Spatial Object Audio Coding (SOAC): How and Why it Works

Christof Faller

October 6, 2007

Audiovisual Communications Laboratory, EPFL Lausanne, Switzerland

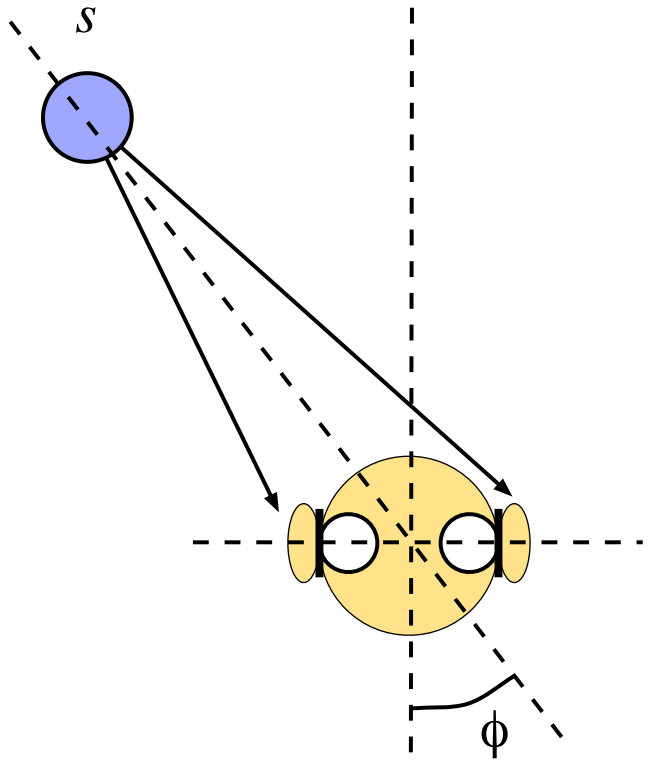
Joint Source Coding

Contents:

- Spatial Hearing Assumptions
- Spatial Object Audio Coding (SOAC)
- Conclusions

Spatial Hearing

Ear-input signals:

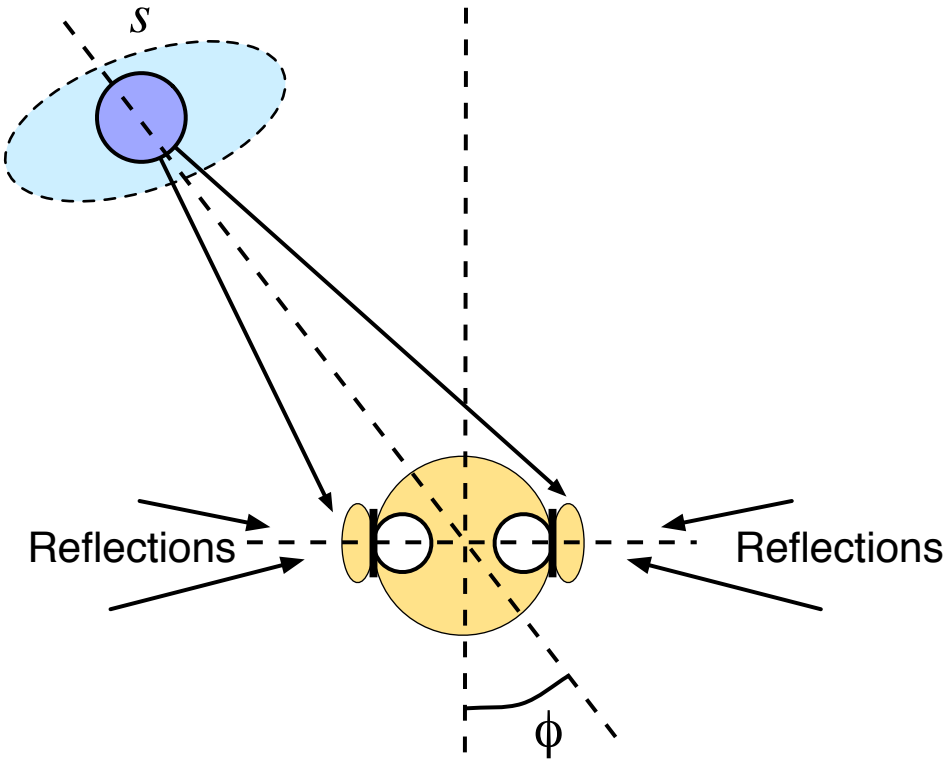


Distance difference:

→ Time difference (ITD)

Head shadowing:

→ Level difference (ILD)



Lateral reflections:

