Transaural Rendering

CMSC 828D / Spring 2006

Lecture 19

Recap (1)

- Acoustic source produces a sound
- This sound arrives at the ear
- Sound is modified because of scattering:
  - Environmental reflections (room acoustics)
  - Anatomical reflections (HRTFs)
- So sound at the eardrum is not the same as sound produced by the source

Recap (2)

- HRTF defines how sound is modified by anatomical scattering
  - Source is at (r,θ,φ)
  - $H_L(\omega,r,\theta,\varphi) = \Phi_L(\omega,r,\theta,\varphi)/\Phi(\omega)$
  - $H_R(\omega,r,\theta,\varphi) = \Phi_R(\omega,r,\theta,\varphi)/\Phi(\omega)$
  - $\Phi_L(\omega,r,\theta,\varphi)$ is the potential at the left eardrum
  - $\Phi(\omega)$ is the potential at the center of the head as if the head were absent

Recap (3)

- HRTF describes how spectrum is modified for sources at various directions
- Spectrum changes are picked by the brain to evoke perception of direction
- So we can “trick” the brain
- Identical stimuli cause identical perceptions
  - Present some waveforms to the eardrums
  - User gets a corresponding perception

Recap (4)

- Present to the ears the sound filtered with HRTF for direction (θ,φ)
- Listener *should* perceive the sound as coming from that direction
- But… various difficulties arise

Individualization

Recap (5)

- Problems complicating such an easy solution:
  - Individualization
  - Environment
  - Dynamics
- All these can be taken care of properly
- When this is done, signals $x_L(t)$ and $x_R(t)$ to be delivered to eardrums are produced…

