ARC 3D Audio Colloquium

Blauert: Technology of binaural listening:

- Extracting information from binaural signals
- Optimization of binaural algorithms
- Binaural dereverberation

Xie: HRTF and VAD

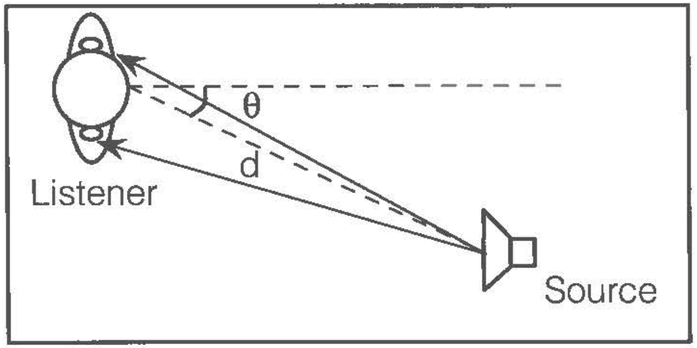
- Binaural headphone reproduction

Jakob Vennerød

SINTEF ICT Acoustics

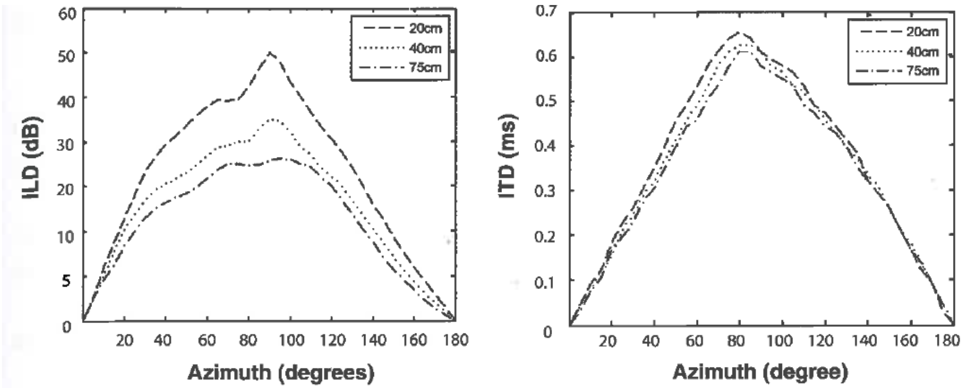
Extracting Sound-Source-Distance Information from Binaural signals

- Humans with normal hearing can estimate the distance to a sound source with reasonable accuracy. Why?
- How can we use this knowledge to estimate sound source distance computationally?
- Distance perception relies on binaural cues (in particular when the source is close to the listener) and monaural cues (near- and far-field sources).



Distance perception factors

- Stimulus spectral content / envelope
- Sound reflections and direct-to-reveberant ratio (DRR)
- A priori knowledge of stimuli presentation level
- Azimuthal location



- Visual information about possible sound sources
- Over-/underestimation
  - $r' = kr^{\alpha}$
- Binaural cues – the ILD can be up to 50 dB at a distance of 20cm.

Distance-estimation methods

- Important: Estimate Direct-to-Reverberant Ratio (DRR) – *but we then need to know the reverberation time!*
- Several other methods have been proposed, but most of them requires training in the room.
- With enough training, one can eliminate the need for a priori knowledge
  - Machine Learning
  - Statistical properties of the binaural signals
  - Effective measure: Binaural Spectral-Magnitude-Difference Standard Deviation (BSMD-STD)
- Coarse distance estimation > 90 % performance (Georganti, May, van de Par, Mourjopoulos, 2013). Applies only for small distances, performance degrades when the room acoustics change

