$\Bigg(\bigg((\cdot)\bigg)\Bigg)$

AES 22nd, Espoo, Finland, June 2002.

CODING PRINCIPLES FOR VIRTUAL ACOUSTIC OPENINGS

Aki Härmä
Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Espoo, Finland

This work was mainly done at *Media Signal Processing Research, Agere Systems, Murray Hill, NJ, USA (old Bell Labs dep.)*

http://www.acoustics.hut.fi/~aqi

Summary

- Acoustic opening: history and application ideas
- The coding problem
- Generic formulation for a coder
- Different approaches: pros and cons
- Conclusions
- Future work

Acoustic opening

- Multichannel audio communications system.
- Imitation of an open window between two rooms.
- Transducer arrays
- Macroscopic wavefield reconstruction: sweet all space, naturalness.

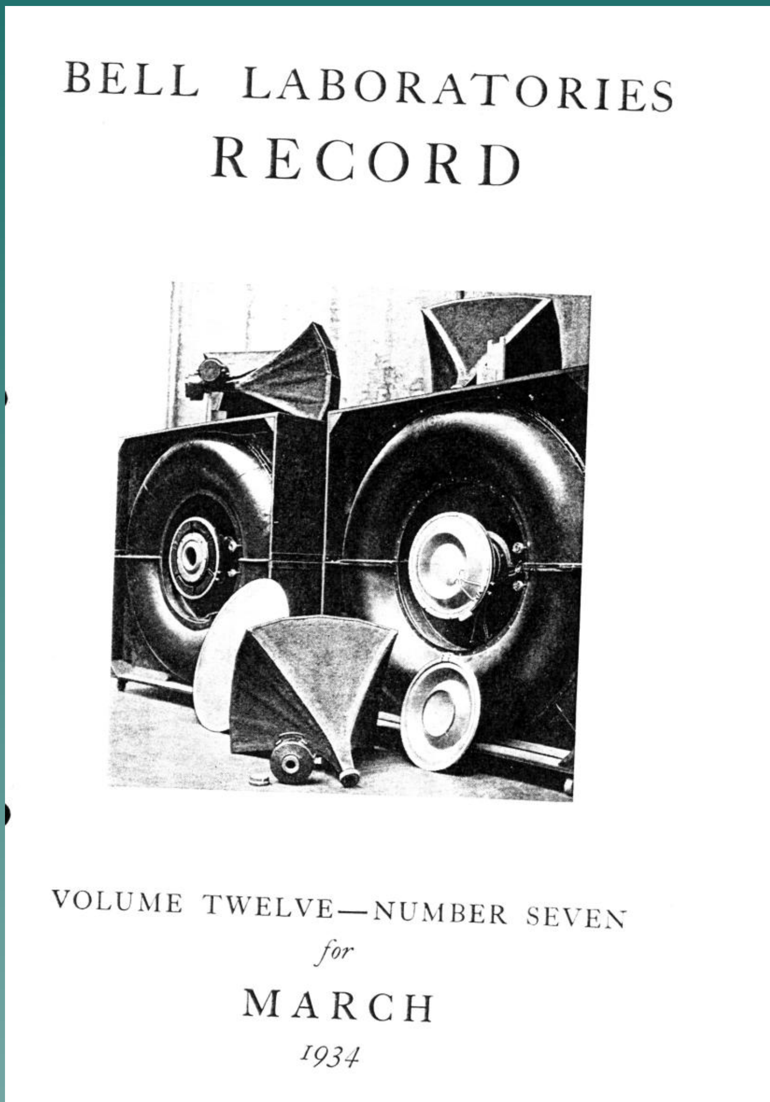
Acoustic opening

- Multichannel audio communications system.
- Imitation of an open window between two rooms.
- Transducer arrays
- Macroscopic wavefield reconstruction: sweet all space, naturalness.
- *Huyghen's principle*
- Acoustic holography
- **Wave field synthesis (WFS)** (*Berkhout (1993) et al.*)

Acoustic opening

- Multichannel audio communications system.
- Imitation of an open window between two rooms.
- Transducer arrays
- Macroscopic wavefield reconstruction: sweet all space, naturalness.
- *Huyghen's principle*
- Acoustic holography
- **Wave field synthesis (WFS)** (*Berkhout (1993) et al.*)
- Introduced almost 70 years ago (*Snow (1934), 'Auditory Perspective',...*)

Snow (1934), 'Auditory Perspective',...