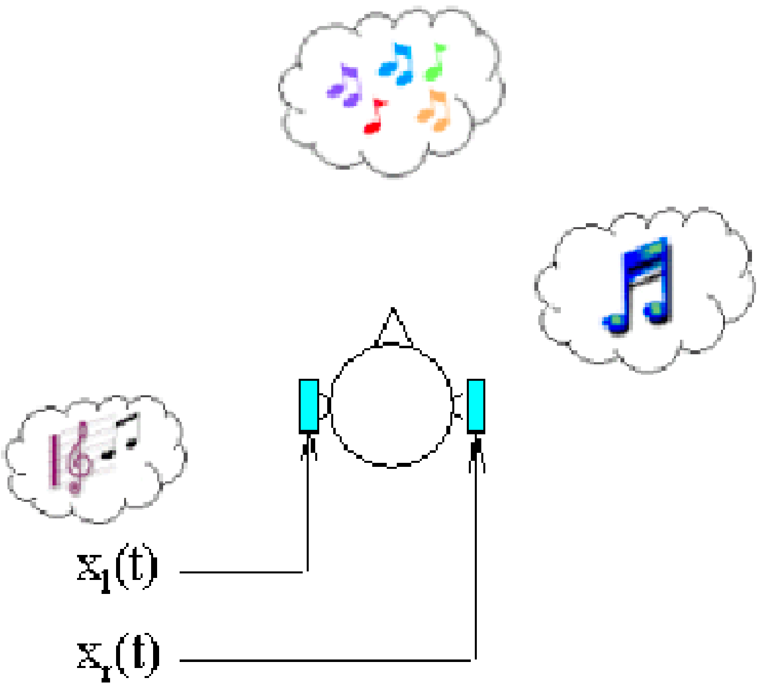
Binaural Delivery

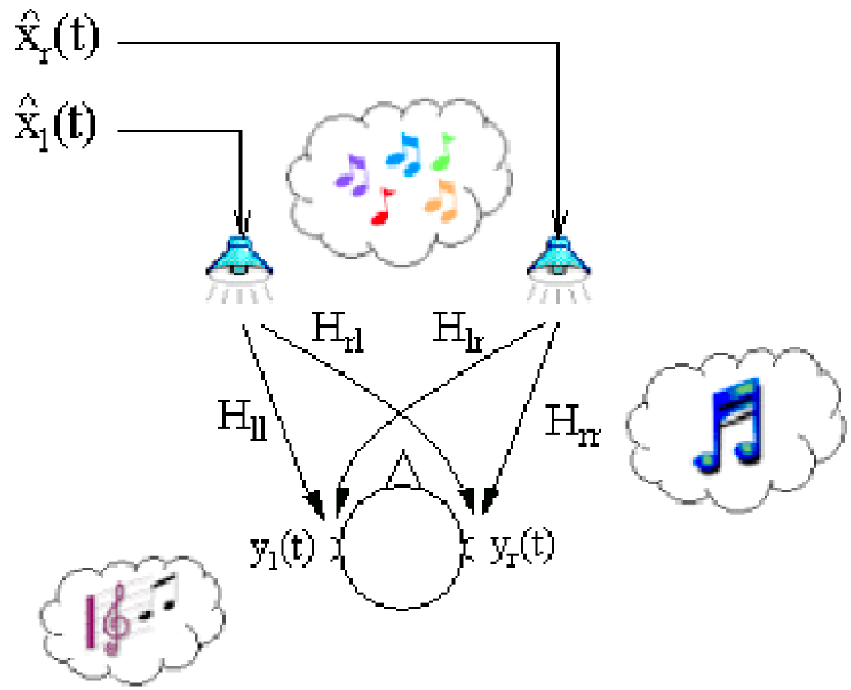
- We're given signals to be delivered to the left and right ears
- Put headphones on person
- Possibly compensate for
  - Ear canal response
  - Headphone response
  - Headphone-to-head coupling

Transaural Delivery

- Use loudspeakers
- “Natural” solution for desktop computers
  - Monitor and two speakers on the sides
- More challenging case
  - Crosstalk
  - User positioning (“sweet spot”)

Illustration





Binaural presentation

Transaural presentation

Transaural Setup

- Assume that signals $\hat{x}_L(t)$ and $\hat{x}_R(t)$ are played via loudspeakers
- Assume symmetric listening case
- $H_{LL}=H_{RR}=H_i$, $H_{LR}=H_{RL}=H_c$
- Then,
  - $Y_L(\omega)=\hat{X}_L(\omega)H_i(\omega)+\hat{X}_R(\omega)H_c(\omega)$
  - $Y_R(\omega)=\hat{X}_L(\omega)H_c(\omega)+\hat{X}_R(\omega)H_i(\omega)$

Matrix Form

- In matrix form, $Y = H\hat{X}$, where
- $Y = \begin{bmatrix} Y_L \\ Y_R \end{bmatrix}$, $H = \begin{bmatrix} H_i & H_c \\ H_c & H_i \end{bmatrix}$, $\hat{X} = \begin{bmatrix} \hat{X}_L \\ \hat{X}_R \end{bmatrix}$
- We need to choose $\hat{X}$ so that $Y = X$

Crosstalk Cancellation Filter

- Filter X to generate such $\hat{X}$ that $H\hat{X} = X$
- Crosstalk can be represented as filter
- Knowing crosstalk filter, design a *crosstalk-canceling* filter G for X, $\hat{X} = GX$
  - So that the crosstalk, when occurs later in the delivery path, undoes the effect of G
- We need simply $G=H^{-1}$
- Then $Y=H\hat{X}=HGX=HH^{-1}X=X$

Crosstalk Cancellation

- $H^{-1} = \dfrac{1}{H_i^2 - H_c^2} \begin{bmatrix} H_i & -H_c \\ -H_c & H_i \end{bmatrix}$

