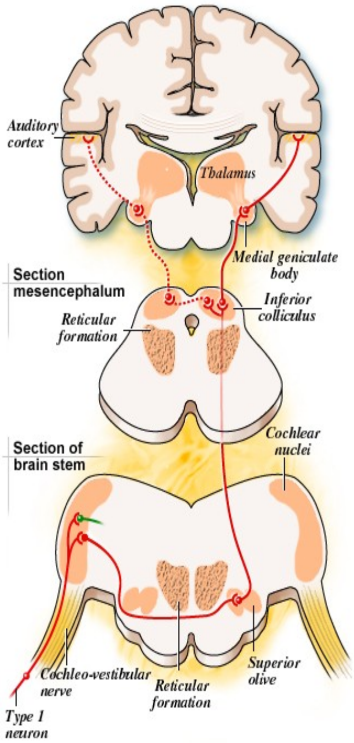
Neural Adaptation

- Typical neural response pattern:
  - **Onset**: High spike rate (and strong phase locking)
  - **Ongoing portion**: saturation at lower spike rate
  - **Offset**: strong reduction and subsequent increase up to spontaneous spike rate
- Contrast enhancement
  - Accentuation of dynamic changes of the input



Central Processing Stages

- Evaluate all incoming information:
  - Neural spike patterns as a function of CF
  - Integration of information across CFs
  - (Integration of information across ears)
- Incorporate *a priori* information
- Involve learning and context effects
- Central detector is assumed to optimally extract the information relevant in a particular task/situation



[Port, 2007]

Basic Auditory Functions

- Absolute threshold:
  - Audibility of frequency components
- Frequency selectivity:
  - Separation of frequency components
  - Reflected in spectral (frequency) masking effects
- Temporal resolution:
  - Reflected in temporal masking effects
- Intensity discrimination, loudness perception, and modulation perception
- Pitch perception

Modeling of Auditory Perception: Why?

- Predict the results from a variety of experiments within one framework
- Explain the functioning of the auditory system
- Help generating hypotheses that can be explicitly stated and quantitatively tested

Types of Auditory Models

- Biophysical:
  - Detailed processes of components and structures
- Physiological:
  - Functionality of components and structures
- Mathematical/Statistical:
  - Abstract representation of auditory processing
- Perceptual/Effective:
  - "Effective" signal processing of stages of auditory system based on perceptual experiments

The Computational Auditory Signal-Processing and Perception (CASP) Model

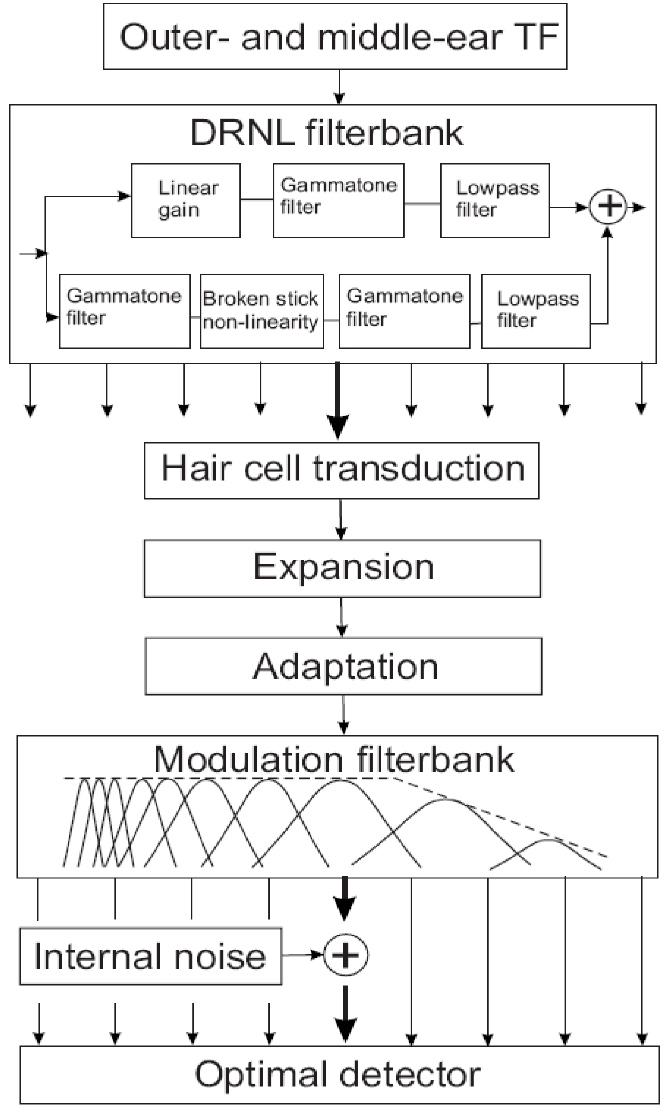
- Simulates the experimentally observed input-output behavior in humans
- No explicit modeling of precise biophysical mechanisms
- Currently focusing on modeling of:
  - Masking phenomena (spectral and temporal)
  - Intensity discrimination
  - Modulation perception
- Based on Dau et al. (1996, 1997) and Jepsen et al. (2008)

