Aalto University
School of Electrical
Engineering

Communication acoustics
Ch 13: Auditory modeling

Pedro Llado, Ville Pulkki

Department of Signal Processing and Acoustics
Aalto University, Finland

September 22, 2022

Auditory models

- Auditory system can not be explained with mathematical expressions accurately
- E.g., the formula presented for loudness don’t explain many phenomena in hearing
- DSP models of hearing have proven to be effective in modeling
- A large number of different types of auditory models have been developed
- We go through some basic versions of most-well-known models

This chapter

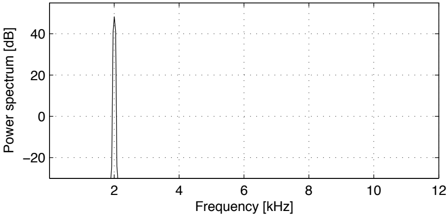
- Simple psychoacoustic models;
- Filter bank models;
- Cochlear models;
- Hair-cell models;
- Models for cognitive processing;
- Models of binaural interaction.

Simple psychoacoustic model with DFT

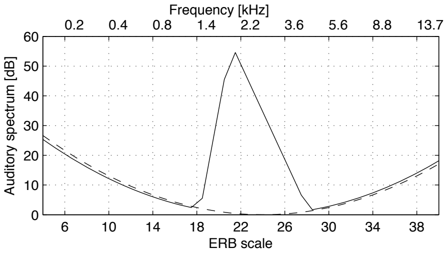
- Windowing with 25 ms Hamming window
- DFT (discrete Fourier transform)
- Power spectrum
- Magnitude response of ear canal and middle ear
- Spreading of excitation in frequency

Sine input

- Power spectrum

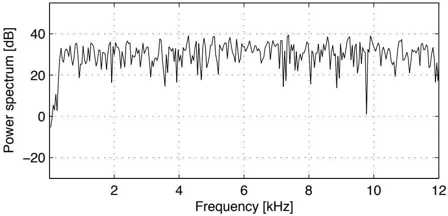


Auditory spectrum

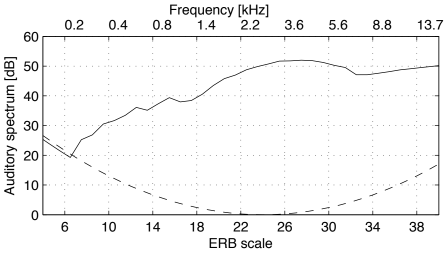


White noise input

- Power spectrum



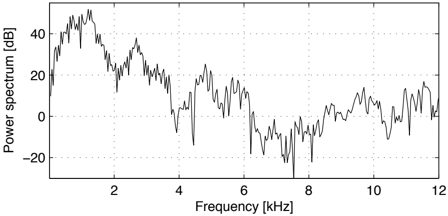
Auditory spectrum

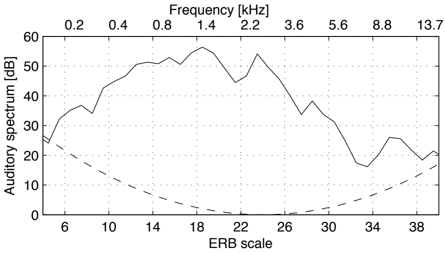


/a/ input

- Power spectrum

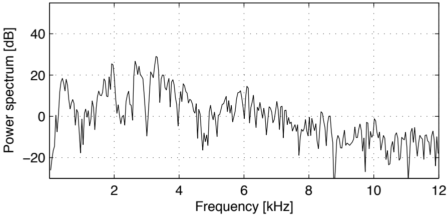
Auditory spectrum



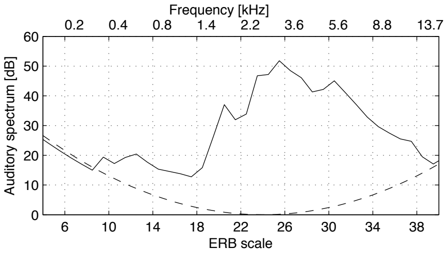


/s/ input

- Power spectrum

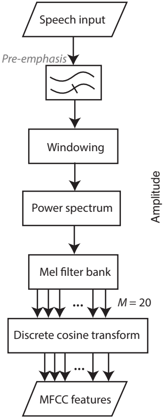


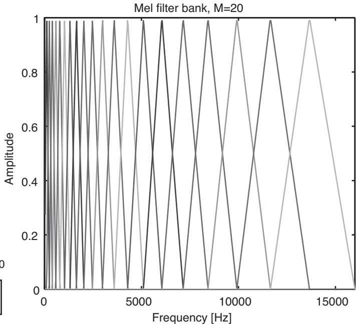
Auditory spectrum



Mel cepstral coefficients

- Utilized widely in speech recognition
- Similar processing with DFT-based auditory models