BSMD-STD

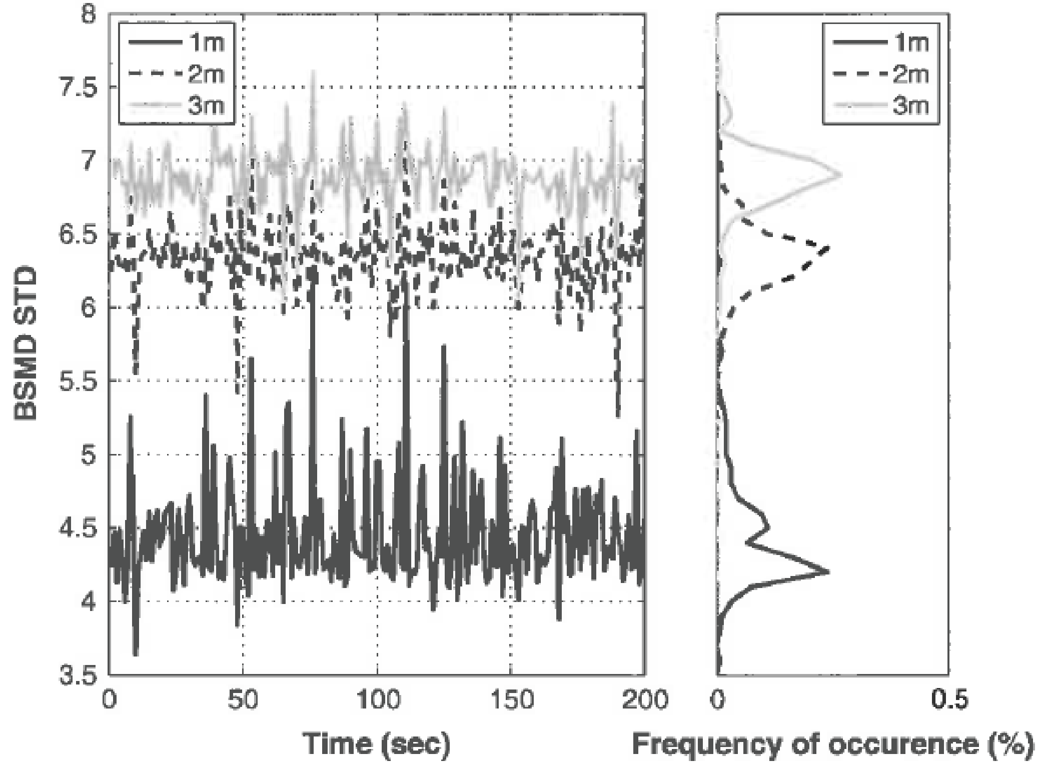


Fig. 10 *Left* BSMD–STD extracted from speech signals recorded at different source/receiver distances in room with a reverberation time of $T_{60} = 0.89\,\mathrm{s}$. *Right* The corresponding histogram of the extracted BSMD–STD values

Optimization of Binaural Algorithms for Maximum Predicted Speech Intelligibility

- Binaural unmasking of speech-in-noise improves the SNR of about 10dB when the speech and noise are separated by 90 degrees azimuth
- Increasing speech intelligiblity is a difficult problem, solved primarily in research by
  - beamforming algorithms,
  - blind source separation (BSS) or
  - multi-channel Wiener filters

Binaural statistics

- The IPD of the envelope represents a meaningful location feature throughout the entire spectrum, in addition to IPD at LF and ILD at HF. However it is sensitive to noise.

- The standard deviation of binaural parameters is generally higher for lateral sources

- Directional hearing aids alter the front-back ambiguity ("cone of confusion"), increases IPD and reduces ILD

- Relying on these parameters *only,* results in speech processing algorithm degradation when noise is present. We have to include additional binaural parameters.

