Distributed Spatial Audio Coding in Wireless Hearing Aids

Olivier Roy

LCAV - I&C - EPFL

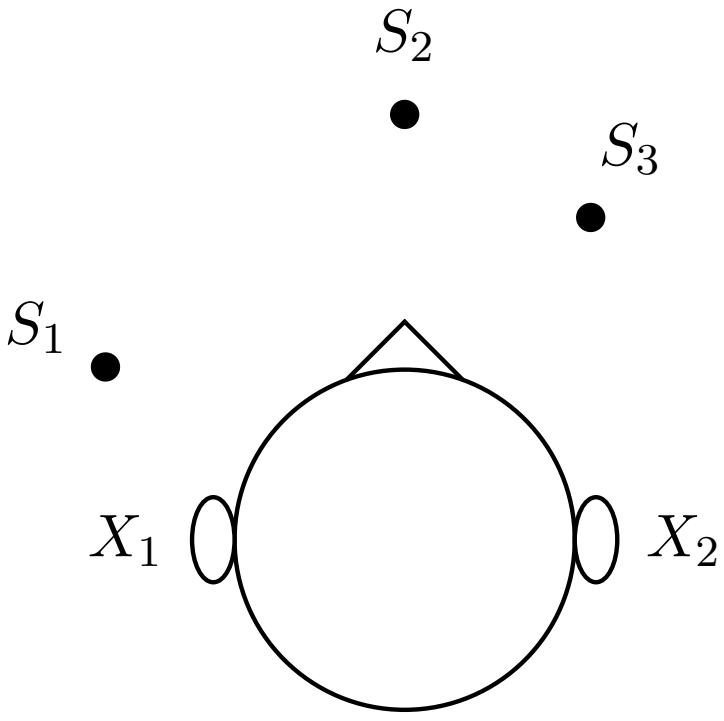
Joint work with Martin Vetterli

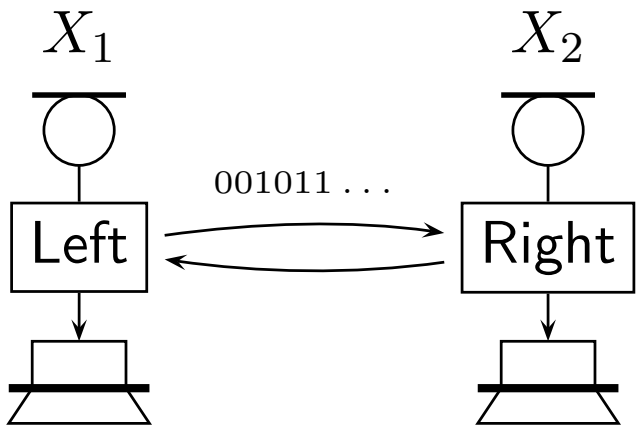
WASPAA, New Paltz, NY - October 23, 2007

Outline

Setup

- Binaural hearing aids with wireless link





Motivations

Computation of binaural cues

- Scene analysis
  - Classification
  - Source localization
  - Voice activity detection
- Parametric spatial audio coding
  - Binaural Cue Coding (centralized)
  - Acoustic sensor networks, binaural hearing aids (distributed)

Motivations

Ultimate goals

Distributed computation of binaural cues

Distributed parametric spatial audio coding

Motivations

Ultimate goals

Distributed computation of binaural cues

Distributed parametric spatial audio coding

In this presentation

Some preliminary results

Binaural Cues - Generalities

- Time-frequency representation, one value per critical band $\mathcal{B}_l$
- Inter-channel level difference (ICLD)

$$\Delta p[l] = p_1[l] - p_2[l]$$

where

$$p_m[l] = 10 \log_{10} \left(\frac{1}{|\mathcal{B}_l|} \sum_{k \in \mathcal{B}_l} |X_m[k]|^2 \right) \quad \text{for } m = 1, 2$$