Basic Experiment for CASP: Masking Task

- Detection of “target” signal (T) in presence of masker signal (M)
- Three-alternative force choice (3-AFC)
- Random position of T
- Listener task: Which of the three intervals contained T?
- Example trial:

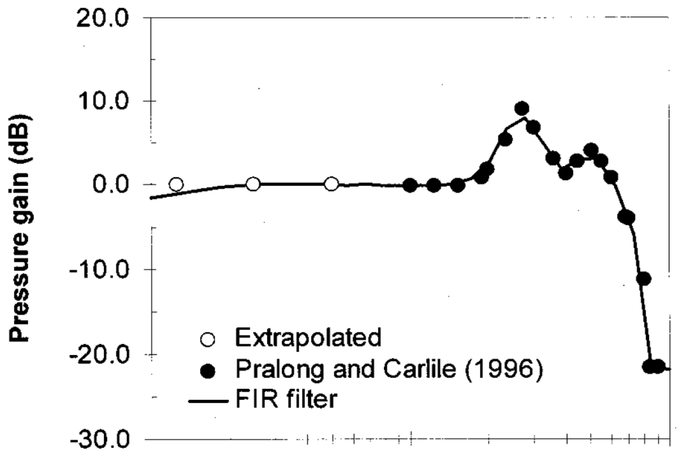
M	M+T	M

Model Structure

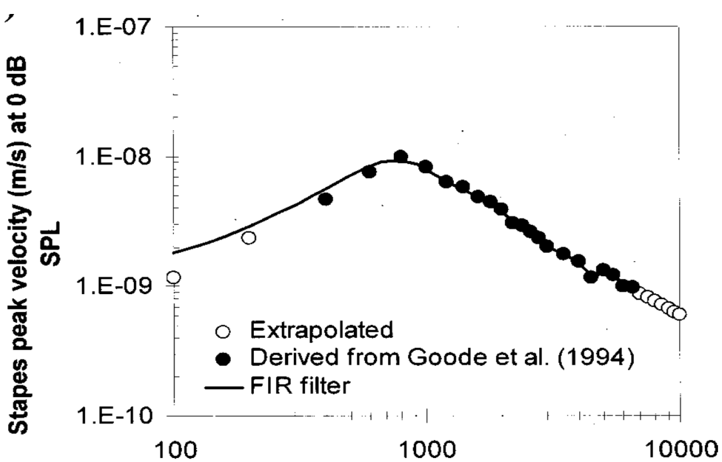


[Jepsen et al., 2008]