- Can also do expression for non-symmetric setup
- Slightly more complicated

Structural Chart

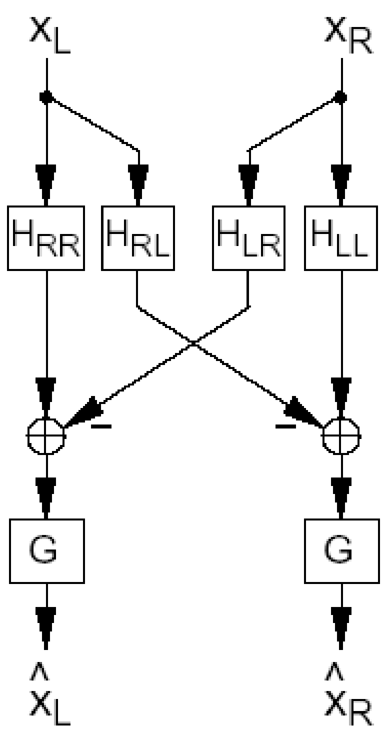


Figure 4: General transaural filter, where $G = 1/(H_{LL}H_{RR} - H_{LR}H_{RL})$.

(from Gardner 1995)

Shuffler Implementation

- Can be easier to implement
- More elegant solution

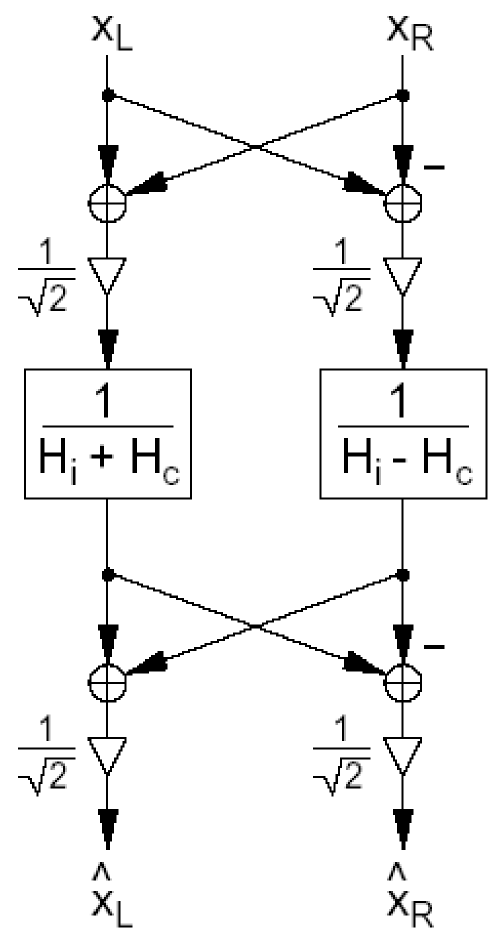
- Shuffler matrix D = $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

- Note that $D^{-1}=D$ and $D^{-1}H^{-1}D=\begin{bmatrix} (H_i+H_c)^{-1} & 0 \\ 0 & (H_i+H_c)^{-1} \end{bmatrix}$

- Denote $Q= D^{-1}H^{-1}D$

Shuffler Implementation

- Can then write $\hat{X} = H^{-1}X = DD^{-1}H^{-1}DD^{-1}X = D(D^{-1}H^{-1}D)D^{-1}X = DQD^{-1}X = DQDX$
- Easier implementation
- (from Gardner 1995)



Applications

- Now we can render transaurally *any* signal that was prepared for binaural delivery
  - Just process it with filter and output via speakers
- Transaural rendering is targeted mainly at desktop applications
  - Typical placement of speakers is at ±30 degrees
  - User is stationary, is centered, and is facing the monitor

Limitations

- High-quality audio scene is achieved only at “sweet spot”
- Scene breaks down completely with 30 cm left or right motion or 30 degree rotation
- Reasonably tolerant to front-back motion
- Head tracking can be done to adapt filters in real time

HRTF Approximation

- Personalized HRTF is often not available
- Use approximate models of HRTF
  - E.g., use sphere HRTF
  - Valid only for low frequencies
  - But at high frequencies, crosstalk is low anyway in typical setup
  - Head shadows left ear from right loudspeaker, and vice versa