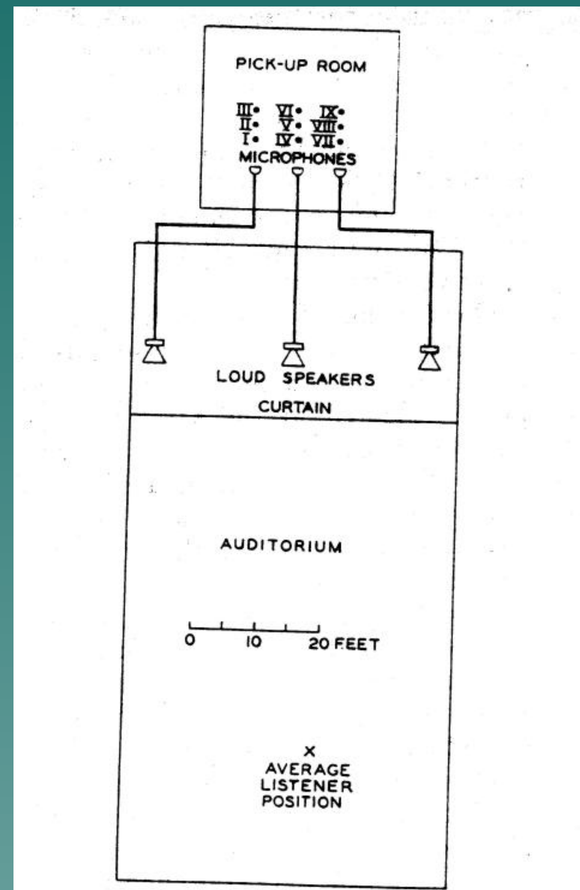
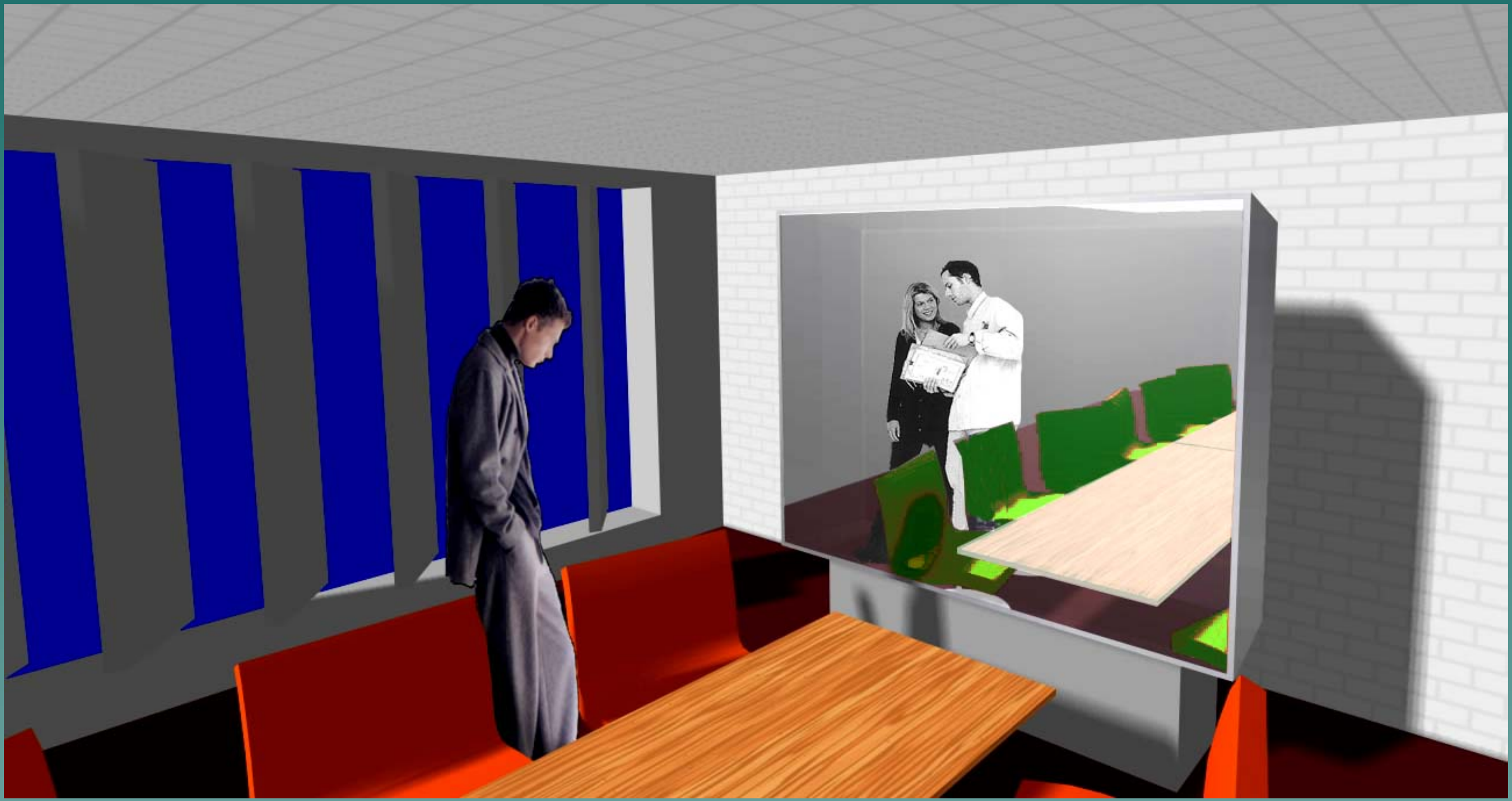