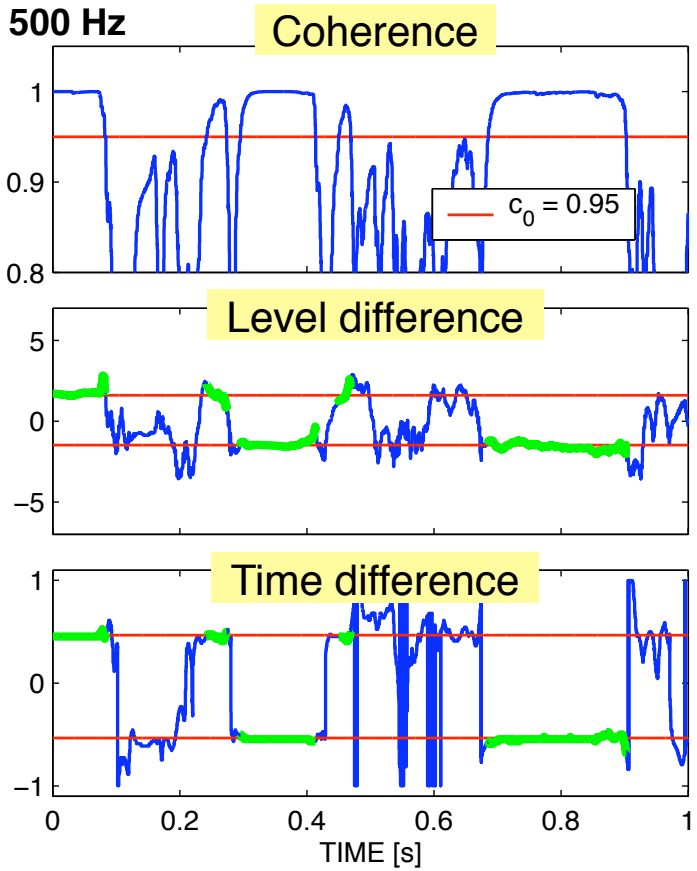
→ Coherence (IC)

Spatial Hearing Assumptions

Spatial Audio Coding:

Stereo Signal

Cues as a function of time and frequency determine the auditory spatial image



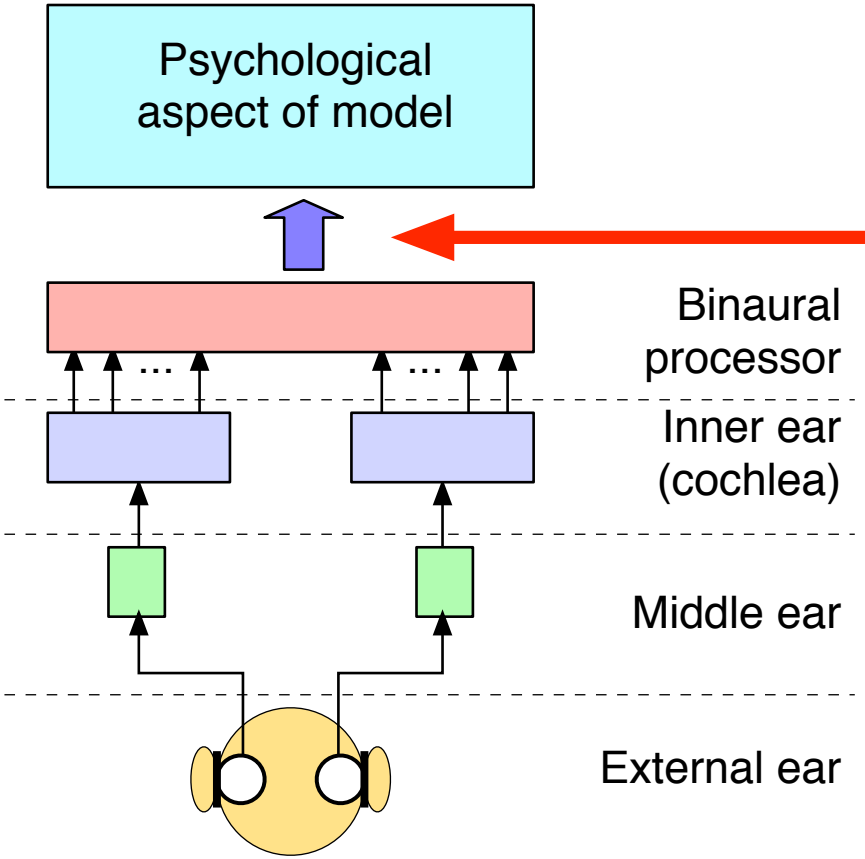
Mono Signal

Cue Synthesis

Perceptually Similar Stereo Signal

Spatial Hearing Assumptions

Perception of the horizontal auditory spatial image:



Assumption:
Horizontal auditory spatial image is largely determined by:
-time & level difference and coherence
-in critical bands

