Transaural Audio - The reproduction of binaural signals over loudspeakers

Fabio Kaiser

Outline

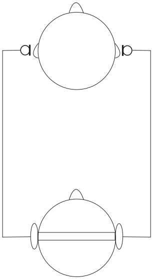
Outline

Binaural technology

- 3D immersive audio technology
- Provides interaural cues for sound localization
- Reproduction with headphones

⇒ Why not loudspeakers?

⇒ Crosstalk!!!

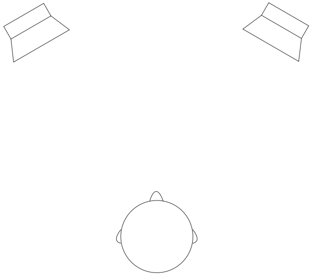


Binaural technology

- 3D immersive audio technology
- Provides interaural cues for sound localization
- Reproduction with headphones

⇒ Why not loudspeakers?

⇒ Crosstalk!!!

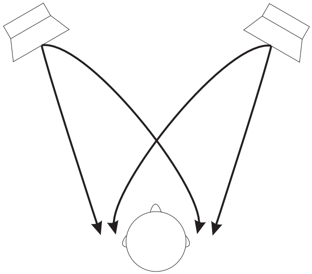


Binaural technology

- 3D immersive audio technology
- Provides interaural cues for sound localization
- Reproduction with headphones

⇒ Why not loudspeakers?

⇒ Crosstalk!!!

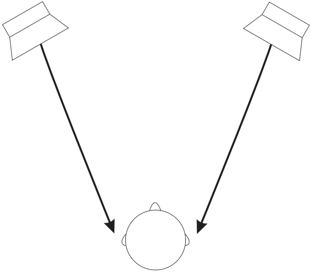


Binaural technology

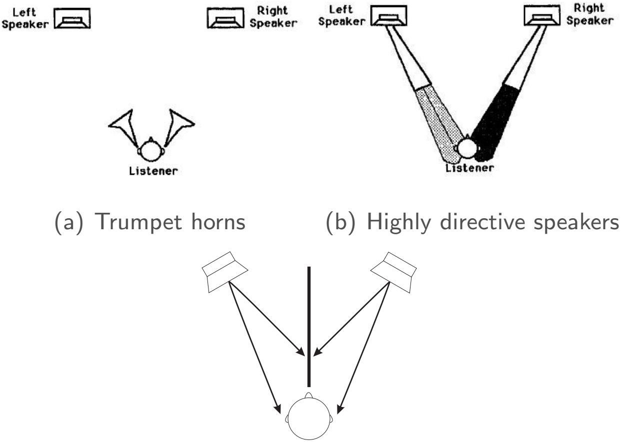
- 3D immersive audio technology
- Provides interaural cues for sound localization
- Reproduction with headphones

⇒ Why not loudspeakers?

⇒ Crosstalk!!!



Acoustical means for crosstalk cancellation



(a) Trumpet horns

(b) Highly directive speakers

(c) Physical barrier

Fig. 1: A few suggestions for acoustical CTC [Boc86]

This work

$\Rightarrow$ Signal processing techniques for crosstalk cancellation

Introduction to transaural audio

- Virtual headphones

⇒ Crosstalk cancellation (CTC) + direct path cancellation

- History
  - Bauer 1961 [Bau61]
  - Atal and Schroeder 1973 - First implementation [Sch73]
  - Cooper and Bauck 1989 - Shuffler filters [CB89]
  - Kirkeby et al. 1998 - Stereo Dipole [KN98]
  - Takeuchi et al. 2000 - Optimal Source Distribution (OSD) [TN00]

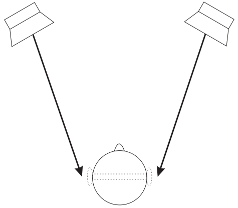
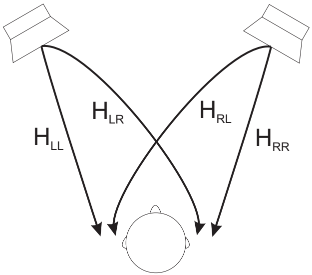