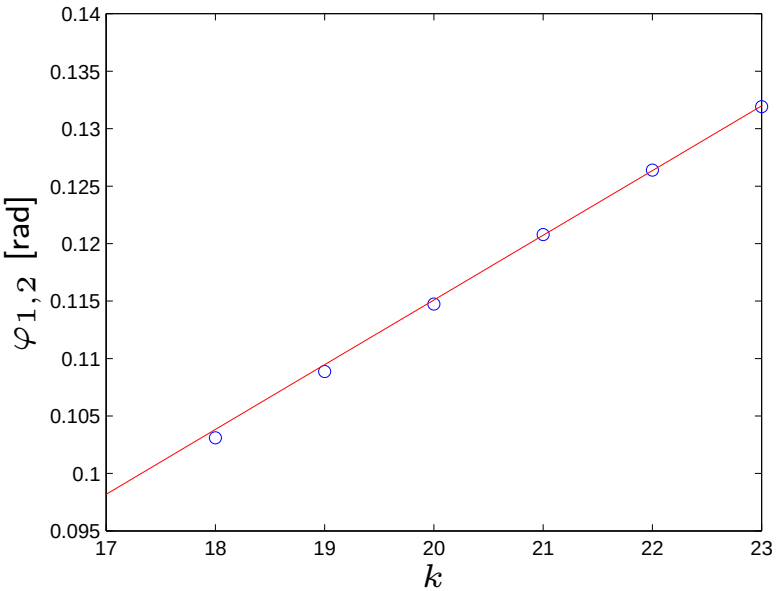
Binaural Cues - Generalities

- Inter-channel time difference (ICTD)
  - Phase difference

$$\varphi_{1,2}[k] = \arg X_1[k] X_2^*[k]$$

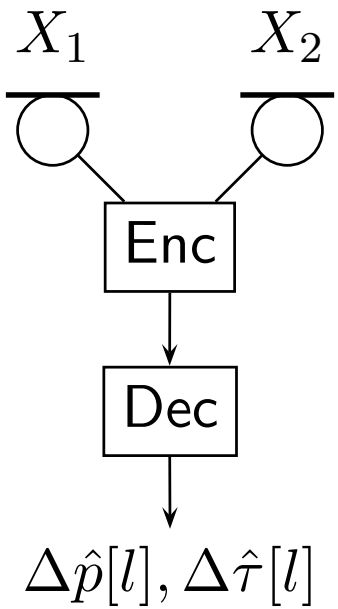
  - Mean-square fitting

$$\Delta\tau[l] = \frac{N}{2\pi} \frac{\sum_{k\in\mathcal{B}_l} k\, \varphi_{1,2}[k]}{\sum_{k\in\mathcal{B}_l} k^2}$$



Binaural Cues - Centralized Case

- Centralized coding



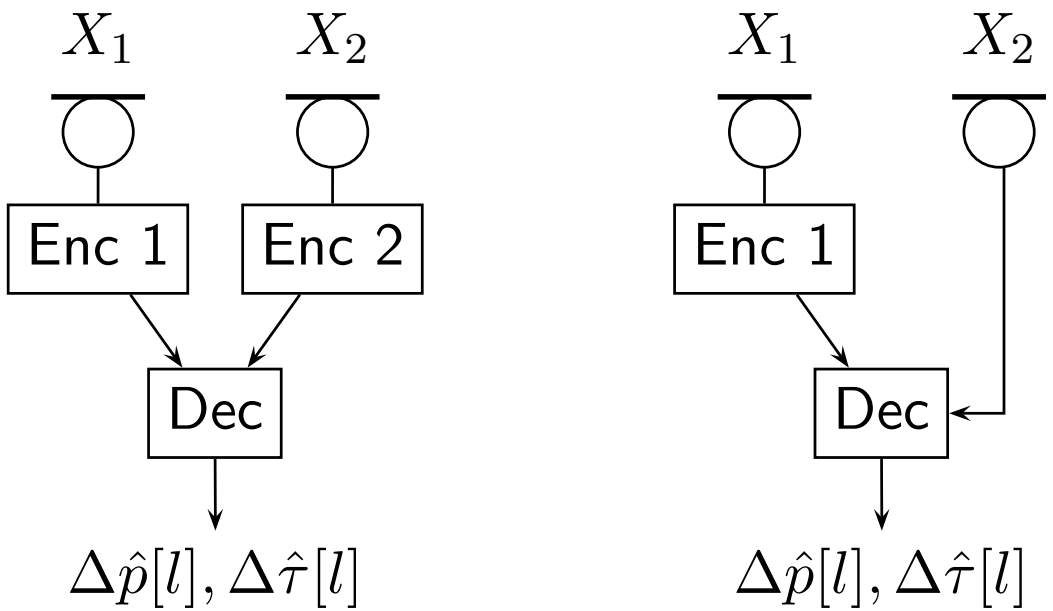
  - Both X_1 and X_2 available
  - Critical band averaging before transmission
  - Spatial correlation taken into account for coding
    e.g. ICLD