Fig. 1—For experimental purposes, pick-up microphones were mounted in a sound-proof room and connected by independent circuits to loud speakers on the stage of the auditorium

- Teleconference - communications



- Music, performing arts, remote rehearsal, artistic exhibits, interior design, entertainment, games.

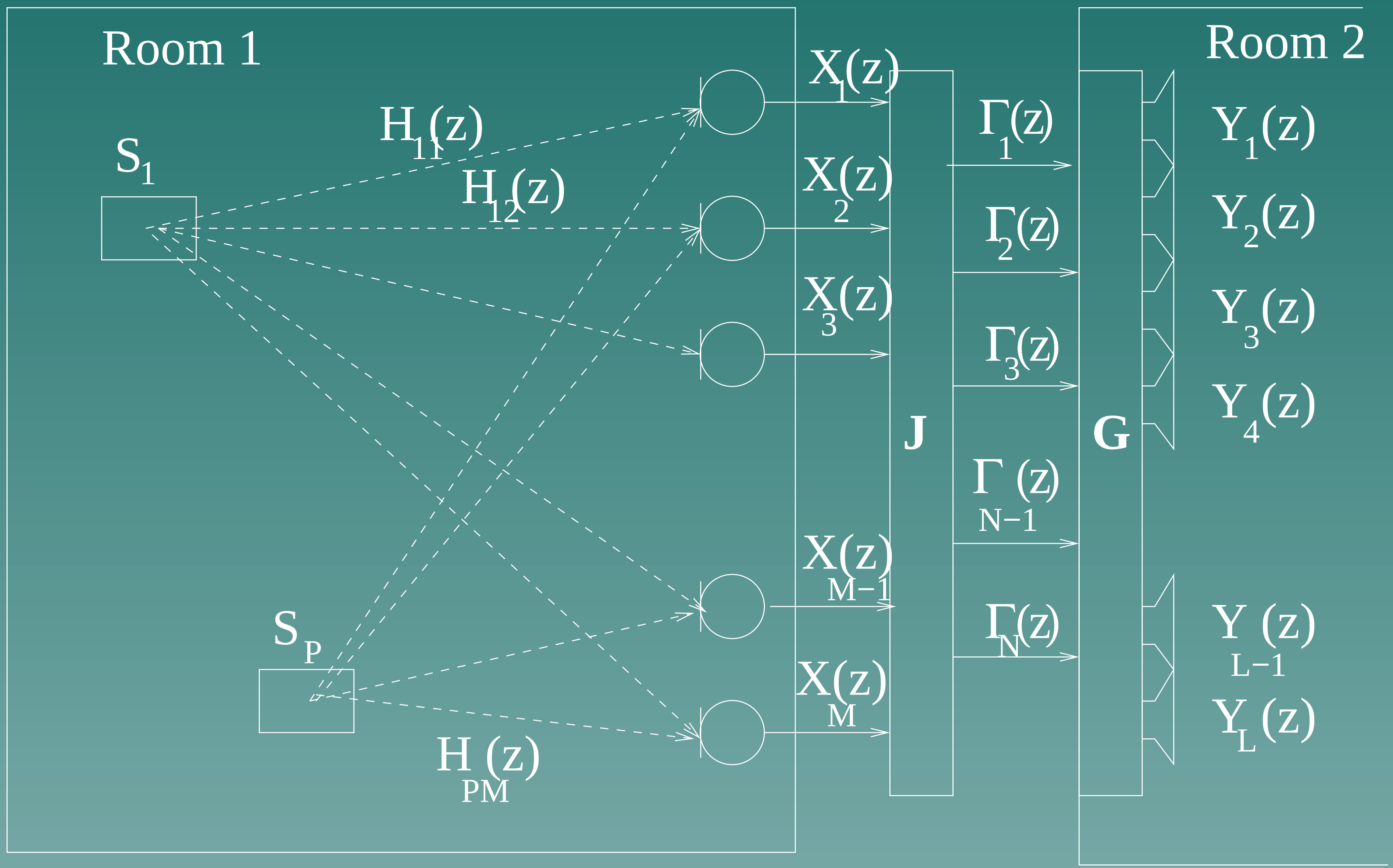
The coding problem

- *... beyond realms of possibility at present ... (Lipshitz, 1986)*
- But now,
  - ⋆ Multichannel (e.g., 48-ch) PC sound boards
  - ⋆ Cheap high-quality electret microphones
  - ⋆ New transducer technologies (DML, EMfi, etc.)