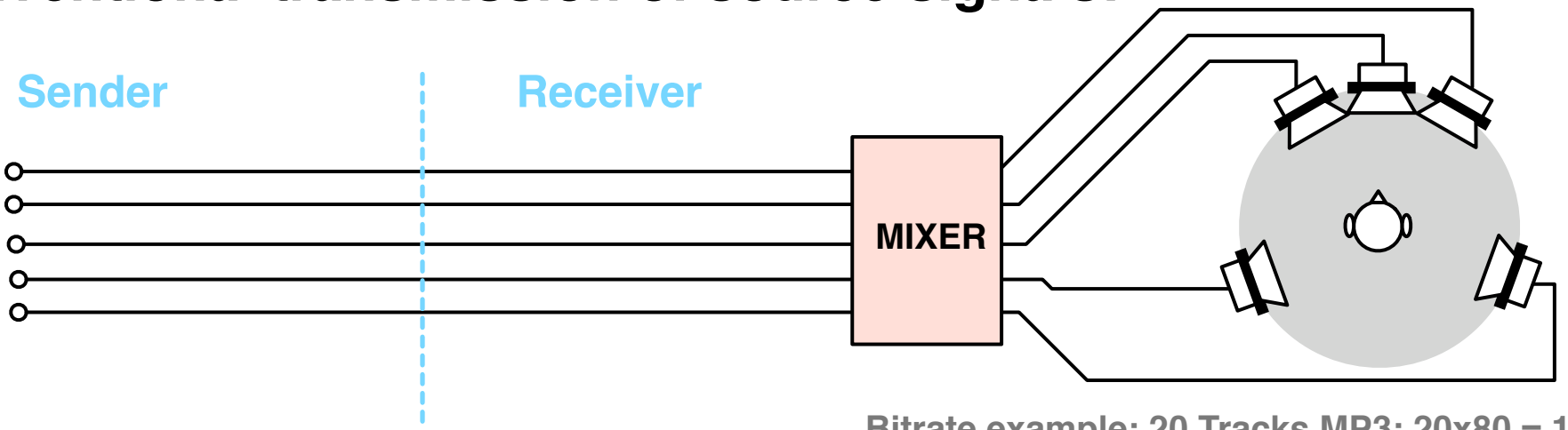
Spatial Object Audio Coding (SOAC)

Contents:

- Spatial Hearing Assumptions
- Spatial Object Audio Coding (SOAC)
- Conclusions

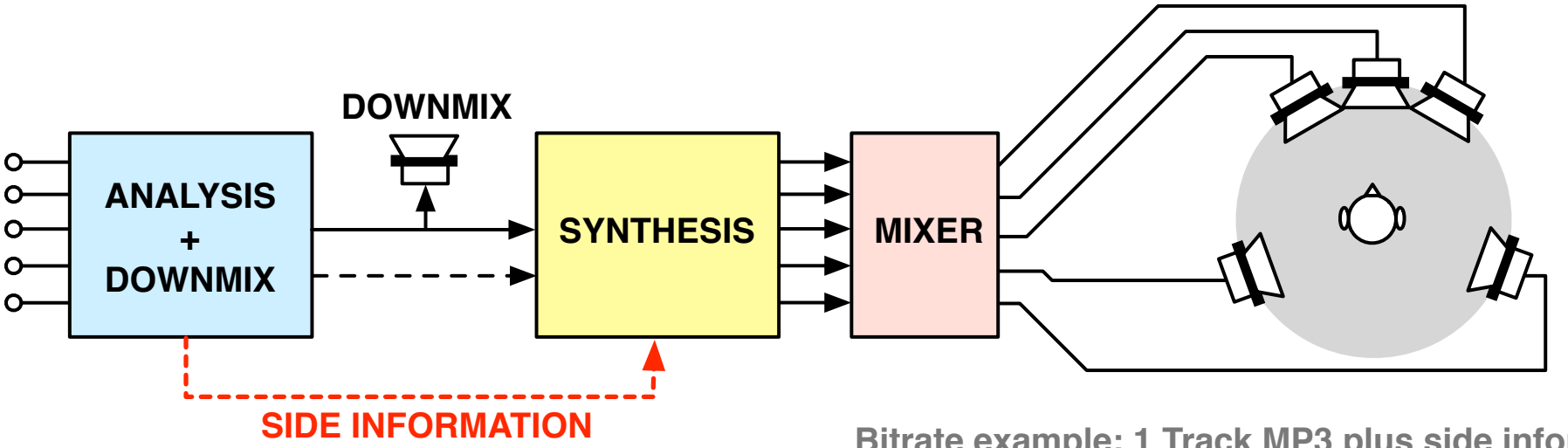
Spatial Object Audio Coding (SOAC)

Conventional transmission of source signals:



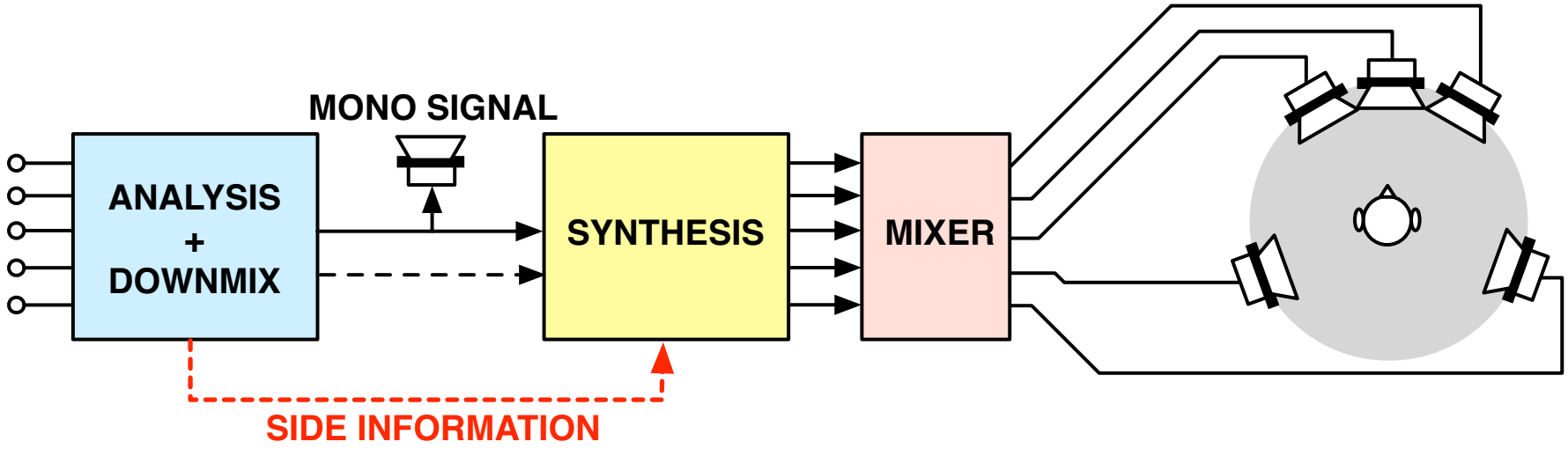
Bitrate example: 20 Tracks MP3: 20x80 = 1600 kbit/s

Proposed joint-source coding:



Bitrate example: 1 Track MP3 plus side information:
80 + 20x3 = 140 kbit/s

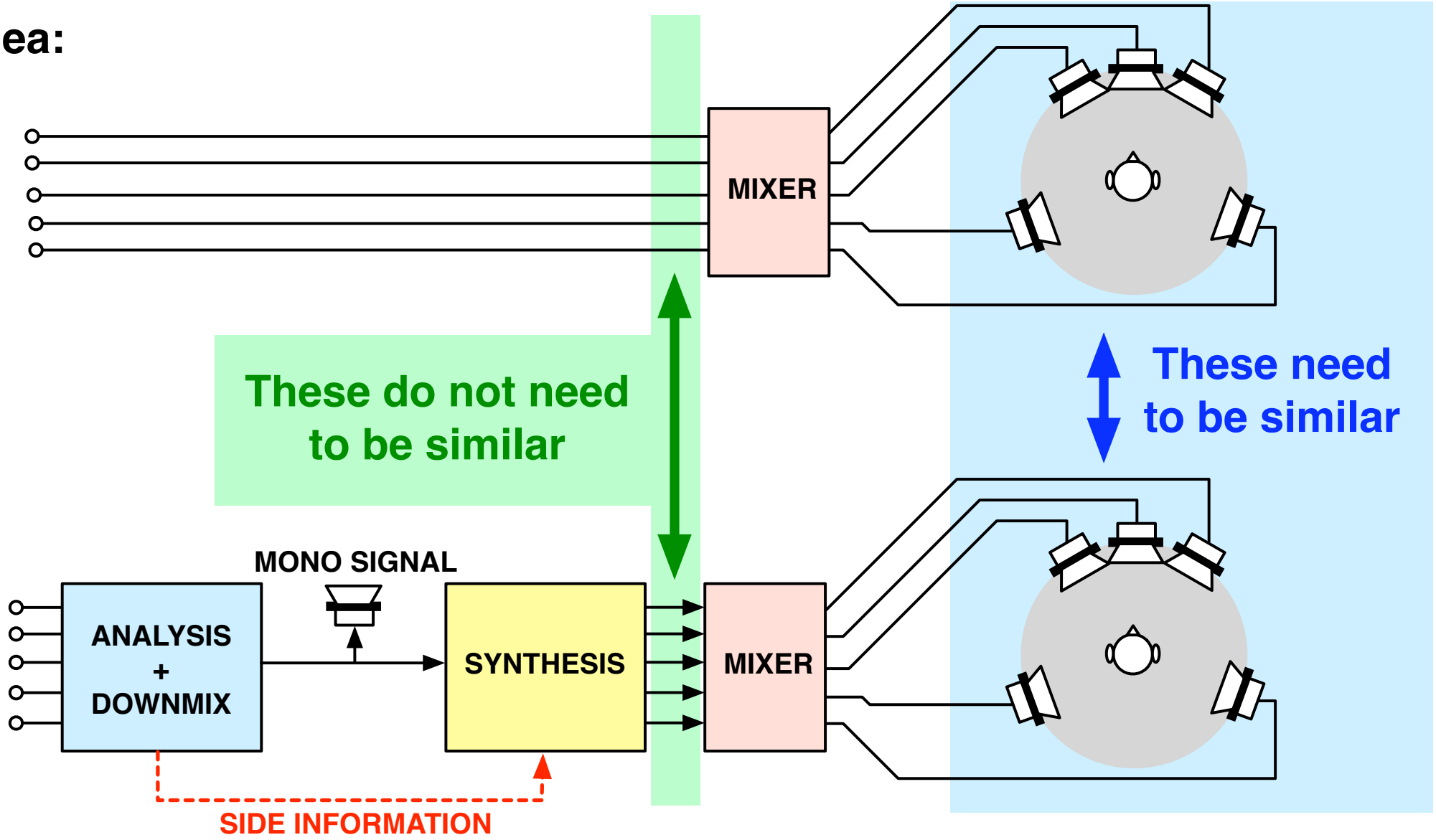
Spatial Object Audio Coding (SOAC)