Fig. 2: Virtual headphones

Problem formulation (1)

- Matrix formulation (frequency domain)
- The head transfer matrix **H** contains the acoustic paths inherent to the plant



$$\mathbf{e} = \mathbf{H}\mathbf{y} \tag{1}$$

$$\mathbf{y} = \begin{bmatrix} y_L \\ y_R \end{bmatrix} \quad , \quad \mathbf{e} = \begin{bmatrix} e_L \\ e_R \end{bmatrix} \quad , \quad \mathbf{H} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \tag{2}$$

- **y**...loudspeakers signals
- **e**...ear signals

Problem formulation (2)

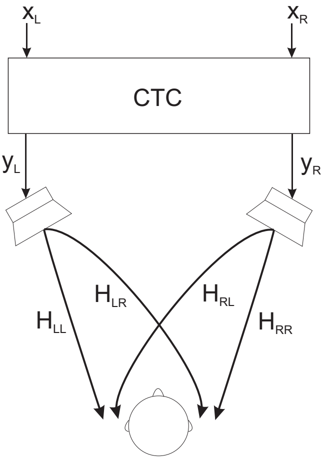
- Process input signals $\mathbf{x} = \begin{bmatrix} x_L \\ x_R \end{bmatrix}$
- Cancellation matrix **C** cancel for the head transfer matrix **H**

$$\mathbf{e} = \mathbf{HCx} \tag{3}$$

$$\mathbf{C} = \mathbf{H}^{-1} = \frac{1}{D} \begin{bmatrix} H_{RR} & -H_{RL} \\ -H_{LR} & H_{LL} \end{bmatrix} \tag{4}$$

$$D = H_{LL} H_{RR} - H_{LR} H_{RL} \tag{5}$$

⇒ **Inversion problem**



Design questions

- Static vs. dynamic
- Two-channel vs. Multi-channel
- Location of transducers
- Matched vs. mismatched plant
- Implementation method and computational complexity

Outline

Inversion of non-minimum phase filters - Theory (1)

- Unilateral z-transform of a discrete-time signal $h[n]$ (FIR)

$$H(z) = \sum_{n=0}^{\infty} \frac{h[n]}{z^n} \tag{6}$$

- The direct inverse is

$$G(z) = \frac{1}{H(z)} \tag{7}$$

- If one zero of $H(z)$ lies outside the unit circle

$\Rightarrow$ $G(z)$ is instable

Inversion of non-minimum phase filters - Theory (2)

- Head-related impulse response (HRIR)
  - Finite impulse response (FIR)
  - Typically 128-512 samples
- Mixed phase characteristics
  - Non-minimum phase part due to reflections in the pinna
  - Sound energy focused to the ear canal after direct sound arrived

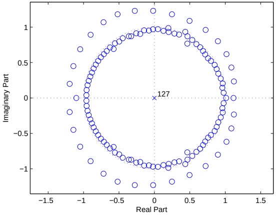


Fig. 3: Zero-pole plot of a HRIR ($\phi = 30°, \theta = 0°$, length is 128 at 44.1kHz)

Inversion of mixed phase filters - Theory (3)

- Choose region of convergence (ROC) to include unit circle
- Inverse $g(n)$ then is
  - Stable
  - Two-sided
  - Infinite

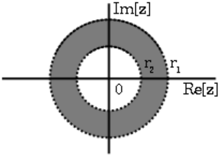


Fig. 4: ROC of a two-sided sequence

- Window $g(n)$ to obtain a FIR filter

$$\tilde{g}[n] = g[n] \cdot w[n], w[n] = \begin{cases} 1, L < n < M \\ 0, otherwise \end{cases} \quad (8)$$