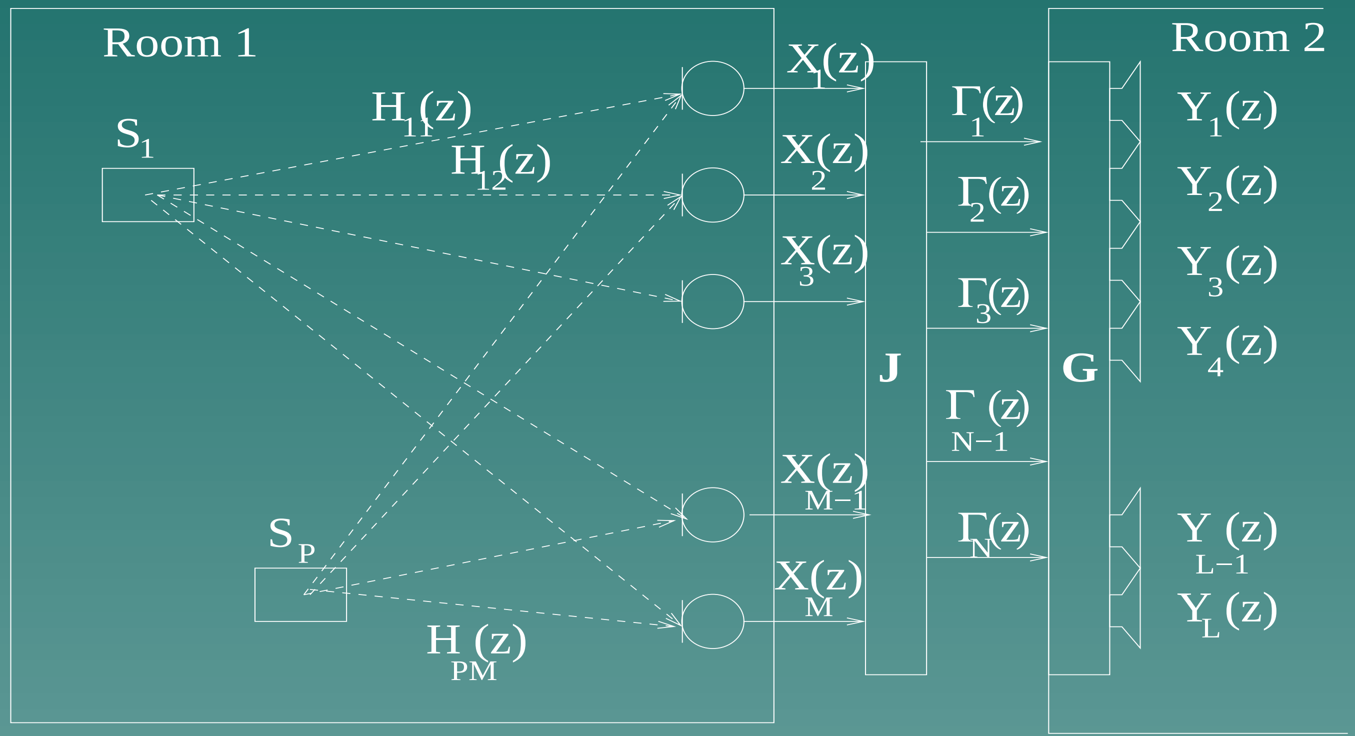
The coding problem

- *... beyond realms of possibility at present ... (Lipshitz, 1986)*
- But now,
  - ⋆ Multichannel (e.g., 48-ch) PC sound boards
  - ⋆ Cheap high-quality electret microphones
  - ⋆ New transducer technologies (DML, EMfi, etc.)
- Bitrate is still a major problem
- Large number of highly correlated nonidentical signals.
- Goal: Perceptually sufficiently accurate representation of a wavefield in an opening.

Generic formulation



Generic formulation



Encoder:

$$\mathbf{\Gamma} = \mathbf{J}\mathbf{X}$$

Decoder:

$$\mathbf{Y} = \mathbf{G}\mathbf{\Gamma}$$

- $\mathbf{\Gamma}$ - set of **generating** signals (coded and transmitted)
- $\mathbf{G}$ - **reconstruction** filters (wavefield parameters)
- **Transducer invariance**: bit-rate constant $\forall M, L$.

1. Scalar matrix $\mathbf{G}$

For example, sum-difference decoder (M/S stereo)

$$\begin{bmatrix} Y_r(z) \\ Y_l(z) \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} \Gamma_s(z) \\ \Gamma_d(z) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \Gamma_s(z) \\ \Gamma_d(z) \end{bmatrix}$$

Generalized sum-difference (Walsh-Hadamard) decoder

$$\begin{bmatrix} Y_1(z) \\ \vdots \\ Y_L(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & -1 & -1 \\ & \vdots & \vdots & \vdots & \\ 1 & -1 & \cdots & 1 & -1 \end{bmatrix} \begin{vmatrix} \Gamma_1(z) \\ \vdots \\ \Gamma_L(z) \end{vmatrix}$$

Encoding $\mathbf{\Gamma} = \mathbf{J}\mathbf{X} \approx \mathbf{G}^{-1}\mathbf{X}$

Decoding $\mathbf{Y} = \mathbf{G}\mathbf{\Gamma}$

Cosine transform

$$\mathbf{G} = \left[\begin{array}{c} \text{Array of} \\ \text{Cosine basis} \\ \text{functions} \end{array} \right]$$