Classical Binaural ASA algorithms

- ASA = *Auditory Scene Analysis*
- Three basic groups:
  - *Carrier-Level-Phase* (CLP) – Calculates source direction and a resulting freq/time dependent gain factor (*filter mask)* to enhance speech
  - *Carrier-Coherence* (CC) – Based on L/R coherence with gain factors proportional to coherence
  - *Envelope-Level-Time* (ELT) – Based on ITD/ILD of waveform envelope of the fundamental frequency of speech.
- These algorithms have been shown to improve speech intelligibility in binaural HAs
- Normally based on short-time binaural parameters (e.g. 8-16 ms)

Algorithm optimization

- Genetic algorithms
  - Fast convergence
  - Practical and psychoacoustically grounded solutions
  - May give insight in the ranking of low-level binaural cues in the hearing system
- With incoherent noise, CC and CLP algorithms give 5-15 % increase in speech intelligibility, while no improvement is gained with ELT.
- With coherent noise, CC is not applicable. If there is only one coherent interference, ELT may work.
- Bivariate (IPD + ILD) model perform better than a univariate model (IPD@LF, ILD@HF)
- CLP is inferior in most situations

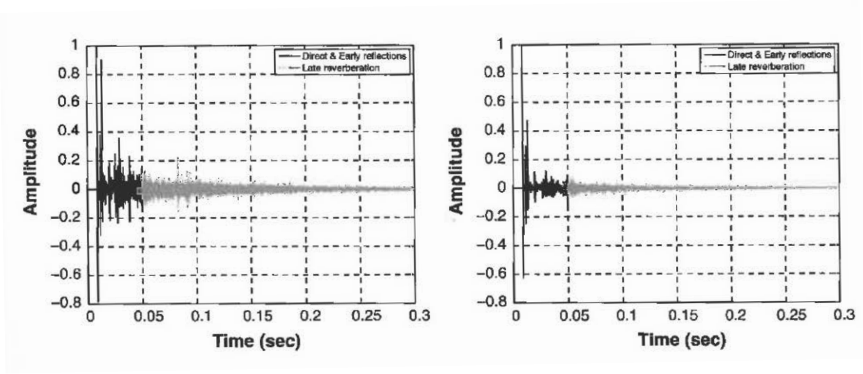
NASA spacecraft antenna

Binaural dereverberation

- BRIRs can be decomposed into two parts which represent the direct sound + early reflections, and late energy (reverberation)

$$h_i(n) = h_{i,e}(n) + h_{i,l}(n)$$

- Convolution with these IRs gives the signal at the ear. Thus we can say that the ear signal has two components $x_{i,e}(n)$ and $x_{i,l}(n)$.
- In most dereverberation applications, these signals are normally treated separately.



Speech signals in rooms

- Early reflections (<50-80 ms) improve speech intelligibility
- Late reverberation (>80 ms) has a negative effect
- The binaural auditory system suppresses early reflection coloration and late reverberation. Auditory masking masks many reflections
- DRR provides a very reliable cue for distance perception
- Direct + early refl. = high IC, Reverberation = low IC