$$\Delta p[l] \in \left[\Delta p_{min}[l] \, , \Delta p_{max}[l] \right]$$

    $\Longrightarrow$ scalar quantizer with range $\Delta p_{max}[l] - \Delta p_{min}[l]$

Binaural Cues - Distributed Case

- Distributed coding



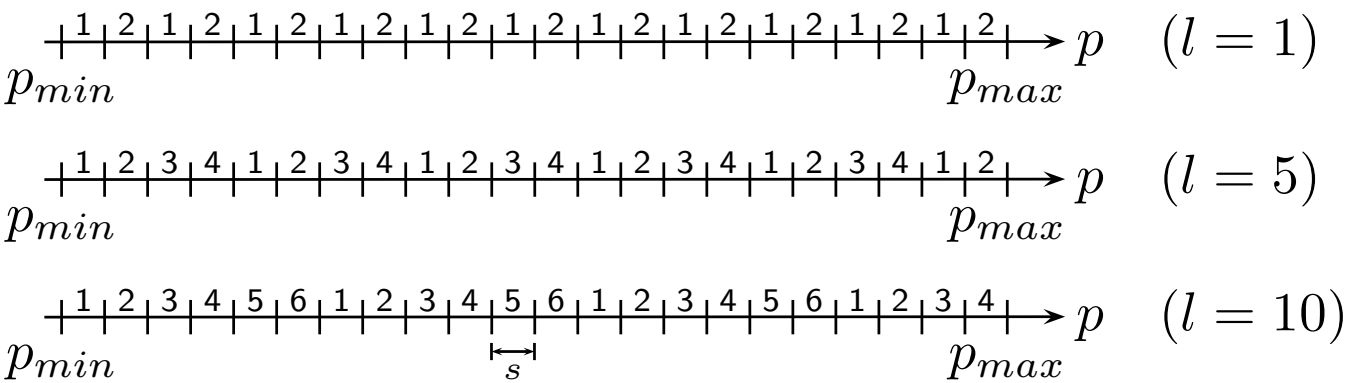
- X_1 and X_2 not anymore available together
- Critical band averaging? Spatial correlation?

Binaural Cues - Distributed Case

- ICLD
  - Scalar quantization of $p_1[n,l]$ and $p_2[n,l]$

$$\begin{aligned} i_1[n,l] - i_2[n,l] &\in \left\{\Delta i_{min}[l], \ldots, \Delta i_{max}[l]\right\} \\ &= \left\{\left\lfloor \frac{\Delta p_{min}[l]}{s} \right\rfloor, \ldots, \left\lceil \frac{\Delta p_{max}[l]}{s} \right\rceil\right\} \end{aligned}$$