Karhunen-Loeve (KLT)

$$\mathbf{G} = \left[\begin{array}{c} \text{Adaptive} \\ \text{eigenfunction} \\ \text{basis} \end{array} \right]$$

Encoding $\mathbf{\Gamma} = \mathbf{J}\mathbf{X} \approx \mathbf{G}^{-1}\mathbf{X}$

Decoding $\mathbf{Y} = \mathbf{G}\mathbf{\Gamma}$

Cosine transform

$$\mathbf{G} = \left[\begin{array}{c} \text{Array of} \\ \text{Cosine basis} \\ \text{functions} \end{array} \right]$$

Karhunen-Loeve (KLT)

$$\mathbf{G} = \left[\begin{array}{c} \text{Adaptive} \\ \text{eigenfunction} \\ \text{basis} \end{array} \right]$$

Pros

- Non-parametric (no params.)
- Transducer invariance easy
- Low complexity
- Adaptive KLT works well!

Encoding $\mathbf{\Gamma} = \mathbf{J}\mathbf{X} \approx \mathbf{G}^{-1}\mathbf{X}$

Decoding $\mathbf{Y} = \mathbf{G}\mathbf{\Gamma}$

Cosine transform

$$\mathbf{G} = \left[\begin{array}{c} \text{Array of} \\ \text{Cosine basis} \\ \text{functions} \end{array} \right]$$

Karhunen-Loeve (KLT)

$$\mathbf{G} = \left[\begin{array}{c} \text{Adaptive} \\ \text{eigenfunction} \\ \text{basis} \end{array} \right]$$