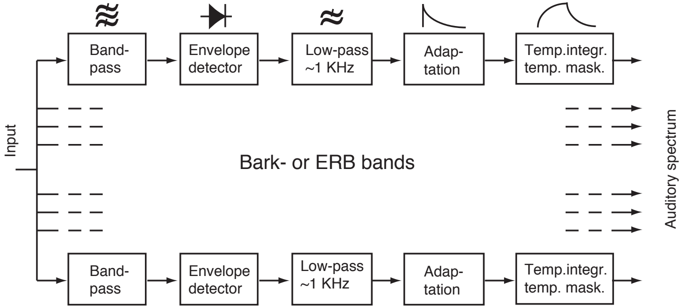


Shortcomings with DFT-based auditory models

- Time window forces equal time resolution at all frequencies, not realistic
- Forward masking and temporal integration not easy to model
- Level-dependent effects

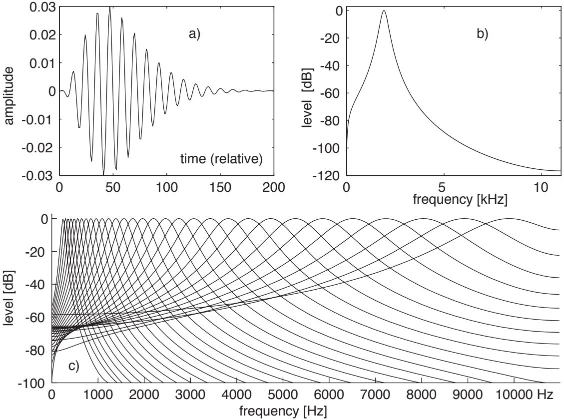
Filter bank models

- Model cochlea by time-domain parallel filters
- Neural phase locking (half-wave rect, low-pass filter)
- Adaptation, temporal integration (e.g., loudness modeling)



Gamma-tone filter bank

- Often used in auditory modeling



Simple auditory model

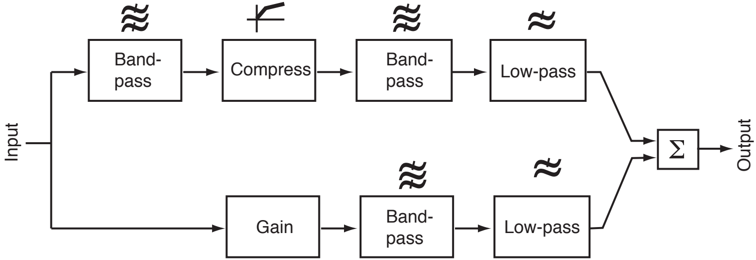
```
%create of a gammatone filter bank using a command from the auditory
%modelling toolbox (http://amtoolbox.sourceforge.net)
[b,a] = gammatone(erbspacebw(fLow,fHigh),fs,'complex');

%processing the signal through the filter bank
filterOut = real(ufilterbankz(b,a,sample));

%emulation of the neural transduction with half-wave rectification and
%low-pass filtering of the filter bank output
rectified = filterOut.*(filterOut>0);

%a first-order IIR filter is used as the low-pass filter
beta = exp(-fCut/fs);
outSig = filter(1-beta,[1 -beta],rectified);
```

Shortcomings in simple filter-bank models

- Better approximation of time-frequency resolution than DFT models
- Time resolution may still be to coarse at high frequencies
- Response to short transients is too slow
- Level-dependent effects are missing

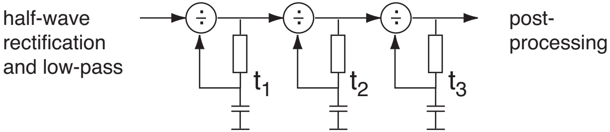
Dual-resonance filter bank

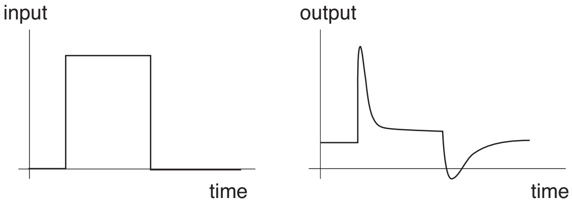
- Implements at least some level-dependent effects in cochlea
- Lower path broader, upper path narrower response