  - Modulo coding approach = index reuse

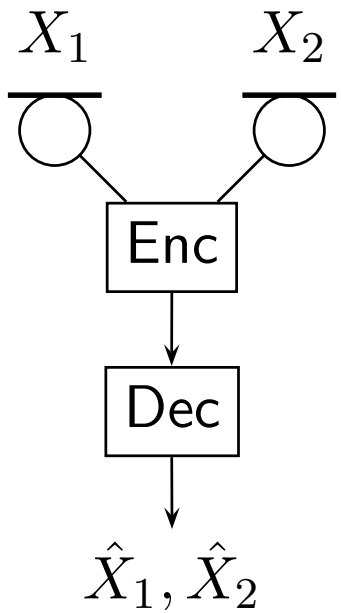


Binaural Cues - Distributed Case

- ICLD (cont'd)
  - Same coding efficiency as centralized case
  - Takes head shadowing into account
  - Assumption must be verified!!
- ICTD
  - Difficult
  - HRTF lookup table

Parametric Spatial Audio Coding - Centralized Case

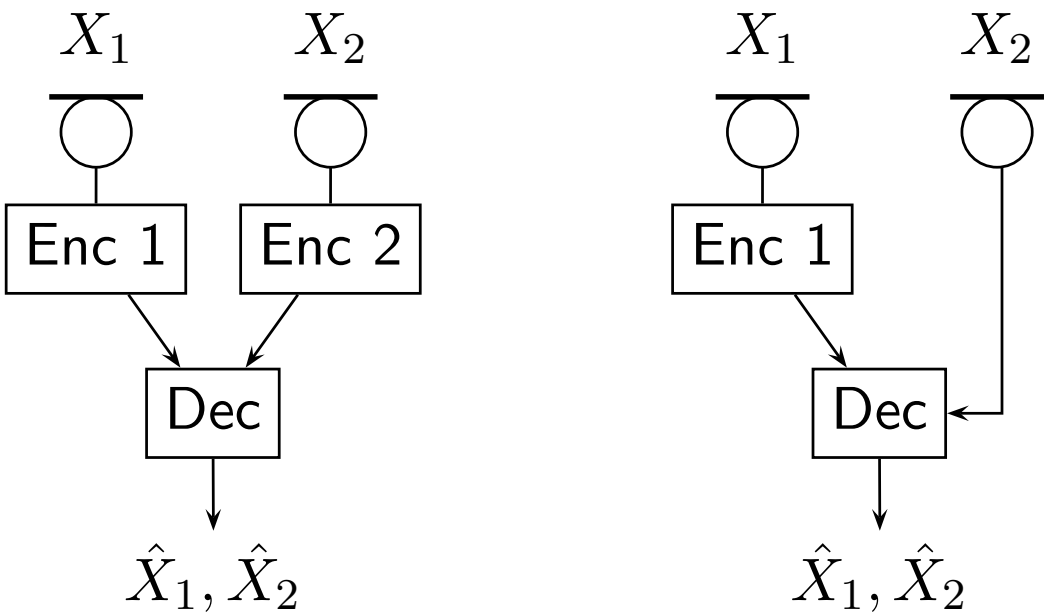
- Centralized coding



  - Binaural Cue Coding [Baumgarte and Faller '03]: downmixed signal + inter-channel cues
  - Multi-channel audio recovered by imposing cues on mono signal

Parametric Spatial Audio Coding - Distributed Case

- Distributed coding



  - Binaural cues computed as explained previously
  - Binaural signal recovered by imposing cues on available signal

Parametric Spatial Audio Coding - Results

- Results (f_s = 20.48 kHz, R = 8 kb/s)
  - Anechoic
    - Sources at 0°and 15°: original & reconstruction
    - Sources at -30°, 0°, 15°: original & reconstruction
  - Reverberant
    - Sources at 0°and 30°, RT 120 ms: original & reconstruction
    - Sources at 0°and 30°, RT 600 ms: original & reconstruction
- Works decently for simple scenarios (no reverberation)
- True ICTDs needed for more realistic scenarios

Thanks for Your Attention

Questions?