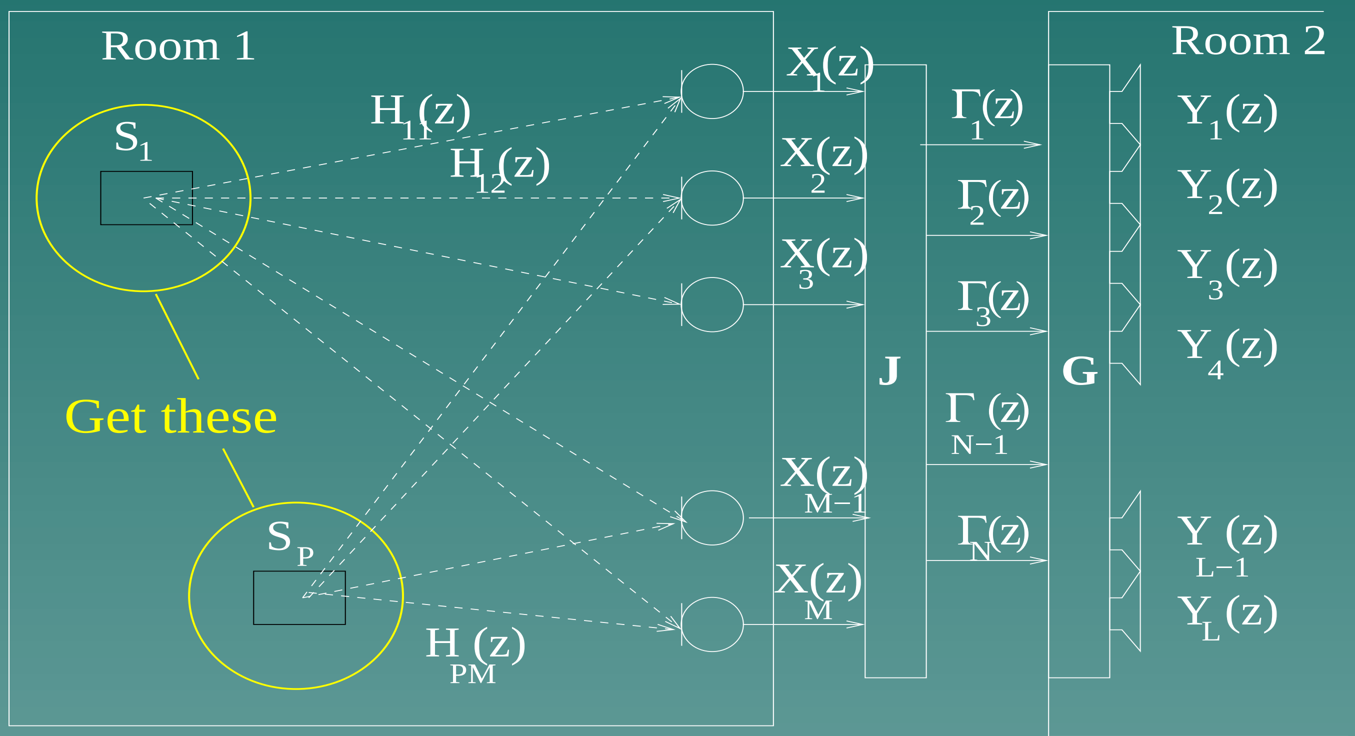
Pros

- Non-parametric (no params.)
- Transducer invariance easy
- Low complexity
- Adaptive KLT works well!

Cons

- Low coding gain
- Perceptual model?
- Blind to the problem
- Instantaneous mixture

2. Source separation approach



Signals:

$\mathbf{X} = \mathbf{HS}$

Encoder:

$\mathbf{\Gamma} = \mathbf{JHS}$

Decoder:

$\mathbf{Y} = \mathbf{G\Gamma}$

- $\mathbf{JH} = \mathbf{I} \Rightarrow \mathbf{\Gamma}$ are dry source signals!
- $\mathbf{J}$ is doing **source separation**
- $\mathbf{G}$ – Acoustics of the transmission room

Pros

- Total control over the problem
- High coding gain (?)
- Meaning of $\mathbf{G}$
- Additional functs.
- Transducer invariance
- Perceptual coding