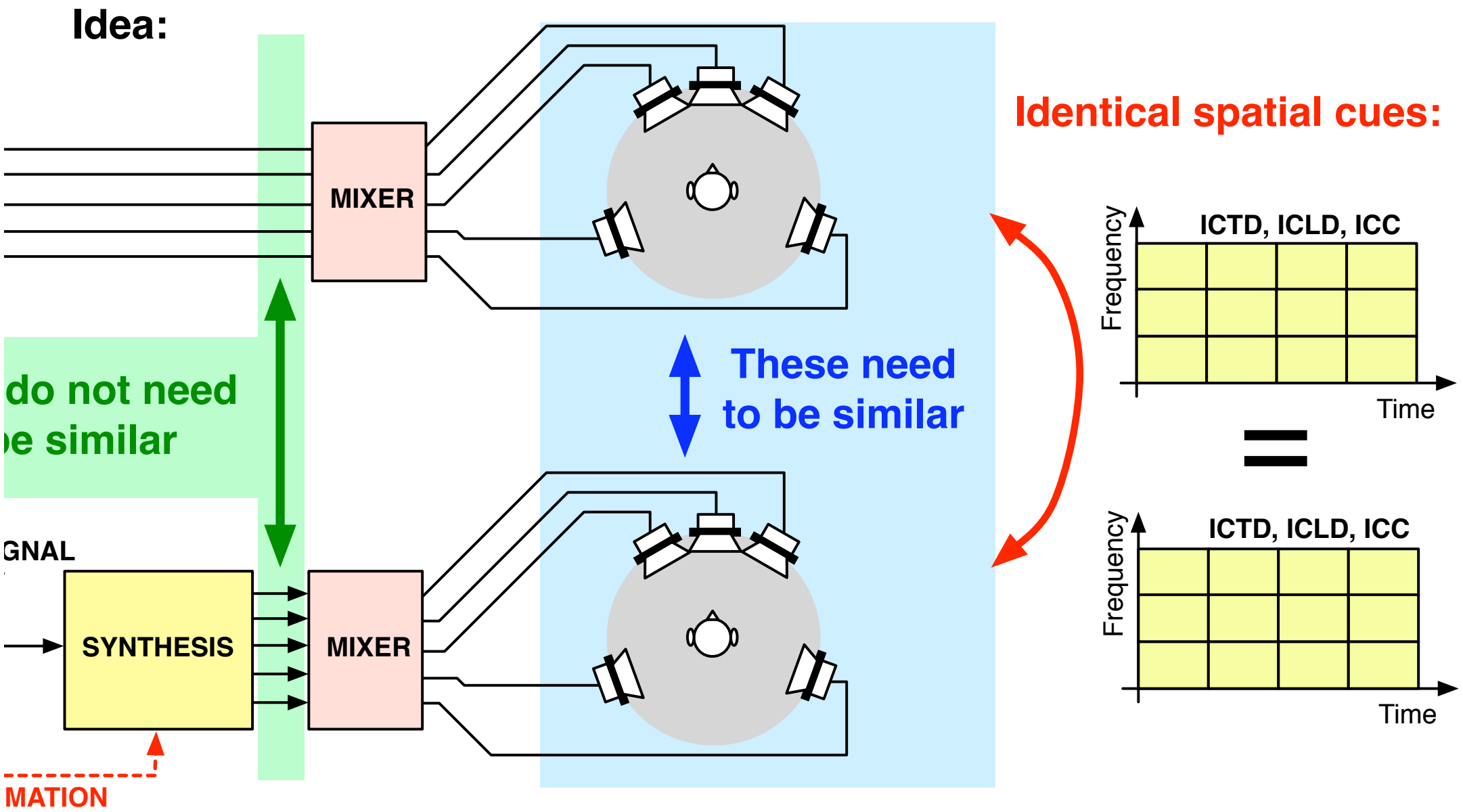
How possible?