Outer- and Middle-Ear Transfer Function



- Outer ear: Typical human headphone-to-eardrum sound pressure gain



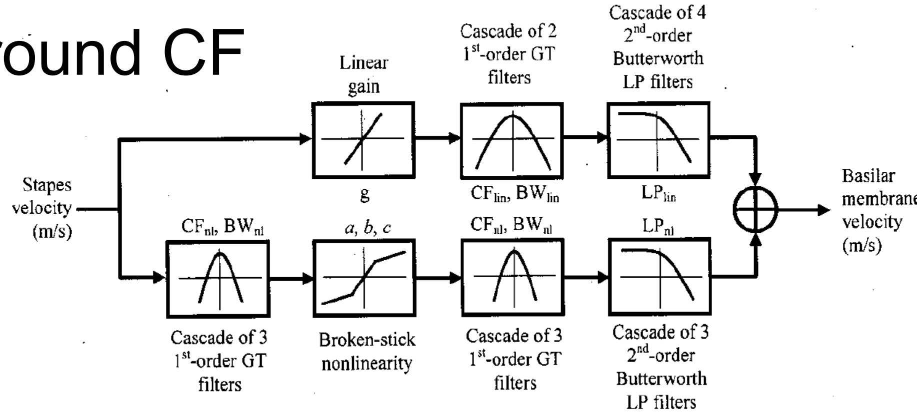
- Middle ear: Stapes peak velocity (m/s) as a function of frequency for SPL = 0 dB

[Lopez-Poveda and Meddis, 2001]

Dual-Resonance Nonlinear (DRNL) Filterbank

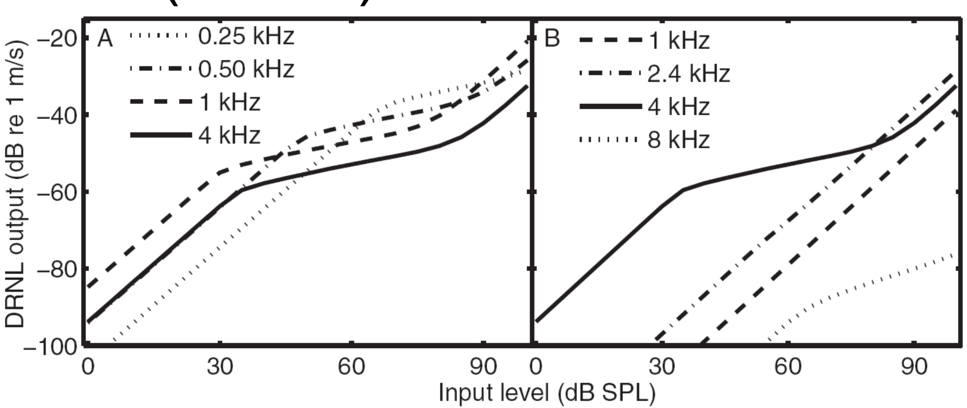
- Mimics the complex nonlinear BM response
  - Compressive for frequencies ±10% around CF
  - Linear outside CF

[Lopez-Poveda and Meddis, 2001]



On frequency (f = CF)

CF = 4 kHz



[Jepsen et al., 2008]