$\Rightarrow$ Maximal energy inside window

- Delay if $\tilde{g}[n]$ is anti-causal

Inversion using the DFT

- Discrete Fourier transform
- Inversion in the frequency domain
- Regularization by limitation of magnitude
- Procedure
  - Compute $H(k)$, the DFT of $h[n]$
  - Invert $H(k)$
  - Limit the resulting magnitude response
  - Compute the inverse DFT
  - Delay result if necessary
- Performance is determined by DFT size (time-aliasing)
- Limitation yields poles with shorter decay time

Least-squares technique (1)

- Error function $e[n]$

$$e[n] = \delta[n] - \tilde{\delta}[n] \tag{9}$$

- $\delta[n]$...desired output
- $\tilde{\delta}[n]$...actual output
- Minimize squared error

$$J = \frac{1}{L}\sum_{n=0}^{L} e^2[n] = \frac{1}{L}\sum_{n=0}^{L}(\delta[n] - \tilde{\delta}[n])^2 \tag{10}$$

Least-squares technique (2)

- Write deconvolution in matrix form

$$\mathbf{d} = \mathbf{H}\tilde{\mathbf{g}} \tag{11}$$

- $\mathbf{d}$... desired output (with delay) of length L
- $\mathbf{H}$...Toeplitz matrix of size $L \times M$
- $\tilde{\mathbf{g}}$...inverse filter of length M
- The error function is then

$$\mathbf{e} = \mathbf{d} - \mathbf{H}\tilde{\mathbf{g}} \tag{12}$$

- Minimize squared error $||e||^2 = e^T e$
- Set gradient to zero
- This yields

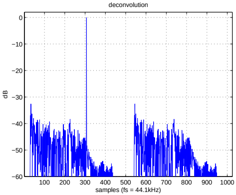
$$\tilde{\mathbf{g}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{d} \tag{13}$$

Minimum phase inversion

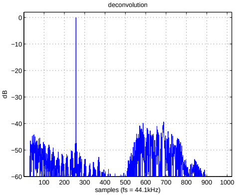
- HRIR is of minimum phase
  - Except for a nearly frequency independent delay
- The delay is ignorable for the inversion (Schroeder)
- Procedure
  - Invert the magnitude response only
  - Compute the minimum phase by using the Hilbert transform of the log-magnitude spectrum

$$arg[H(j\omega)] = -\mathcal{H}\left\{log(|H(j\omega)|)\right\} \tag{14}$$

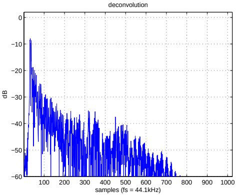
Comparison



(a) DFT inversion



(b) Least-squares



(c) Min phase inversion

Fig. 5: Deconvolution using a filter length of 512 samples.

Outline

Implementation of CTC

Problem

How to implement the crosstalk cancellation matrix **C**?

- First Atal and Schroeder 1973
- Assumed a perfectly symmetric loudspeaker setup
- **H** can be rewritten to

$$\mathbf{H} = \begin{bmatrix} S & A \\ A & S \end{bmatrix}, \mathbf{C} = \frac{1}{S^2 - A^2} \begin{bmatrix} S & -A \\ -A & S \end{bmatrix} \tag{15}$$

- S...ipsilateral HRTF
- A...contralateral HRTF

Implementation of CTC

- Reordering leads to their implementation

$$\mathbf{C} = \frac{\frac{1}{S}}{1 - (\frac{A}{S})^2} \begin{bmatrix} 1 & -\frac{A}{S} \\ -\frac{A}{S} & 1 \end{bmatrix} \qquad (16)$$

- The capital $C = -\frac{A}{S}$, not to confuse with the inverse head transfer matrix $\mathbf{C}$

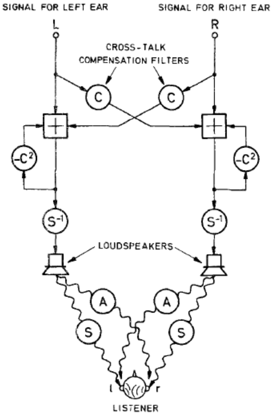


Fig. 6: The crosstalk canceler by Atal and Schroeder [Sch73]