Modeling of neural adaptation

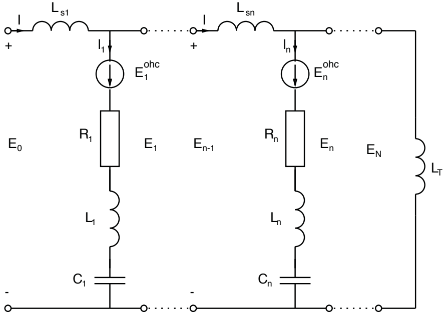
- Serial feedback loops with division





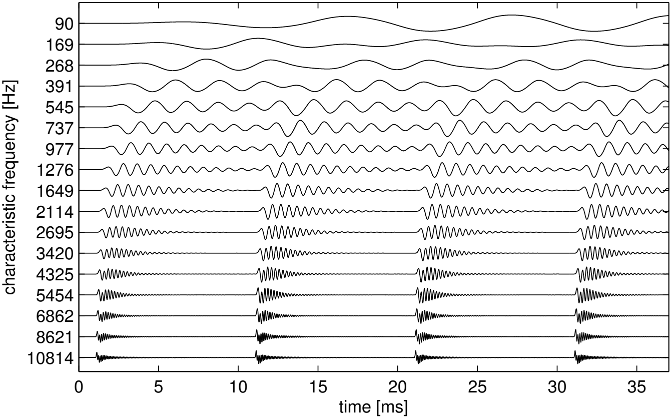
Modeling basilar membrane

- 1D-2D-3D FEM, Transmission line models,
- Electric equivalent models with possible nonlinear components modeling outer hair cells
- Computationally very demanding, becoming more popular



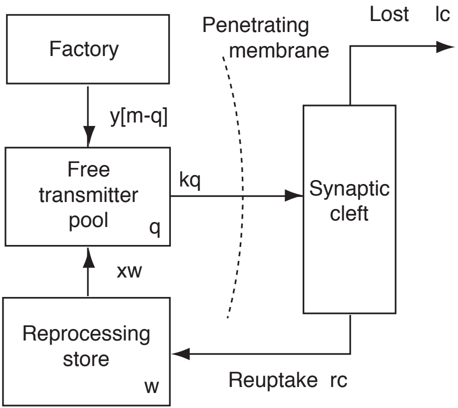
Transmission-line model response to click train

- High-frequency response: peak + noise burst, simulates outer hair cell functionality



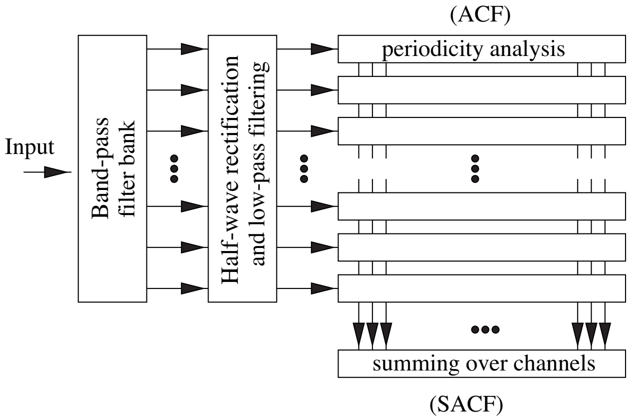
Modeling cells separately

- Processing the neurotransmitters in hair cell
- Analog: two pools of water connected