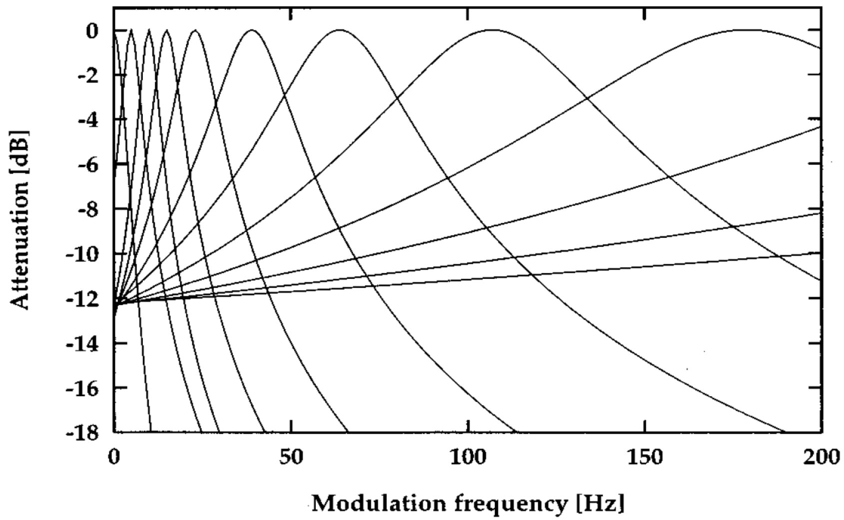
Hair Cell and Auditory Nerve

- Mechanical BM oscillation → Receptor potential
  - Half-wave rectification + 1st order lowpass filter (1000 Hz)
  - Preserves fine structure at low CFs and envelope at high CFs
- Squaring Expansion:
  - Mimics spike rate-vs.-level functions of auditory nerve fibres
- Adaptation
  - Dynamic changes in gain in response to changes in input level

  - Arises at level of auditory nerve
  - Implemented as chain of five nonlinear feedback loops (Dau et al., 1996)

Modulation Processing and Filterbank

- Model of frequency-dependent modulation sensitivity
  - 1st order low-pass filter
- Model of modulation-frequency selectivity
  - Modulation filterbank



[Dau et al., 1997]

Model of Limited Amplitude Resolution

- Gaussian-distributed internal noise
- Added to each channel at output of modulation filterbank
- Noise variance adjusted to predict human intensity discrimination for 1-kHz tone at SPL of 60 dB

Decision Device: Optimal Detector

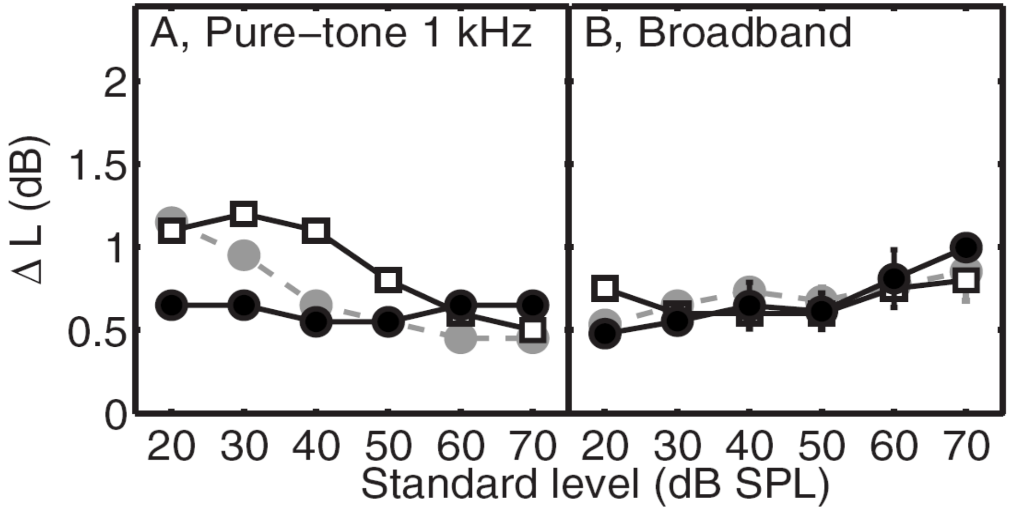
- Assumption: Listener creates "template" of target
  - Template = Int(M + T) – Int(M); T > masked threshold
    - Int(x) = internal representation of x
    - 3-dimensional pattern with axes time, frequency, and modulation frequency
  - Simulation to obtain "difference representation":
    - Int(M) – Int(each interval of AFC task) → T+IN or IN?
    - IN = internal noise
  - Decision based on cross-correlation coefficient between "template" and "difference representation"
  - Interval with largest value assumed to contain target
  - Simulation of experimental procedure

Model Evaluation

- Simulation of psychoacoustic experiments:
  – Using experimental stimuli
  – Applying experimental paradigm
- Presentation of results:
  – Dark filled circles: CASP model
  – Gray filled circles: original model (Dau et al., 1997)
    - Main difference: linear filterbank instead of DRNL
  – Open symbols: experimental data

Intensity Discrimination

- Stimuli:
  - 1-kHz sinusoid, broadband noise
  - Several reference levels
- Poor prediction for pure-tones:
  - Possibly due to lacking evaluation of nonlinear phase effect across CFs