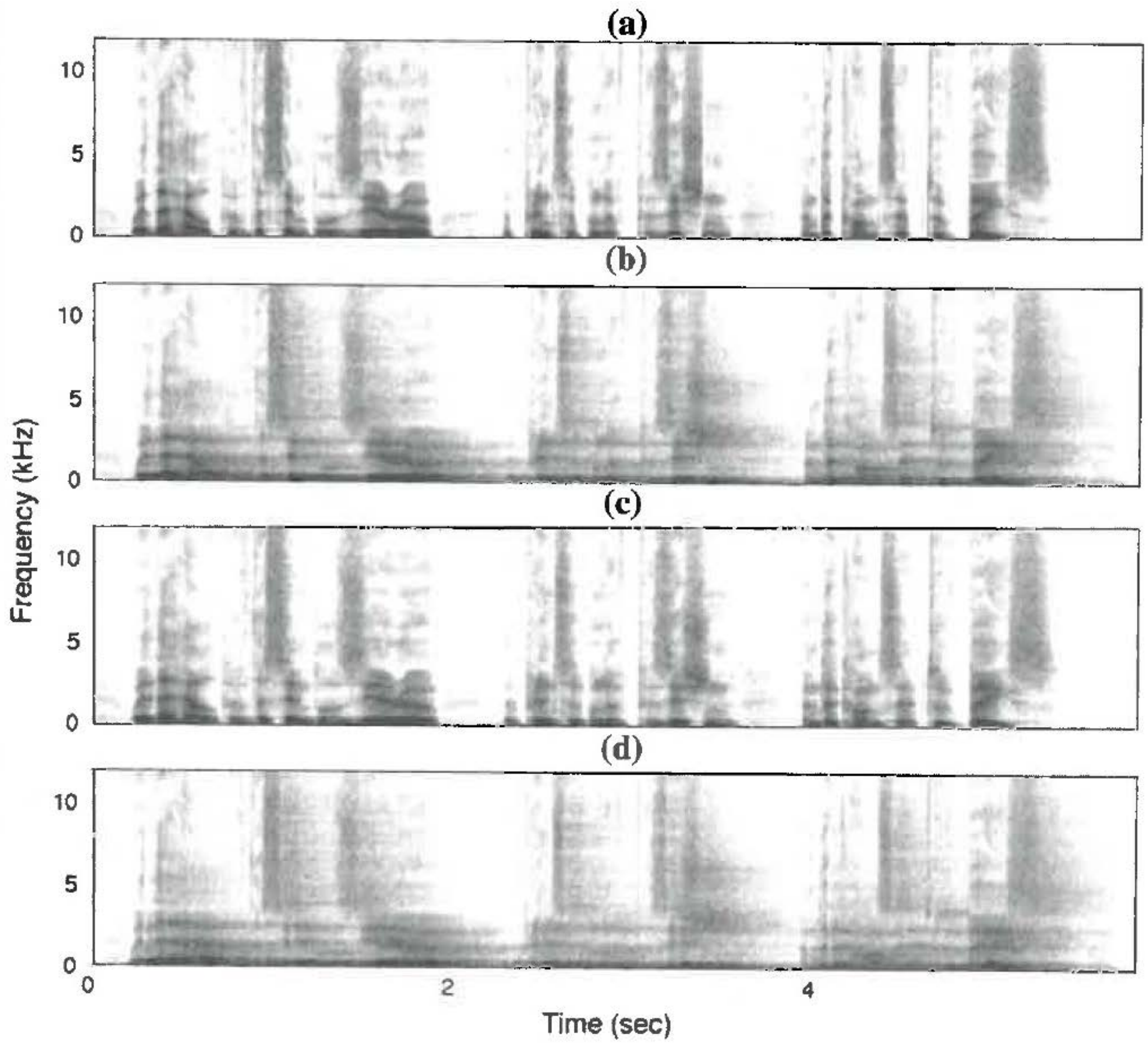
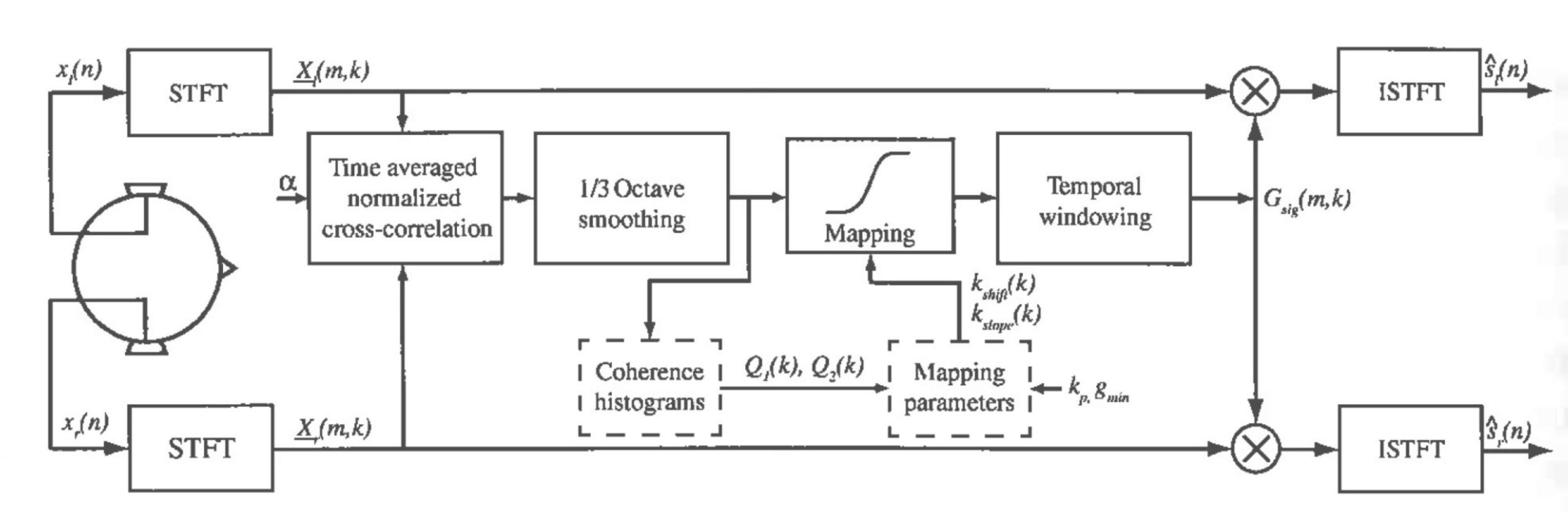