General topology

- Direct implementation of

$$\mathbf{C} = \mathbf{H}^{-1} = \frac{1}{D}\begin{bmatrix} H_{RR} & -H_{RL} \\ -H_{LR} & H_{LL} \end{bmatrix} \tag{17}$$

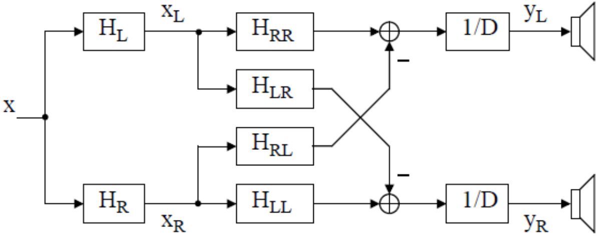


Fig. 7: Block diagram of the general topology including the binaural synthesis stage [Gar97].

Shuffler topology (1)

- Cooper and Bauck 1989
- Reformulation of **H**

Lemma

A $n \times n$ *matrix* **A** *is diagonalizable if it can be written as*

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \tag{18}$$

- **D**...$n \times n$ diagonal matrix containing the Eigenvalues of **A**
- **P**...$n \times n$ matrix containing the corresponding Eigenvectors

Shuffler topology (2)

- Assume $\mathbf{H}$ to be symmetric
- Diagonalize $\mathbf{H}$
- The inverse can be written to

$$\mathbf{H}^{-1} = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\frac{1}{S+A} & 1\\ 1 & \frac{1}{S-A}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}. \tag{19}$$

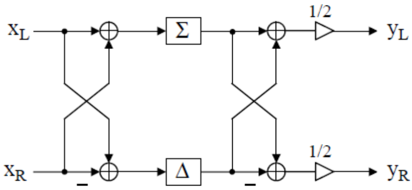


Fig. 8: Shuffler topology. Here Σ denotes the sum filter and Δ the difference filter [Gar97].

Shuffler topology (3)

- Two filters are of joint-minimum phase
  - Excess phase is equal
  - Excess phase is a frequency independent delay
- Sum and difference filter
  - Excess phases ≈ the same
  - Excess phase ≈ frequency independent delay

⇒ Excess phase can be ommited

- Invert the minimum phase part only

Asymmetric shuffler form

- Vandernoot PhD 2001
- Topology is equivalent to shuffler form

$$
\begin{aligned}
H_{LL} &= \frac{H_{RR} + H_{RL}}{D} \\
H_{LR} &= \frac{H_{LR} + H_{LL}}{D} \\
H_{RL} &= \frac{H_{RR} - H_{RL}}{D} \\
H_{RR} &= \frac{-H_{LR} + H_{LL}}{D}
\end{aligned}
\tag{20}
$$

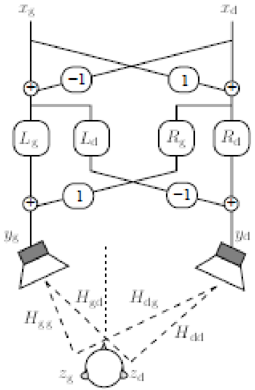


Fig. 9: Asymmetric shuffler topology [Van01].

Fast least-squares deconvolution (1)

- Kirkeby et al. 1998 [KNHOB98]
- Based on the least-squares approximation in frequency domain
- Cost function
  - Performance error
  - Effort penalty

$$J(z) = e^H(z)e(z) + \beta v^H(z)v(z) \tag{21}$$

- β...regularization parameter
- Finding the minimum of the cost function yields

$$\mathbf{C}(z) = \left[\mathbf{H}^T(z^{-1})\mathbf{H}(z) + \beta\mathbf{I}\right]^{-1}\mathbf{H}^T(z^{-1})\mathbf{A}(z) \tag{22}$$

Fast least-squares deconvolution (2)

- Using the fast fourier transform (FFT)

$$\mathbf{C}(k) = \left[\mathbf{H}^T(k)\mathbf{H}(k) + \beta\mathbf{I}\right]^{-1}\mathbf{H}^T(k)\mathbf{A}(k) \qquad (23)$$