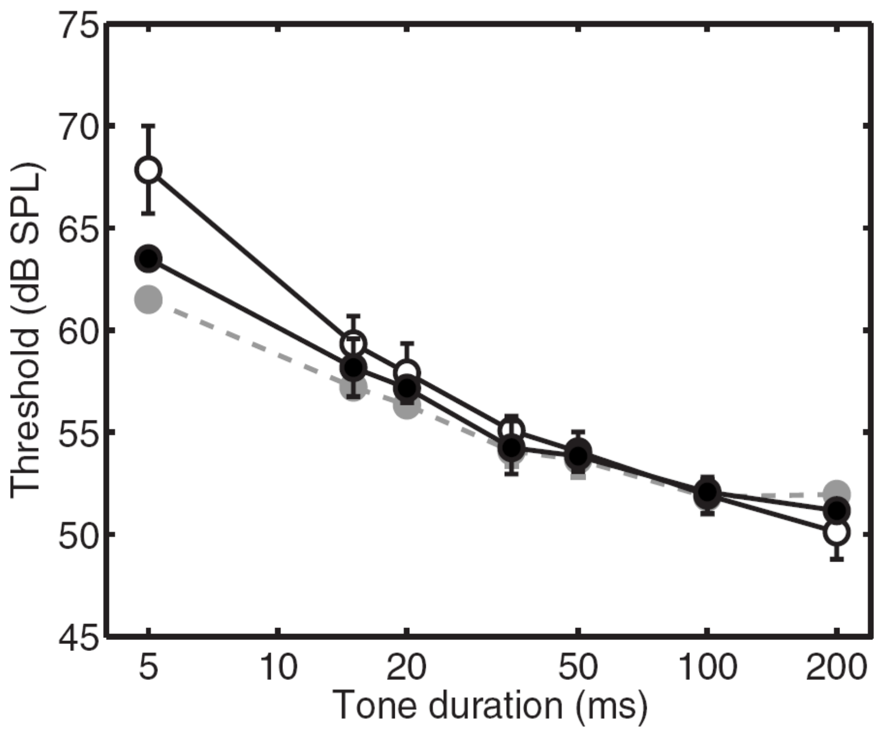


Data from [Houtsma et al., 1980]

Figure from [Jepsen et al., 2008]

Tone-in-Noise Simultaneous Masking

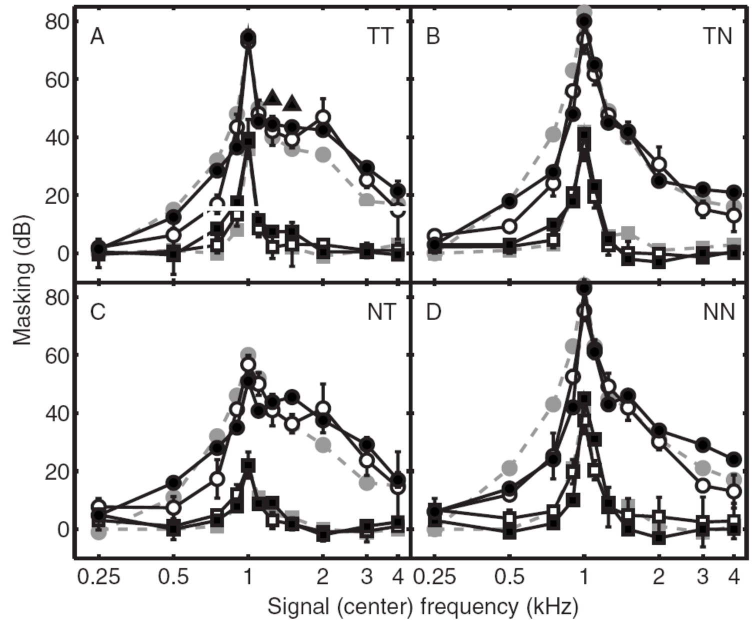
- Target: 2-kHz sinusoid, variable duration
- Masker: band-limited Gaussian noise, 500 ms