Fig. 3 Spectrograms illustrating the effects of reverberation on speech. **a** Anechoic input signal. **b** Reverberant signal. **c** Reverberant signal due to early reflections only. **d** Reverberant signal due to late reverberation only

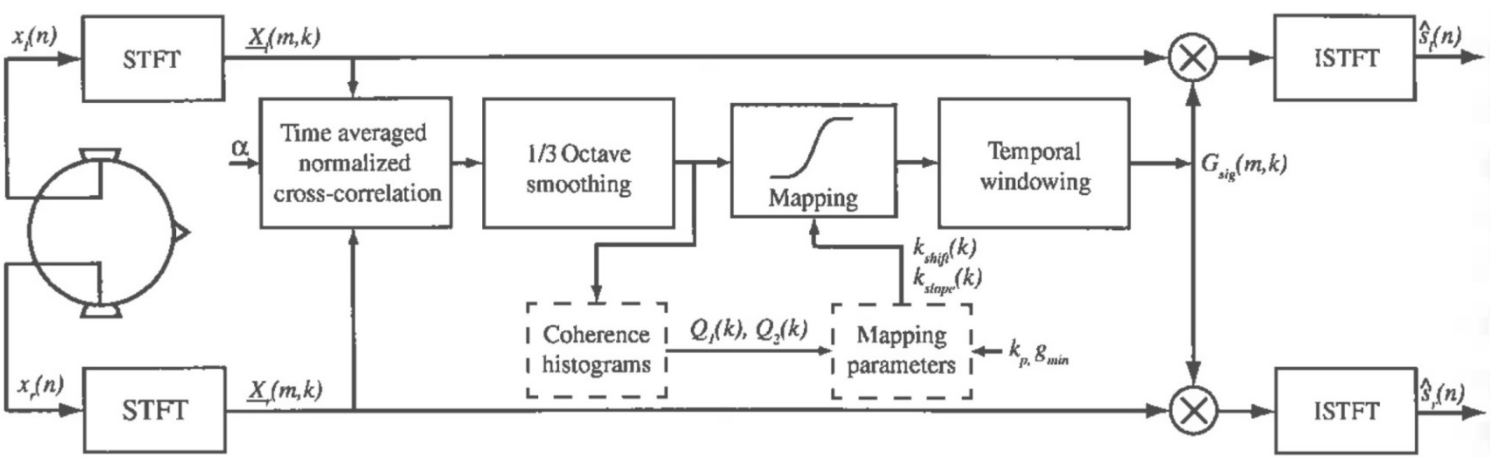
Dereverberation techniques

- Early reflection / coloration
  - Inverse filtering – but only limited usefulness (remove spectral coloration)
  - Cepstral techniques
  - LP-residual enhancement (linear prediction)
- Late reverberation
  - Temporal-envelope filtering (often combined with LP-residual enhancement and spectral subtraction)
  - Spectral enhancement/subtraction (subtract noise from signal)
- Dereverberation based on multiple inputs
  - Based on beamforming
  - Can improve LP-residual enhancement
- Binaural dereverberation
  - Challenge: should preserve ITD and ILD (and spectral cues?)
  - Gain factors based on interaural coherence