Spatial Object Audio Coding (SOAC)

Idea:

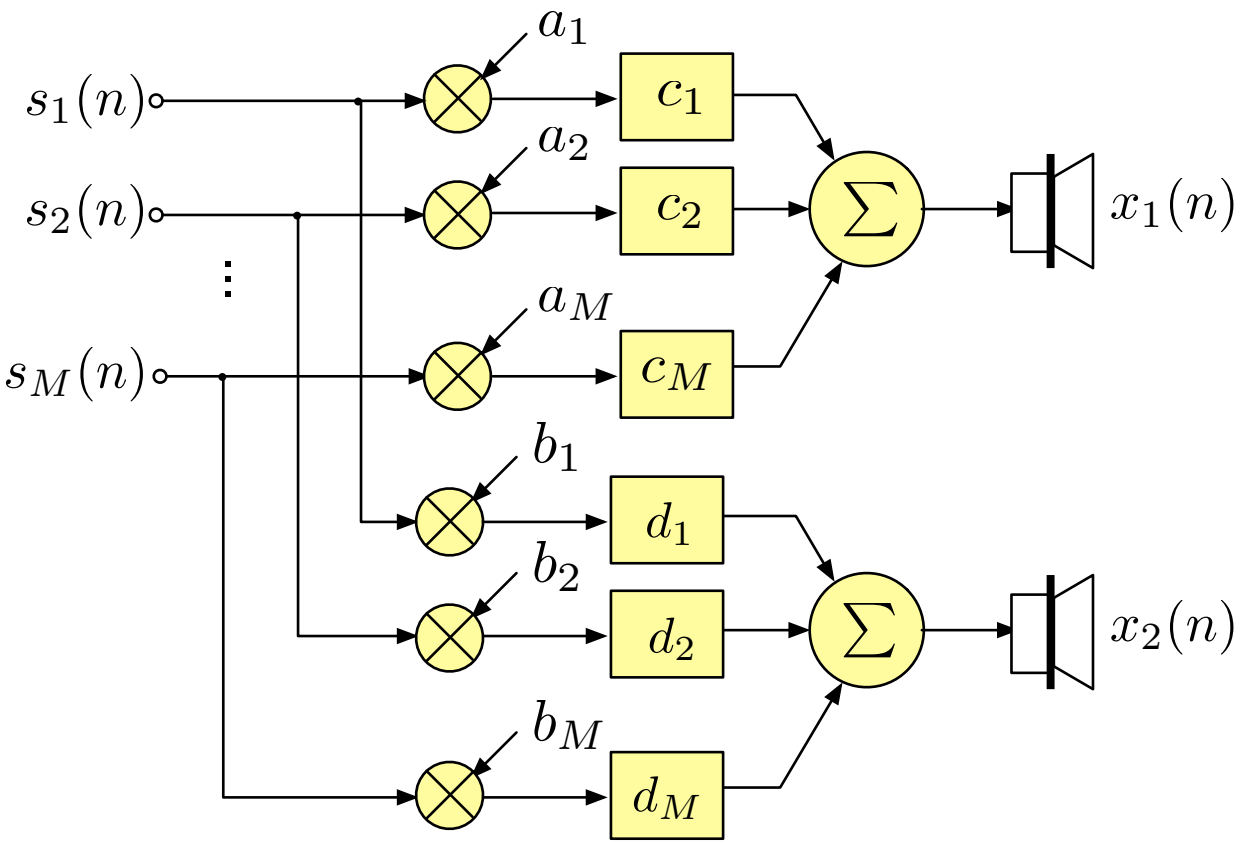


Spatial Object Audio Coding (SOAC)



Spatial Object Audio Coding (SOAC)

Stereo mixer:



Stereo signal:

$$x_1(n) = \sum_{i=1}^{M} a_i s_i(n - c_i)$$

$$x_2(n) = \sum_{i} b_i s_i(n - d_i)$$

Spatial Cues:

$$\text{ICLD} = 10 \log_{10} \frac{\sum_{i=1}^{M} b_i^2 \mathrm{E}\{\tilde{s}_i^2(n))\}}{\sum_{i=1}^{M} a_i^2 \mathrm{E}\{\tilde{s}_i^2(n))\}}$$