- k...k-th frequncy index
- Implementation summery
  - Calculate the N-point FFT of the relevant HRIRs
  - Calculate **C** by applying Eq.(23)
  - Calculate the N-point inverse FFT of **C**
  - Introduce a modeling delay by doing a cyclic shift of m samples of the resulted IRs. A good value for m is suggested by N/2.
- β has to be set appropriately
- $\beta = 0.0001$ is suggested
- Topology is the direct implementation of **C**

Recursive form (1)

- Iwahara and Mori 1978
- Very intuitive
- Derived by Step-by-step analysis of the crosstalk process
- Interaural transfer function $ITF = \frac{H_c}{H_i}$
  - Predicts crosstalk
- Compensates for higher-order crosstalk

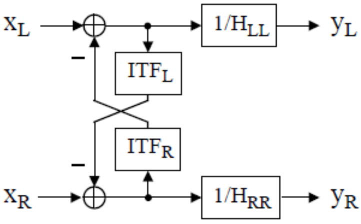


Fig. 10: Block diagram of the recursive topology [Gar97].

Recursive form (2)

- Feedback loop is a power series of ITFs

$$1 + ITF(\omega) + ITF(\omega)^2 + ITF(\omega)^3 + \cdots = \frac{1}{1 - ITF(\omega)} \tag{24}$$

- Instable if absolute values of ITF are
  - Equal to one or
  - Close to one (due to numerical inaccuracies)
- Commercial system *Ambiophonics*
  - Avoids instabilities by band-limitation (250Hz-6kHz)

Summary

- Quantitative comparison

	General	Shuffler	A. shuffler	Fast LS	Recursive
Convolutions	4-6	2	4	4	4
Asymmetry	+	-	+	+	+

Outline

Objective evaluation

- Objective analysis of the conditioning of the problem
- Objective comparison of different implementations
  - Free-field simulation
- HRTF data
  - IRCAM LISTEN database
  - Diffuse field equalized
  - 512 samples

Condition number

- Linear equation system $b = \mathbf{A}x$
  - Well-conditioned $\Rightarrow$ small condition number
  - Ill-conditioned $\Rightarrow$ high condition number
- Measures the sensitivity to numerical errors
- Ratio of the largest to smallest singular value of a matrix

$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \tag{25}$$

- $\kappa(A) = \infty$

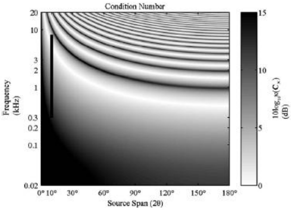
$\Rightarrow$ System is singular

- $\kappa(A) = 1$

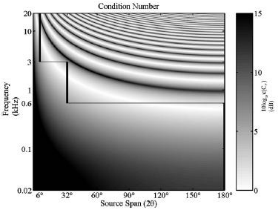
$\Rightarrow$ Perfect inversion is possible

Conditioning of the head transfer matrix

- Ward and Elko 1999 [WE99]
- Condition Number
  - Frequency dependent
  - Source span dependent
- Takeuchi 2000
  - Optimal source distribution (OSD)



(a) Stereo dipole [NR05]



(b) OSD

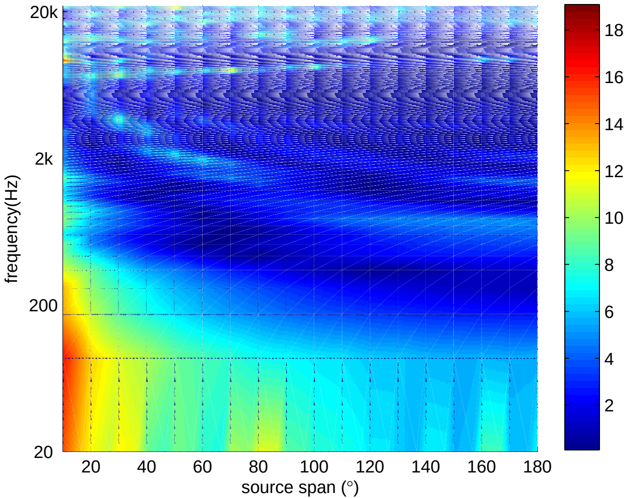


Fig. 11: Condition Numbers as a function of frequency and source span of a measured set of HRTFs. The colorbar indicates $10log_{10}(\kappa(C))$.