Binaural dereverberation



Binaural dereverberation



1. Time-averaged normalized cross-correlation

$$C_{LR}(m,k) = \frac{|\phi_{LR}(m,k)|}{\sqrt{\phi_{LL}(m,k)\phi_{RR}(m,k)}}$$

$$\phi_{LR}(m,k) = \alpha\phi_{LR}(m,k-1) + X_L(m,k)X_R^*(m,k)$$

2. 1/3-octave smoothing of ϕ_{LR}
3. IC estimates are mapped (sigmoidal mapping) to the gain function

$$G_{sig}(m,k) = \frac{(1-G_{min})}{1+e^{-k_{slope}(k)(C_{LR}(m,k)-k_{shift}(k))}} + G_{min}$$

   where k_{slope} and k_{shift} are determined from IC-histogram distributions.
4. Temporal windowing (ISTFT + padding + STFT) to suppress potential aliasing effects

Spectral subtraction framework

- Relies on
  1. Estimating the reverberation decay
  2. Estimating the reverberation power in the current time frame
  3. Subtraction of the estimated reverberation signal
- Can be adopted from mono to binaural dereverberation
  - Identical processing should be applied to L/R channels
  - Can either compute a reference signal from both channels, or compute G_L and G_R separately and compute the L/R gain with mean, max or min of the two channel gains.

Objective and perceptive measures

- Objective methods is an open research issue. Most measures requires knowledge about the source signal.
- In addition, most objective measures do not take the binaural auditory dereverberation or precedence effect into account
- Perceptive measures have only sporadically been used to evaluate dereverberation
- *Multiple-stimuli-with-hidden-reference-and-anchor* test (MUSHRA) is successful for detecting small impairments
- Interaural Coherence method gives good results, especially for close sources

Book:

Head-Related Transfer Function and Virtual Auditory Display

Bosun Xie

HRTF Filter Models and Implementation

- HRTFs must be realised with digital filter models
- Objective: minimise

$$\min \epsilon_{\sum S} = \min \left[\sum_{f_k} \left| H(f_k) - \hat{H}(f_k) \right|^2 \right]$$

  where $\hat{H}(f_k)$ is the frequency response of the filter
- IIR filers are in general less computationally expensive, but more difficult to design stable
- Most of the HRIR energy is located in a 1-1.5 ms window -> short FIR filters are applicable (no localisation degradation with 10 ms filters)
- HRTFs can be decomposed into
  - a minimum-phase function $H_{min}(\theta, \phi, f)$,
  - an all-pass function $\exp[h\psi_{all}(\theta, \phi, f)]$,
  - a linear-phase function $\exp[-j2\pi f T(\theta, \phi)]$

Auditory properties

- An optimal HRTF filter may not be preferred, more concern on errors in auditory perception
- HRTF frequency smoothing
- Auditory weighting, e.g.

$$\min \epsilon_{\sum S} = \min \left[\sum_{f_k} W(f_k) \left| H(f_k) - \widehat{H}(f_k) \right|^2 \right]$$

$$W(f) = \frac{1}{\Delta f_{CB}} \text{ or } W(f) = \frac{1}{ERB}$$

Methods for HRTF filter design

- FIR representation
  - Time windowing $\hat{h}(n) = h(n)w(n)$
  - Frequency sampling method $\hat{h}(n) = IDFT[H(k)]$
  - Interaural transfer function and Wiener filtering (eliminates direction-independent components of the HRTFs)
- IIR representation
  - Prony / Yule-walker algorithms
  - Balance Model Truncation
  - Logarithmic Error Criterion
  - Common-acoustical-pole and Zero Model of HRTFs
- Frequency warping