Signal Processing Issues

- Need to divide by HRTF
  - Division by a “non-minimum-phase” filter lead to instability
  - Either convert HRTF to min-phase format
  - In this case some information is lost but it is believed to be perceptually unimportant
  - Or use iterative methods to devise cross-cancellation filter
  - So implementation is somewhat complicated
  - See literature for some pointers

Existing Applications

- Technology is used in cell phones for surround sound
- Motorola E398 and ROKR-E1 cell phones
  - Non-individualized HRTF
  - Two loudspeakers on the sides
- Desktop applications as well

Brief Literature List

- W. G. Gardner, “Transaural 3-D Audio”, M.I.T. Media Lab Technical Report #342, 1995.
- C. Kyriakakis, P. Tsakalides, and T. Holman, “Surrounded by sound”, IEEE Signal Processing Magazine, January 1999.
- T. Funkhouser, J.-M. Jot, and N. Tsingos, “Sounds good to me!”, SIGGRAPH 2002 Course Notes.
- A. Mouchartis, P. Reveliotis, and C. Kyriakakis, “Inverse filter design for immersive audio rendering over loudspeakers”, IEEE Transactions on Multimedia, vol. 2(2), June 2000.

Web Pointers

- http://www.sonicspot.com/guide/3daudio.html
- http://www.sensaura.com/whitepapers/pdfs/dev009.pdf
- www.sonaptic.com

Multi-speaker Presentation

- Another way to deliver spatial content
- Place a user in the room
- Place many speakers around him/her
- Play sound from necessary direction
  - No HRTF filtering at all
  - Sound just comes from the correct place

Simplest Case: Stereo

- Based heavily on the way human brain processes the sound
  - Combination of two sources is perceived as “phantom image”
  - Located between true sources

Stereo Setup

- Place two speakers at ±30 degrees
- Same loudness in both speakers: perceived source is in the middle
- Increase loudness in one speaker, or
- Delay one speaker w.r.t. another one:
  - The perceived position shifts towards the louder or the earlier speaker
- Also suffers from "sweet spot" problem

Multi-speaker Setup

- Stereo setup can be extended
- Have many speakers covering whole space
- Vector-Based Amplitude Panning (VBAP)
- To render sound at direction D:
  - Find three speakers at directions D_i closest to D
  - Compute gains G_i so that $\Sigma G_i D_i = D$
  - Play sound from these speakers with gains G_i

VBAP

- Good for rendering “far” virtual sources
- Impossible for the virtual source to enter reproduction region
  - I.e. to be closer to the listener than the actual loudspeaker
- Also, implementation problems often arise
  - An array of 128 loudspeakers was built once
  - It very effectively focused all noise in the room to the room center

5.1 and Similar Setups

- Consumer-level surround sound systems
- All speakers are in the same plane
  - Unable to reproduce elevation
- Large angles between speakers
  - Hard to create stable phantom images
- So these setups are more for “enveloping” the user rather than for accurate spatial presentation

5.1 Setup

- Two frontal speakers are at ±30 degrees
  - Compatibility with stereo setup
- Center channel is in front
  - These three form “conventional” stereo setup
- Two “surround” speakers are at ±110 deg
  - So they are placed on the sides
  - Intention is to generate supporting ambience, effects, or “room impression”
- No intention to do full 360 deg spatial audio

Wave-field Synthesis

- Kirchoff principle:
  - Any point on a waveform acts as a secondary source
- One can reconstruct the wave field *exactly* within a region
  - Model or measure the wave on the boundary
  - Place many sources at the region boundary
  - Make them emit the same wave into the region

Wave-field Synthesis

- Advantages:
  - Excellent quality
  - Exact reconstruction everywhere in the region
  - No more sweet spot problem!
- Disadvantages:
  - Many speakers necessary
  - Heavy setup and heavy processing
  - Still can’t simulate source(s) within the region
- … A separate lecture on WFS will follow

Conclusion

- Binaural rendering
- Transaural rendering
- Multi-speaker rendering (VBAP)
- Wave-field synthesis
  - Important to understand advantages and limitations of each of these approaches
- Next lecture: Room Acoustics!