$$\text{ICTD} = \max_d |\Phi(n, d)|$$

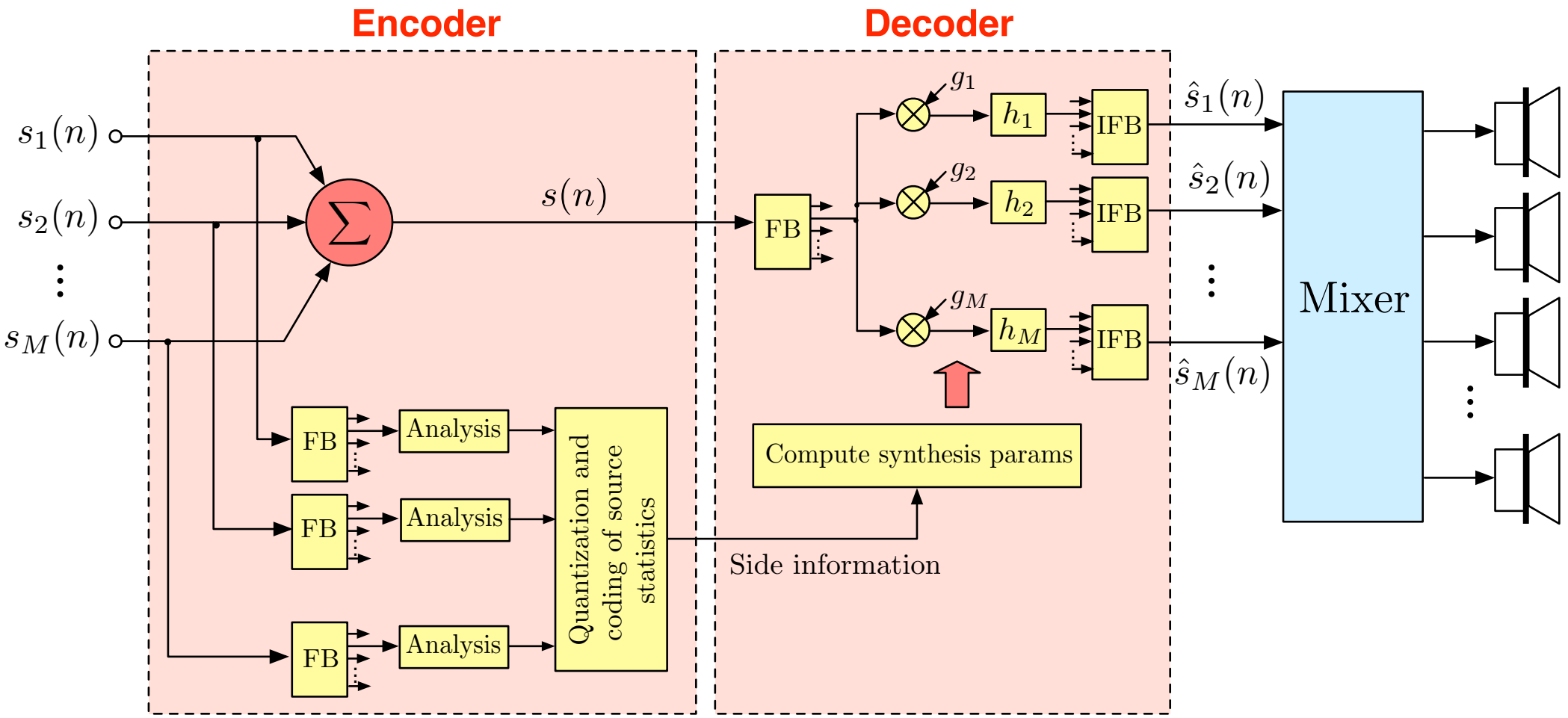
$$\text{ICC} = \arg\max_d \Phi(n, d)$$

Normalized cross-correlation function:

$$\Phi(n, d) = \frac{\sum_{i=1}^{M} a_i b_i \mathrm{E}\{\tilde{s}_i^2(n)\} \Phi_i(n, d - \tau_i)}{\sqrt{(\sum_{i=1}^{M} a_i^2 \mathrm{E}\{\tilde{s}_i^2(n)\})(\sum_{i=1}^{M} b_i^2 \mathrm{E}\{\tilde{s}_i^2(n)\})}} \quad \text{with} \quad \Phi_i(n, e) = \frac{\mathrm{E}\{s_i(n) s_i(n + e)\}}{\mathrm{E}\{s_i^2(n)\}}$$

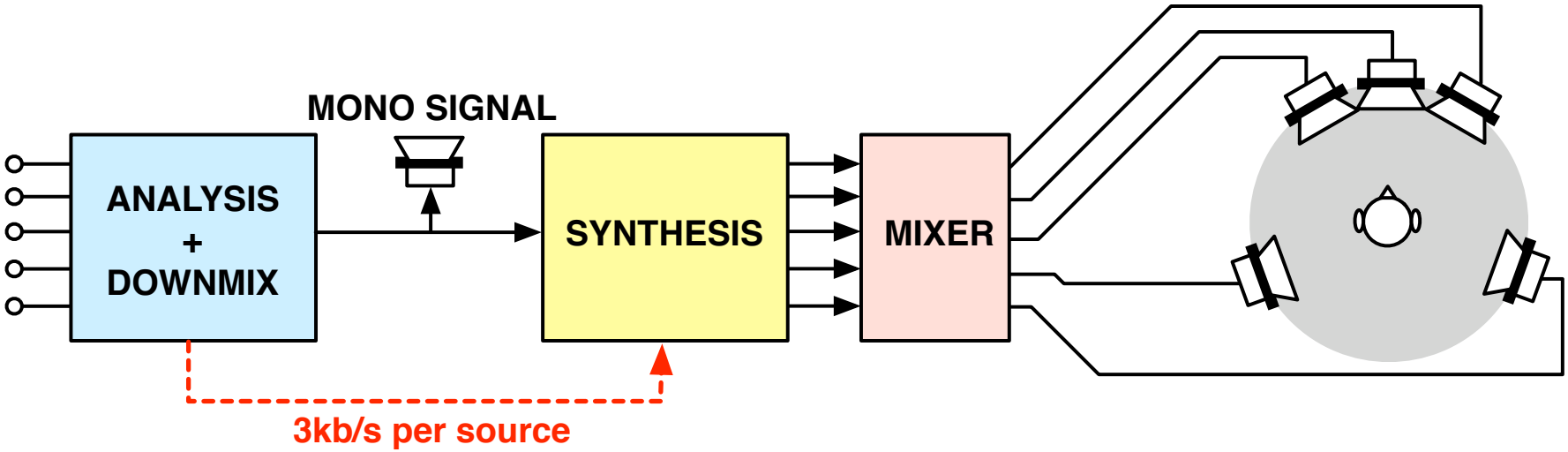
Spatial Object Audio Coding (SOAC)

