Parametric Coding of Spatial Audio

Ph.D. Thesis

Christof Faller, September 24, 2004

Thesis advisor: Prof. Martin Vetterli
Parametric Coding of Spatial Audio

Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Audio Coding

Audio coding

Convert audio signal into a representation suitable for:
- Transmission
- Storage

Minimize bitrate
- Optimal for storage
- Optimal for transmission
Lossless coding

50% Memory/bitrate reduction
(Put music of 10 CDs on 1 CD)

Redundancy reduction:
- Time to frequency transform
- Entropy coding
Perceptual audio coding

\[ x(n) \rightarrow \text{Encoder} \rightarrow \text{Decoder} \rightarrow \hat{x}(n) \]

90% Memory/bitrate reduction (Put music of 10 CDs on 1 CD)
Audio Coding

Parametric audio coding

$x(n)$ → Encoder → Decoder → $\hat{x}(n)$

>90% Memory/bitrate reduction

Drawback: Quality loss
Receiver properties...

...and source properties...

...have been extensively considered in audio coding for monaural audio signals.

This thesis: Extensively consider receiver and source properties for stereo and multi-channel audio signals.
Contents:

- Audio Coding and Thesis Motivation
- **Background**
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Spatial audio playback

Two-channel stereo
(headphone and loudspeaker playback)

5.1 Surround

Spatial audio playback:
- perception of an auditory spatial image
Interaural differences

One source, free-field

Source azimuth and ITD/ILD:
(Narrowband signal)

→ distance difference, head shadowing
→ interaural time and level difference
(ITD and ILD)
Background

**Inter-channel differences:**
- Inter-channel time-difference (ICTD)
- Inter-channel level-difference (ICLD)
- Inter-channel coherence (ICC)
Mixing stereo signals:
Background

Other auditory spatial image attributes:

Auditory event distance
- Power of ear-input signals
- Ratio of power of direct to reflected sound

Auditory event width
- Lateral fraction

Listener envelopment
- Late lateral energy fraction

Spatial audio playback:
These attributes are controlled by adding reflections to the signal channels.
Contents:

- Audio Coding and Thesis Motivation
- Background
- **Binaural Cue Coding (BCC)**
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Binaural Cue Coding (BCC)

Conventional multi-channel transmission

Proposed multi-channel transmission

MONO SIGNAL

ANALYSIS + DOWNMIX

SYNTHESIS
Binaural Cue Coding (BCC)

Estimation of inter-channel cues: ICTD, ICLD, and ICC

**ICTD/ICLD:**
1-3 kb/s (per channel pair)

**ICC:**
1-3 kb/s
Binaural Cue Coding (BCC)

ICTD/ICLD/ICC synthesis

\[ s(n) \rightarrow FB \rightarrow \tilde{s}(k) \]

\[ \vdots \]

\[ d_i(k) \rightarrow a_i(k) \rightarrow h_i(k) \rightarrow \hat{x}_i(k) \rightarrow IFB \rightarrow \hat{x}_i(n) \]

\[ d_C(k) \rightarrow a_C(k) \rightarrow h_C(k) \rightarrow \hat{x}_C(k) \rightarrow IFB \rightarrow \hat{x}_C(n) \]

\[ d_i : \text{delays} \]
\[ a_i : \text{scale factors} \]
\[ h_i : \text{filters} \]
Binaural Cue Coding (BCC)

ICTD, ICLD, and ICC and auditory spatial image attributes

Auditory event localization:
ICTD, ICLD

Effect of late reflections:
(ambience, listener envelopment, auditory event distance)
ICC (de-correlation)
ICLD (reverberation decay)

Effect of early reflections:
(auditory event width, coloration)
ICC (de-correlation)
ICLD (comb filter)
Subjective evaluation: Stereo audio quality

MUSHRA (ITU-R BS.1534), headphone listening, 7 experienced subjects

BCC: "excellent"
Binaural Cue Coding (BCC)

Surround 5.1 Demo:

Original → ANALYSIS + DOWNMIX → Mono → SYNTHESIS → Synthesized

15 kb/s
Binaural Cue Coding (BCC)

Surround 5.1 Demo:

44 kb/s
(lowest bitrate ever demonstrated)

Memory of 1 CD = 32 CDs worth of music!
Not stereo, but surround!

Original

Synthesized
Parametric Coding of Spatial Audio

Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Variations of BCC

"MP3 Surround": Stereo backwards compatible coding of 5.1

Stereo-downmix:

\[ \begin{align*}
Lt &= L + 0.7C + sL \\
Rt &= R + 0.7C + sR
\end{align*} \]

2-to-5 Synthesis:
Variations of BCC

Low complexity **5-to-2 BCC**: Subjective evaluation:
MUSHRA (ITU-R BS.1534)

Hidden reference, Anchor, AAC, 5-to-2 BCC + MP3, Dolby Prologic II

- 256 kb/s
- 192 kb/s
- PCM
Variations of BCC

MP3 Surround Demo:

Lt, Rt, L, C, R, LFE, Ls, Rs
Variations of BCC

Conventional flexible rendering

BCC for flexible rendering

Analysis + Downmix

Mono Signal
Variations of BCC

BCC for Flexible Rendering

Side information:
Time-frequency structure of sum signal
Binaural Cue Coding (BCC)

BCC for Flexible Rendering Demo:

4 Instruments playing together.
Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Model for source localization in complex listening scenarios:
- concurrently active sources
- direct and reflected sound
Source Localization in Complex Listening Scenarios

IC and concurrent sources

\[ e_1(n) = a_{11}s_1(n - d_{11}) + a_{21}s_2(n - d_{21}) \]

\[ e_2(n) = a_{12}s_1(n - d_{12}) + a_{22}s_2(n - d_{22}) \]

\[ \text{IC} = \frac{\max\{a_{11}a_{12}p_1, a_{21}a_{22}p_2\}}{\sqrt{(a_{11}^2p_1 + a_{21}^2p_2)(a_{12}^2p_1 + a_{22}^2p_2)}} \]

IC can be used as a "single-source-active indicator"
IC and reflections

IC can be used as a "first-wavefront indicator"
Cue selection: Use ITD and ILD when IC(n) > c₀
Source Localization in Complex Listening Scenarios

Two concurrent male speech sources, free-field

![Diagram showing two concurrent male speech sources, free-field](image)

- **IC**
- **ILD**
- **ITD**

Parameters:
- $c_0 = 0.95$
- Frequency: 500Hz
- Time: 0 to 1 second
Source Localization in Complex Listening Scenarios

Precedence effect:
(Broadband pulses, 5ms lead/lag delay, free-field)
Source localization in reverberant environment:
(2 male speech sources, 500Hz: RT = 2s, 2kHz: RT = 1.4s)
## Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions
Conclusions

Binaural Cue Coding (BCC):
- low bitrate coding of stereo and multi-channel audio signals
- low bitrate transmission of independent sources
- bridging between different audio formats

Proposed source localization model:
- speculates about role of IC for "cue selection"
- attempts at explaining source localization in complex listening
Future work:

Multi-channel BCC parameters: Different more efficient and flexible parametrization.

BCC applied to binaural recordings: Further investigate.

Psychoacoustic experiments with BCC stimuli.

Source localization model:
- Adaptation of cue selection threshold
- More simulations, psychoacoustic experiments
- Can model be related to other attributes of auditory spatial image?
Acknowledgements:

Martin Vetterli, Peter Kroon, and all colleagues I collaborated with.

Thanks to all of you for coming to this defense!
Hybrid audio coding:
Intensity stereo coding (ISC):

Spectral bandwidth replication (SBR):
Binaural Cue Coding (BCC)

Downmix with equalization

Numerical example:
Binaural Cue Coding (BCC)

Alternative ICTD/ICLD/ICC synthesis

\[ a_i, b_i : \text{scale factors} \]

LR: late reverberation filter

Coherent/incoherent channel power:
Binaural Cue Coding (BCC)

Subjective evaluation: 5-channel audio quality

Hidden reference test (ITU-R BS.1116), loudspeaker listening, 9 subjects
Binaural Cue Coding (BCC)

Subjective evaluation: ICLD time resolution

Auditory spatial image stability

Overall quality

imperceptible
perceptible but not annoying
slightly annoying
annoying
very annoying
Variations of BCC

C-to-E BCC

- Potentially better quality than C-to-1 BCC
- E-channel backwards compatible coding of C-channel audio
Variations of BCC

6-to-5 BCC: 5.1 backwards compatible coding of 6.1

Downmix:

5-to-6 Synthesis:

Dolby Surround EX loudspeaker positioning
Variations of BCC

7-to-5 BCC: 5.1 backwards compatible coding of 7.1

Downmix:

5-to-7 Synthesis:
Variations of BCC

Hybrid BCC

Differences to intensity stereo coding:
- consider more cues (not only intensities)
- use separate filterbank (better performance)
Variations of BCC

Hybrid BCC

anchor 1  anchor 2  $f_0 = 0$ kHz  $f_0 = 1$ kHz  $f_0 = 2$ kHz  $f_0 = 6$ kHz  $f_0 = 16$ kHz  AVERAGE

0 20 40 60 80 100

GRADING

59 kb/s  68 kb/s  74 kb/s  84 kb/s  101 kb/s

Hybrid BCC

\[ x_1(n) \rightarrow FB \rightarrow BCC ENCODER \rightarrow IFB \rightarrow FB \rightarrow BCC DECODER \rightarrow \hat{x}_1(n) \]
\[ x_2(n) \rightarrow FB \rightarrow \vdots \rightarrow IFB \rightarrow FB \rightarrow \vdots \rightarrow \hat{x}_2(n) \]
\[ x_C(n) \rightarrow FB \rightarrow \vdots \rightarrow IFB \rightarrow FB \rightarrow \vdots \rightarrow \hat{x}_C(n) \]

SIDE INFORMATION
Source Localization in Complex Listening Scenarios

Three and five concurrent male speech sources:
(-30°, 0°, 30°, -80°, 80°, free-field)

Without cue selection

With cue selection
Source Localization in Complex Listening Scenarios

**Cue selection:** Three phases of the precedence effect, 500Hz critical band

**Without cue selection:**

- ITD
- ILD

**With cue selection:**

- ITD
- ILD

lead/lag delay 200 ms

$p_0, c_0$ lead/lag delay [ms]
Source Localization in Complex Listening Scenarios

Amplitude panning, free-field, 500Hz critical band

Standard stereo setup
Two male speech sources
+-8dB amplitude panning.