Average best-fit

- Frequency average of condition number
- Find source span with smallest average

- Advantage
  - Conventional transducer system
  - Providing the smallest errors
- Disadvantage
  - Local errors at high frequencies
  - Individual differences

Average best-fit

- Source spans in between $130^\circ - 180^\circ$

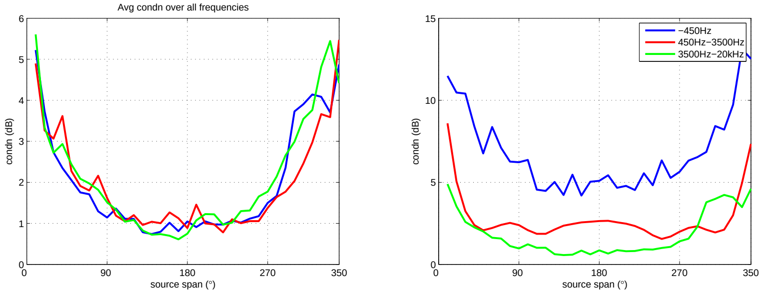


Fig. 12: Frequency average of the condition number. Left, of three different head transfer matrices (subjects). Right, of one head transfer matrix (subject) divided into three sub-bands.

Comparison of implementations - Setup (1)

- Free-field simulation in Matlab
- Measure amount of CTC
- Symmetric source spans only
  - $\pm 5°$, $\pm 30°$ and $\pm 65°$
- Evaluate the different topologies
- Input to system

$$x_L(t) = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$
$$x_R(t) = 0 \tag{26}$$

$\Rightarrow$ Impulse response of system $\mathbf{y}$

- Convolution with plant HRIRs

$\Rightarrow$ Ear signals $\mathbf{e}$

Comparison of implementations - Setup (2)

- Topologies compared
  - General topology (blue)
  - Symmetric shuffler topology (red)
  - Fast least squares deconvolution ($\beta = 0.0001$) (green)
- Asymmetric shuffler topology
  - Output equivalent to general topology
- Recursive topology
  - Instable
- Filter computation
  - Frequency domain
  - 1024-point FFT

Results

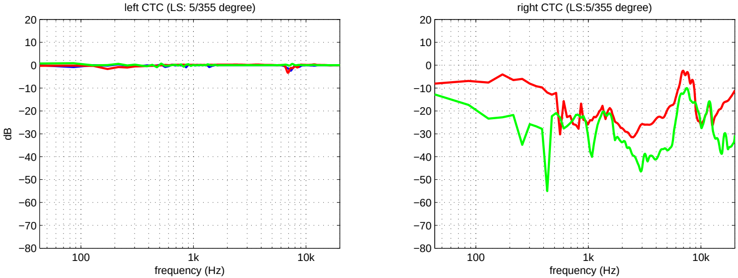


Fig. 13: Magnitude spectra of the acoustic paths between left speaker/left ear (left plot) and left speaker/right ear (right plot). Source span: $\pm5^\circ$ (first), $\pm30^\circ$ (second) and $\pm65^\circ$ (third).

Results

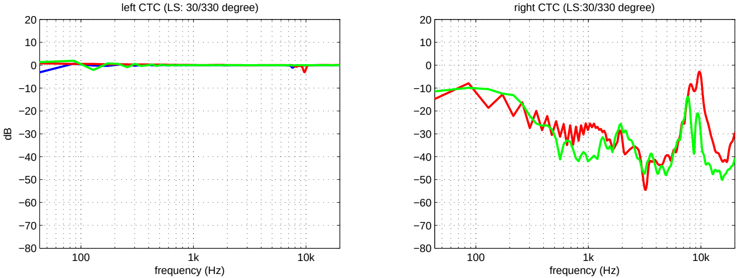


Fig. 13: Magnitude spectra of the acoustic paths between left speaker/left ear (left plot) and left speaker/right ear (right plot). Source span: $\pm 5^\circ$ (first), $\pm 30^\circ$ (second) and $\pm 65^\circ$ (third).