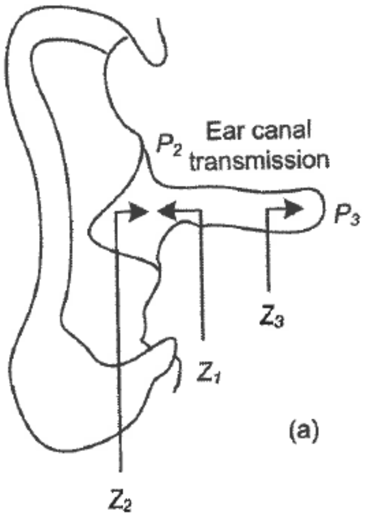
Binaural reproduction through headphones

- Headphone-to-ear canal Transfer Functions (HpTFs)
  - Let Z_1 be the radiation impedance from the ear canal to free air, Z_2 the ear canal input impedance, and Z_4 the headphone impedance seen from the ear canal entrance.

    If $Z_1 \approx Z_4$ and $Z_4 \ll Z_2$ $(f < 1\ kHz)$, we have an open headphone (FEC).

    A commercial "open headphone" is different because it allows sound from the outside to be heard.
  - Compensated by the inverse $F(f) = 1/H_p(f)$. If $H_p(f)$ is minimum-phase, it is invertible and $F(f)$ is causal.

- Compensation of HRTFs with free- or diffuse-field response?



Repeatability of HpTFs

- Circumaural headphones: Standard deviation of 2 – 9 dB at worst (at HF), depending on artifical head and headphones
- Supra-aural headphones: Somewhat larger std.dev. @HF because of pinna deformations
- In-ear headphones (Sennheiser MX500): Std.dev. of less than 1 dB below 10 kHz.
- Pinna deformation is closely related to HpTF repeatability
- Individuality is important; std.dev can be up to 17 dB at 9 kHz (Pralong and Carlile, 1996)