Detailed encoding/decoding:



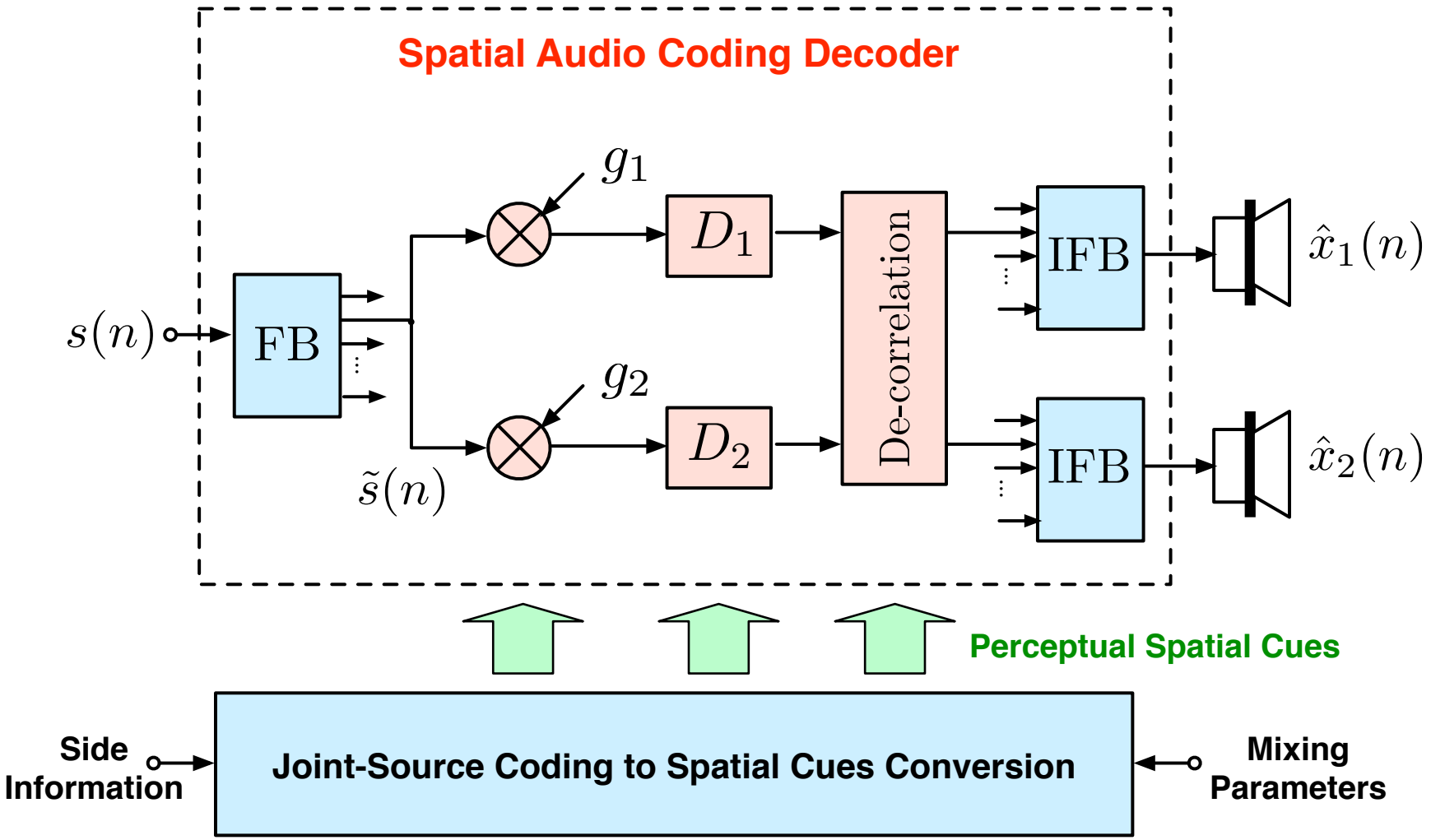
Spatial Object Audio Coding (SOAC)

Demo:



Spatial Object Audio Coding (SOAC)

Direct Mixing without decoding of the N sources:



Spatial Object Audio Coding (SOAC)

Contents:

- Spatial Hearing
- Joint Source Coding
- Conclusions

Spatial Object Audio Coding (SOAC)

Conclusions:

***Joint-Coding* of *Independent Signals* by means of a very simple parametrization of the sources.**

An MEPG Surround decoder can be used for jointly decoding/ mixing the sources.