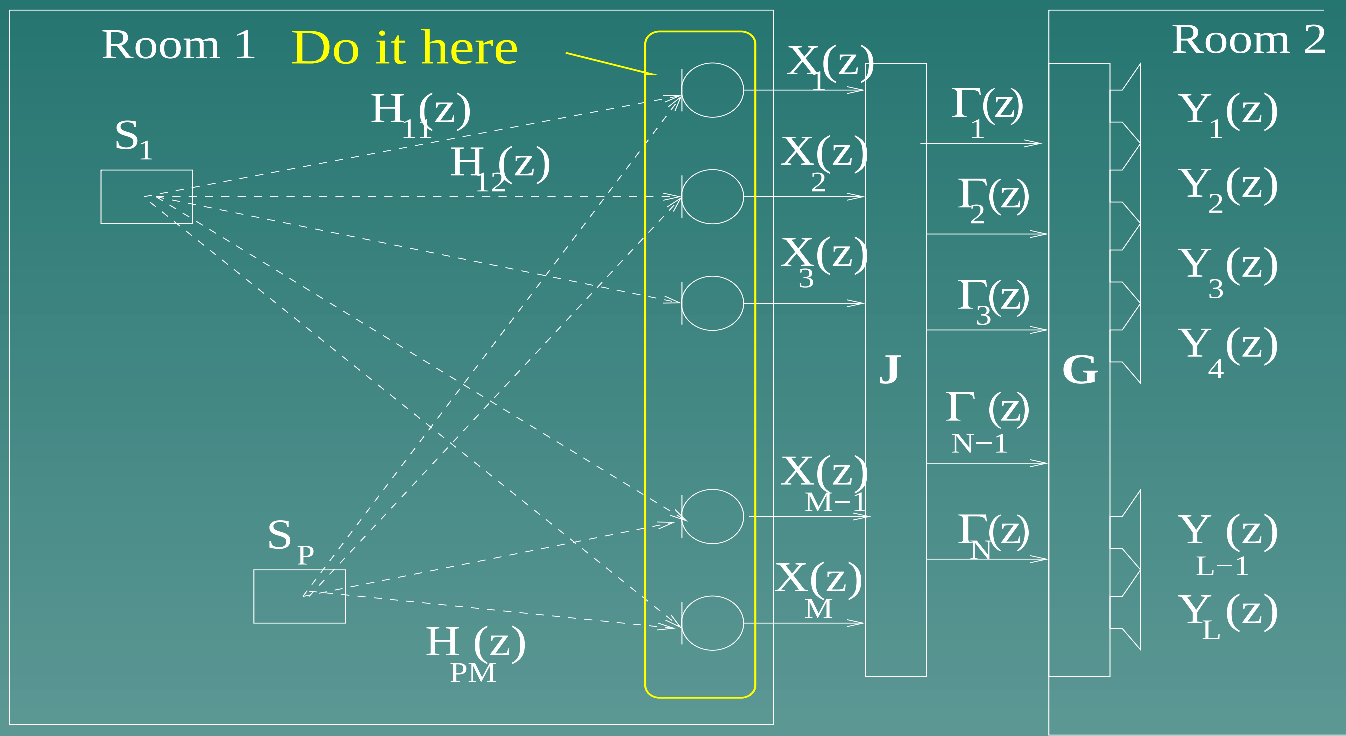
Pros

- Total control over the problem
- High coding gain (?)
- Meaning of $\mathbf{G}$
- Additional functs.
- Transducer invariance
- Perceptual coding

Cons

- Difficult (impossible at extreme)
- When it fails: coloration, artifacts
- Statistical Independence?
- $\mathbf{G}$ still hard to compress
- How many sources? What if $P > M$?
- If not perfect separation, what then?

3. Wavefield parametrization



Encoder:

$$\mathbf{\Gamma} = \mathbf{J}\mathbf{X}$$

Decoder:

$$\mathbf{Y} = \mathbf{G}\mathbf{\Gamma}$$

- Inter-channel prediction formulations
- $\mathbf{\Gamma}$ - clean microphone signals (and prediction errors)
- $\mathbf{G}$ - representation of wavefield

- In theory: one generating signal is enough!

Pros

- Simple formulations exist
- Clean generating signals
- Not making too many assumptions
- Not counting sources
- Perceptual coding

- In theory: one generating signal is enough!

Pros

- Simple formulations exist
- Clean generating signals
- Not making too many assumptions
- Not counting sources
- Perceptual coding

Cons

- Difficult to get high coding gain
- Non-uniqueness and stability problems for multiple Γ_n
- Not using data 'efficiently' (sources, acoustics)
- $\mathbf{G}$ is hard to compress
- Causality problems

Both approaches has many bad cons!

- Getting dry signals in source separation – difficult!
- Blind parametrization of $\mathbf{G}$ in wavefield coding – inefficient