[Jepsen et al., 2008]

Spectral Masking Patterns

- Target: sinusoid or NB-noise, 200 ms
- Masker:
  - sinusoid or NB-noise at 1 kHz, 200 ms
  - Two masker levels: 45 and 85 dB SPL

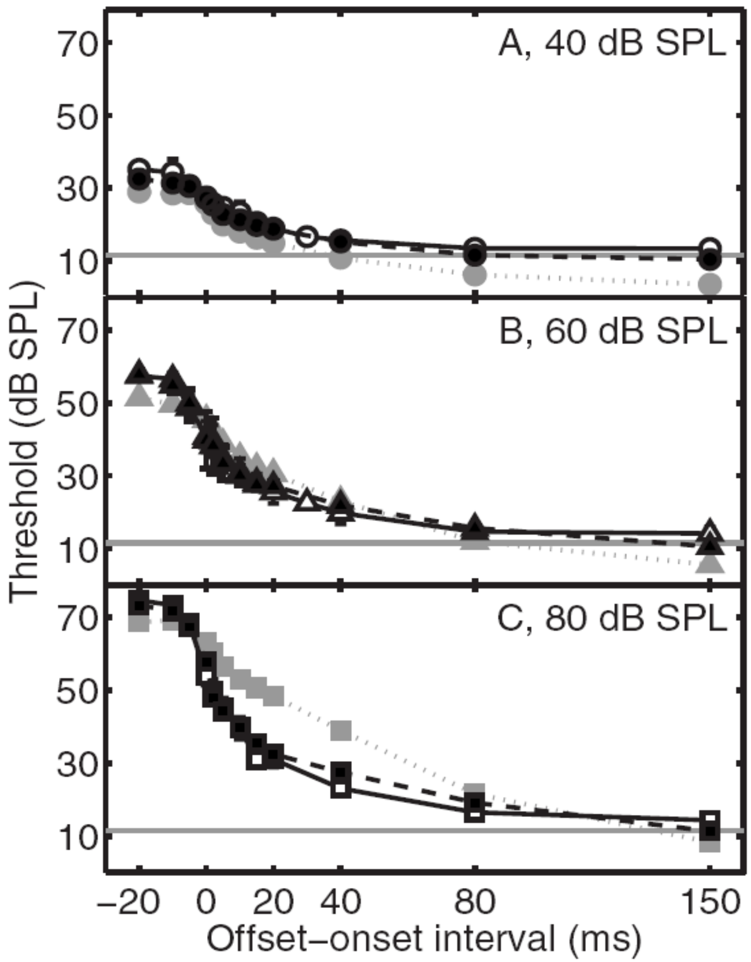


Data from [Moore et al., 1998]

Figure from [Jepsen et al., 2008]

Forward Masking I

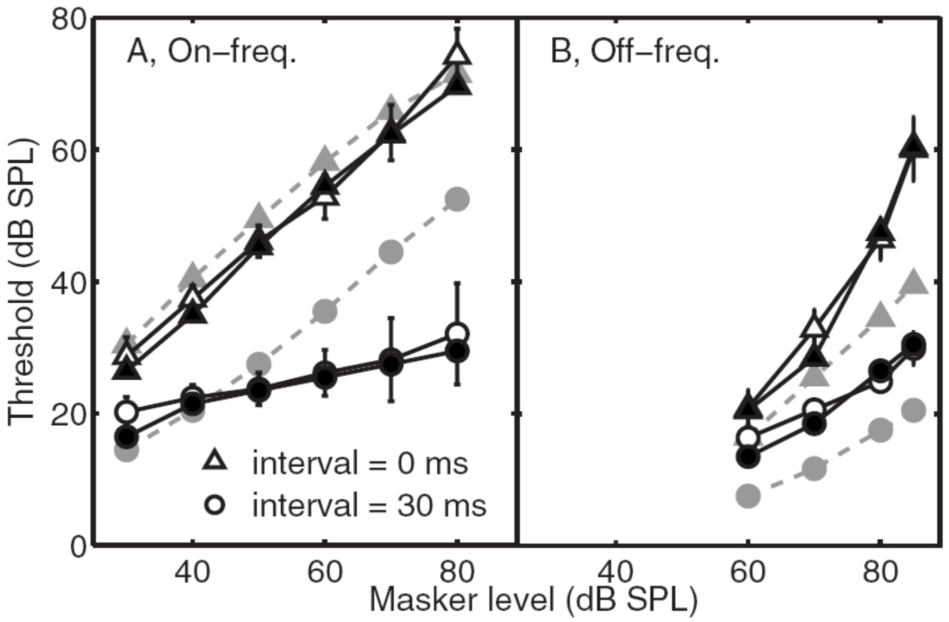
- Target: 4-kHz sinusoid, 10 ms
- Masker: broadband Gaussian noise (0.02-8 kHz), 200 ms
- Variable masker-target intervals