Different headphone types

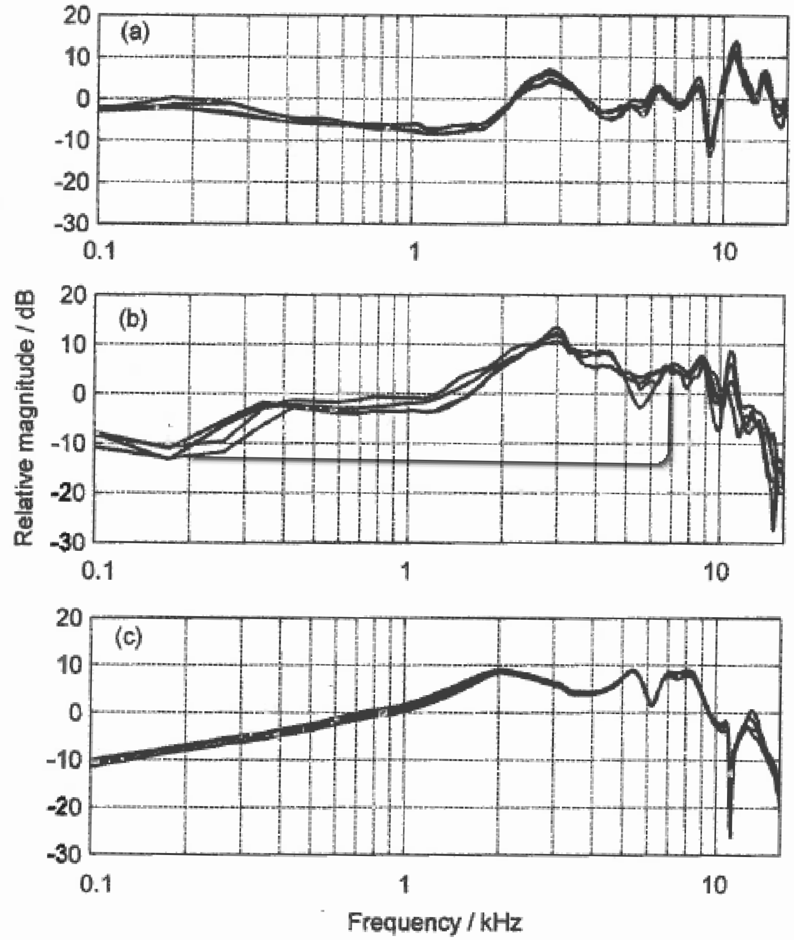
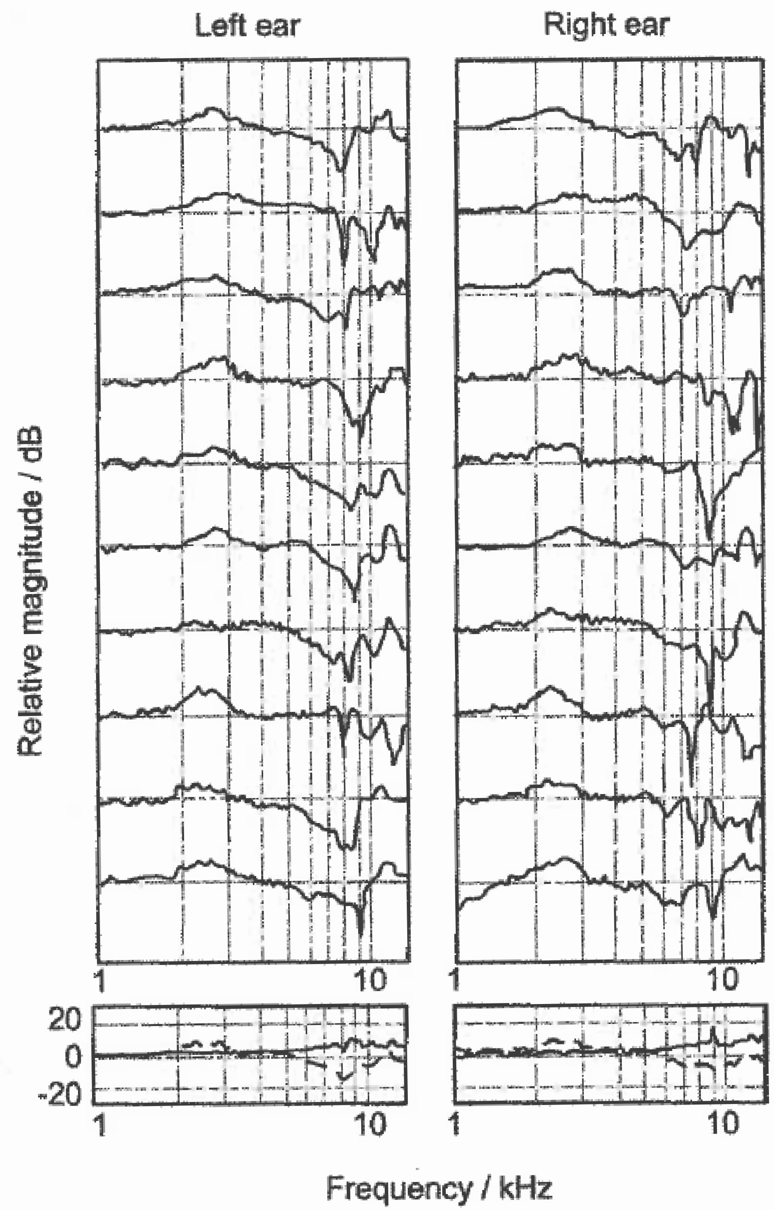


Figure 8.3 Four repeated HpTF measurements for three types of headphones: (a) HD 250 II; (b) MRD 7506; (c) MX 500.

Individual differences



References

- *The Technology of Binaural Listening* – Jens Blauert, 2012 (Ch 7, 11, 14)
- *Sound Source Distance Estimation in Rooms based on Statistical Properties of Binaural Signals* – Georganti, May, van de Par, Mourjopoulos, 2013
- *Head-Related Transfer Function and Virtual Auditory Display,* 2nd ed., Bosun Xie, 2013 (Ch. 5, 8)