Both approaches has many bad cons!

- Getting dry signals in source separation – difficult!
- Blind parametrization of $\mathbf{G}$ in wavefield coding – inefficient
- We need a **hybrid**:
  - ⋆ $\mathbf{\Gamma}$ – clean microphone signals
  - ⋆ $\mathbf{G}$ – parametrization based on source locations and acoustics.
  - ⋆ See article for a simple example: wavefield coder for a **single source in a free field**.

Subband formulations

NOTE: All techniques mentioned in the article can be applied separately at different subbands

- The order of problem decreases
- 'Perceptual' source separation in the (ERB/Bark) domain
- Omit irrelevant data (perceptual model)
- Scalar matrix processing makes more sense
- Tolerate spatial aliasing at high freqs.
- Localization different at low and high frequencies
- **But:** Increased algorithmic delay

Conclusions

- Acoustic opening: 70 yrs old idea
- Present hardware: standard audio stuff
- Digital audio coding: new problem, many open questions
- Generic formulation, alternative approaches
- We recommend: **hybrid subband processing**

Future work

- Prototypes being prepared
- Precedence effect (perceptual theory)
- Perceptual evaluation at large listening area (AES 21st)
- Full-duplex two-way system:
  - ⋆ Feedback and Acoustic Echo Canceling, AEC
  - ⋆ Design of a coder simplifies AEC
  - ⋆ Integrate a coder and AEC
- 1-D slit to a 2-D window
- Related activity: CARROUSO, USC (Immersive audiolab), BeingThere, etc.