[Jepsen et al., 2008]

Forward Masking II

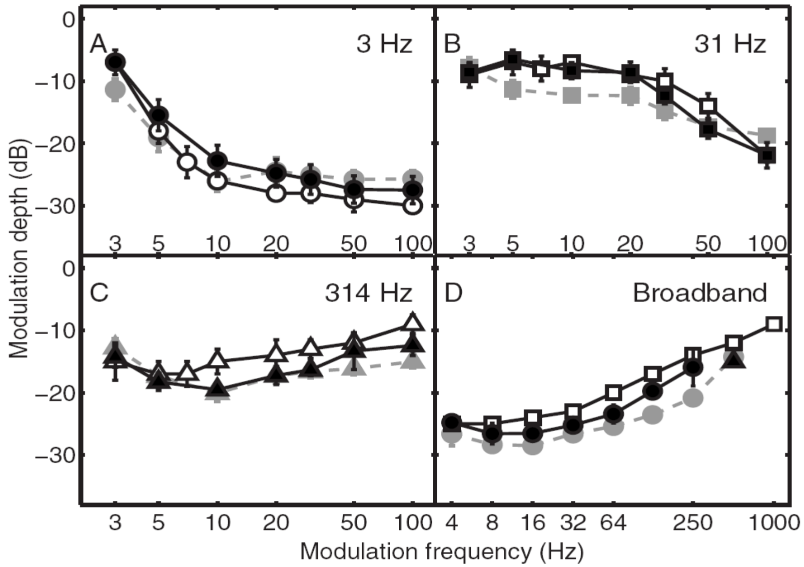
- Target: 4-kHz sinusoid, 10 ms
- Masker: variable levels, 0 or 30-ms separation
  - On-frequency: 4-kHz sinusoid
  - Off-frequency: 2.4-kHz sinusoid



[Jepsen et al., 2008]

Modulation Detection

- Stimuli:
  - Narrow-band carrier: band limited Gaussian noise, centered at 5 kHz, bandwidth = 3, 31, 314 Hz
  - Broadband carrier: Gaussian noise
  - Modulator: pure tone, different frequencies

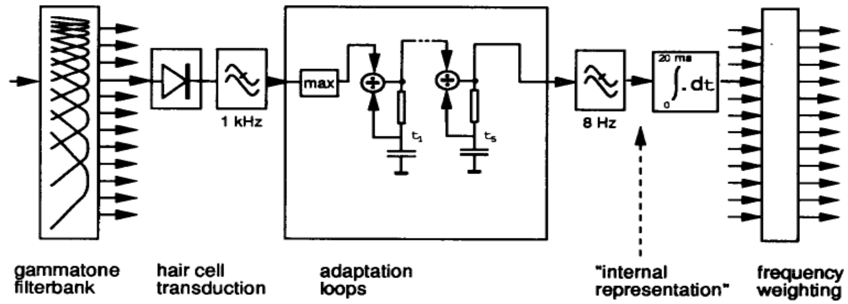


Data from [Dau et al., 1997] and [Viemeister 1979]