Results

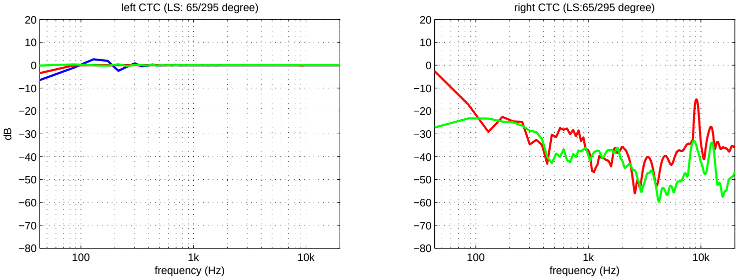


Fig. 13: Magnitude spectra of the acoustic paths between left speaker/left ear (left plot) and left speaker/right ear (right plot). Source span: $\pm 5^\circ$ (first), $\pm 30^\circ$ (second) and $\pm 65^\circ$ (third).

Outline

Conclusions and prospects (1)

- Design questions
  - Two-channel static transaural system

Position of loudspeakers

- OSD
- Average best-fit
- Multi-channel approach [HBC07]

Implementation methods/topologies

- General topology (direct implementation)
- A couple of alternative methods exist
- Adaptive approach [NHE92]
- Dynamic approach [Len06]

Conclusions and prospect (2)

Matched vs. mismatched CTC filters

- Mismatched filters deteriorate performance [ACB+07]
- Plant identification

Influence of reflections

- Severe early reflections lead to distortion [Sae04]

Main challenges

- Robustness against movements
- Sound coloration

Thank you for your attention!

Questions?

M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1056–1069, 2007. [Online]. Available: http://link.aip.org/link/?JAS/121/1056/1

B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *J. Audio Eng. Soc*, vol. 9, no. 2, pp. 148–151, 1961. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=471

D. B. Bock, Timothy M.; Keele, "The effects of interaural crosstalk on stereo reproduction and minimizing interaural crosstalk in nearfield monitoring by the use of a physical barrier: Part 1," in *Audio Engineering Society Convention 81*, 11 1986.

[Online]. Available:
http://www.aes.org/e-lib/browse.cfm?elib=5010

D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *Journal of the Audio Engineering Society*, vol. 37, no. 1–2, pp. 3–19, 1989.

W. Gardner, "3-d audio using loudspeakers(phd)," Ph.D. dissertation, MIT Media Lab, 1997.

Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-d sound reproduction," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 649–652, 2007.

O. Kirkeby and P. A. Nelson, "The "stereo dipole" - a virtual source imaging system using two closely spaced loudspeakers," *Journal of the Audio Engineering Society*, 1998.

O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on speech and audio processing*, vol. 6, no. 2, pp. 189–194, 1998.

T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *Journal of the Audio Engineering Society*, 2006.

P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *Signal Processing, IEEE Transactions on*, vol. 40, no. 7, pp. 1621–1632, 1992.

P. A. Nelson and J. F. W. Rose, "Errors in two-point sound reproduction," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 193–204, 2005.

M. R. Schroeder, "Computer models for concert hall acoustics," *Am. J. Phys.*, vol. 41, pp. 461–471, 1973.

T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *J. Acoust. Soc. Am.*, 2000.

G. Vandernoot, "Caracterisation et optimisation de la restitution haute-fidelite en vehicule," Ph.D. dissertation, UNIVERSITE PARIS 6, 2001.

D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," *Signal Processing Letters, IEEE*, vol. 6, no. 5, pp. 106–108, 1999.