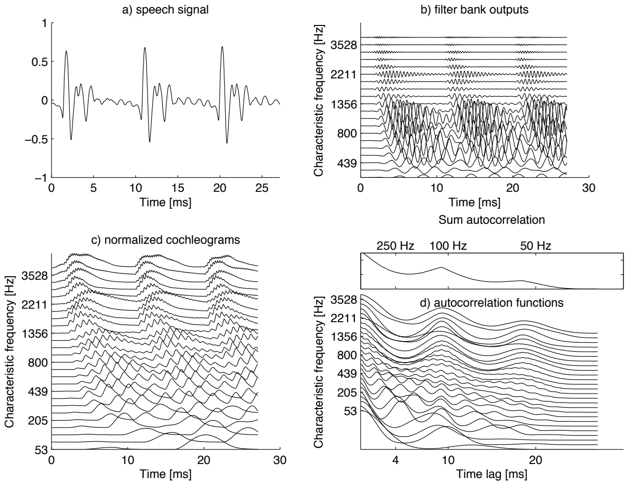


Periodicity analysis, functional model

- Assumes time-domain analysis of the pitch
- Sum autocorrelation function

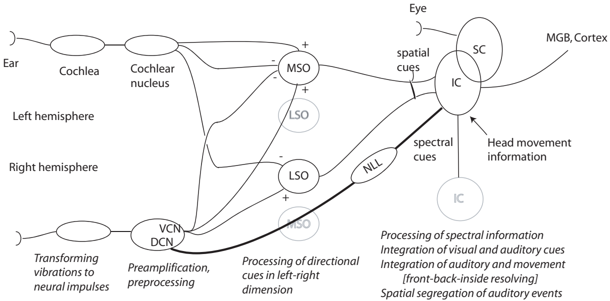


Periodicity analysis, example



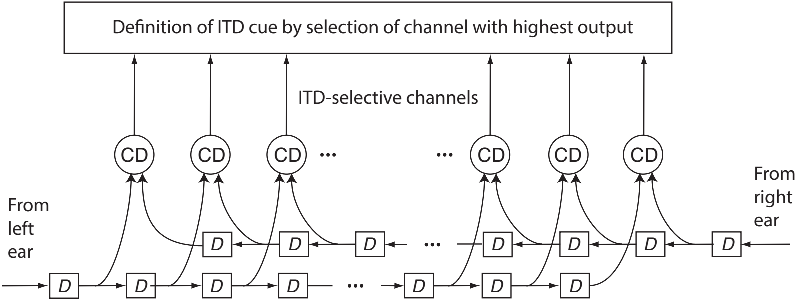
Models of spatial hearing

- Spatial hearing is based on signal analysis in the brains
  - Decoding of monaural spectral cues
  - Decoding of binaural cues
  - Binaural detection of signals
- Distance perception, no models available



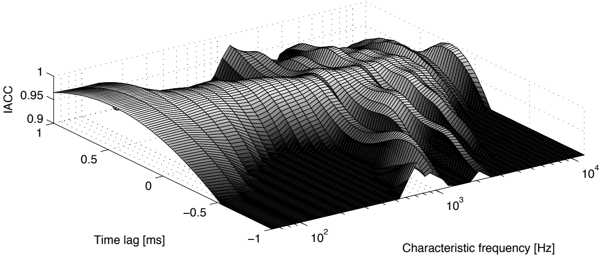
Delay-network-based models

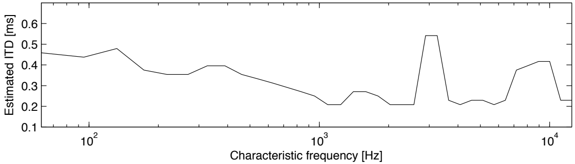
- Signals from the ears meet in neurons, topographic mapping
- Most active output defines the azimuth (of confusion cone)



Delay-network response to source in free field

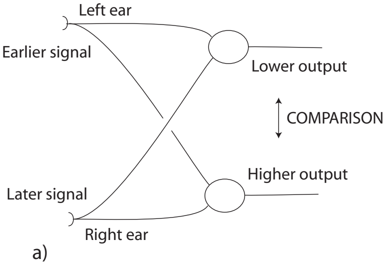
- Speech source in 30°





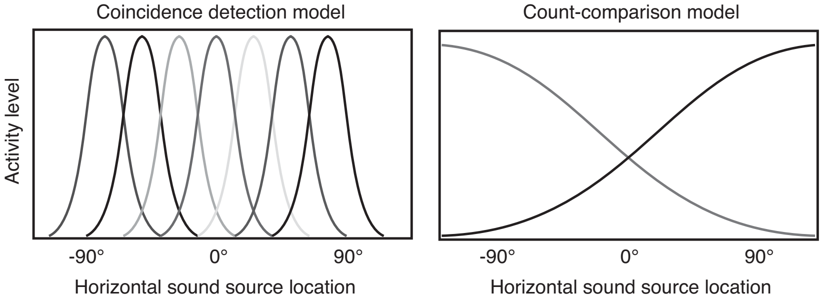
Count-comparison models

- An estimate of left-right coordinate is computed
- Not topographic mapping, only computation of spatial information, "where is the source?"



Difference between principles of binaural models

- The first neurons sensitive to ITD have different sensitivities assumed by the models
- Debate is going on. Some neurophysiological evidence that delay-lines exist in avians, and count-comparison mechanisms in mammals



Applications of auditory models

- Audio coding
  - Psychoacoustic or perceptual models of masking
- Sound quality modeling
  - Modeling of perceived differences
  - Criteria for audio reproduction
  - Binaural audio quality
- Speech recognition
  - Advanced front-end models
- Advanced hearing aids
  - Cochlear implants

References

These slides follow corresponding chapter in: Pulkki, V. and Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, 2015, *where also a more complete list of references can be found.*