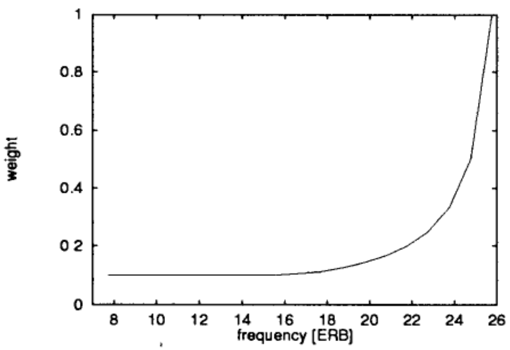
Figure from [Jepsen et al., 2008]

Perception of Sound Quality: Design

- Subjective data:
  - *Low-bit-rate speech*
  - *Mean Opinion Score* (MOS)
- Objective speech quality:
  - Early version of CASP model
  - Based on Dau et al. (1996)
  - Compares internal representation of coded and reference signal:
    - Distance measure based on correlation coefficient: *qc*
    - Employs band weighting function



Auditory processing model (based on Dau et al., 1996)

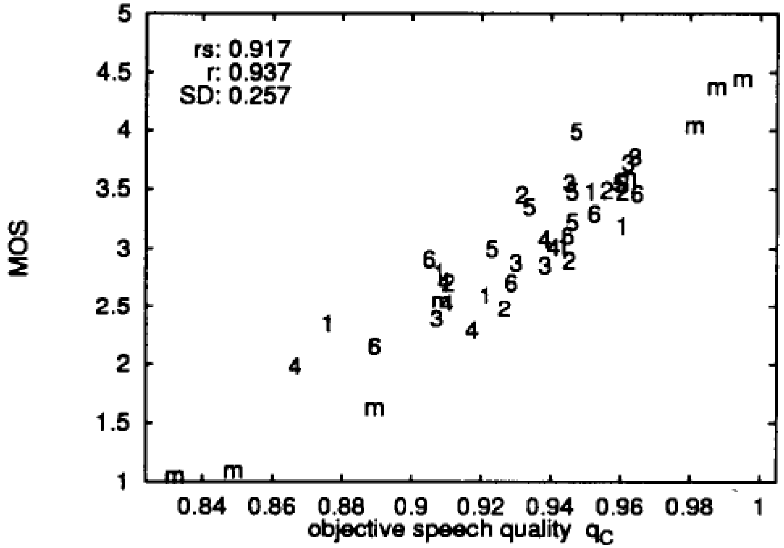


Band weighting characteristic

All data from [Hansen and Kollmeier, 2000]

Perception of Sound Quality: Results

- Good prediction of subjective data (r = 0.94)
- Best prediction with:
  - Strong weighting of high-frequency bands
  - CASP parameters optimized for psychoacoustic data



Data from [Hansen and Kollmeier, 2000]

Power of the Model

- Combines information across:
  - Frequency channels
  - Modulation channels
  - Time
- Allows to model complex spectro-temporal interactions in hearing like:
  - Co-modulation masking release
  - Two-tone suppression
  - Temporal or spectral integration effects (including multiple-looks processing)
- Able to deals with arbitrarily complex stimuli

CASP: Applications and Possible Extensions

- Modeling of hearing loss:
  - Allows studying consequences of dysfunction of specific mechanisms/processing stages like:
    - Damage of outer hair cells
    - Lack of phase locking
- Can be used as front-end for models of other auditory functions like:
    - Pitch perception
    - Sound localization and binaural signal detection
    - Auditory scene analysis (ASA)
    - Speech recognition
    - Subjective evaluation (sound quality assessment)

Summary & Conclusions

- The processing of the auditory system is highly nonlinear and complex
- CASP:
  - Perceptual model for the peripheral auditory system
  - Modular and easily extendable
  - Able to simulate many auditory phenomena
- CASP implementation:
  - in MATLAB available upon request (Morten Jepsen)
    - http://caspmodel.sourceforge.net/
  - An open-source MATLAB implementation of the model soon available: AMToolbox @ sourceforge.net