Aalto University

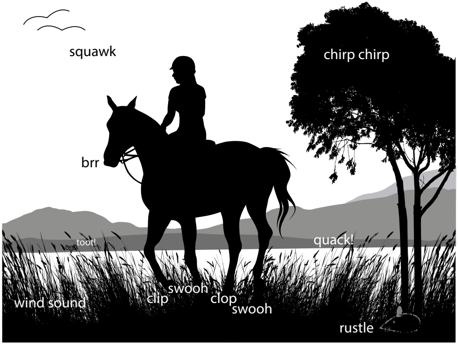
Parametric Time-Frequency-Domain Spatial Audio – Delivering Sound According to Human Spatial Resolution

Ville Pulkki

Acoustics lab
Department of Signal Processing and Acoustics
School of Electrical Engineering
Aalto University, Helsinki, Finland

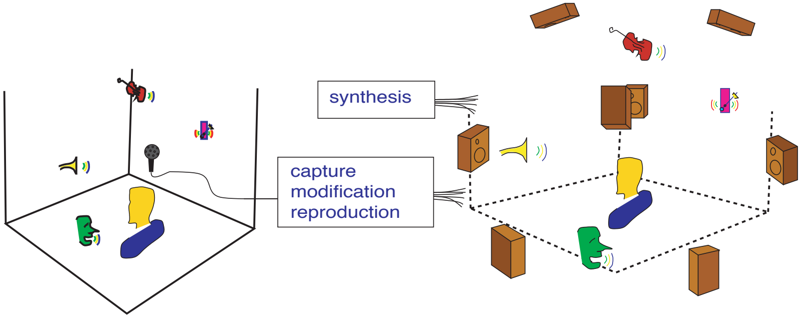
16-Oct-2017

Spatial sound

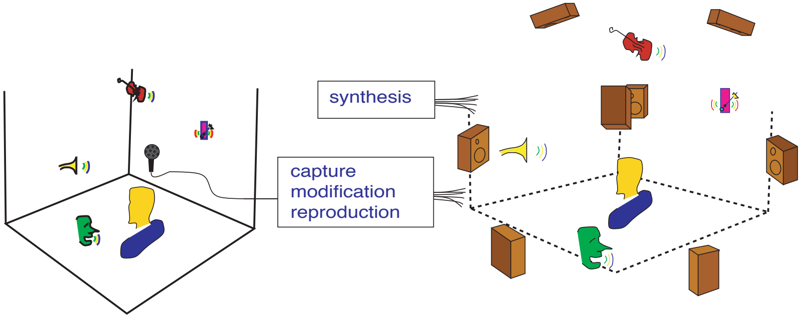


- Spatial: Where is it? / How far is it?
- Spectral: What is it?

Reproduction of spatial sound

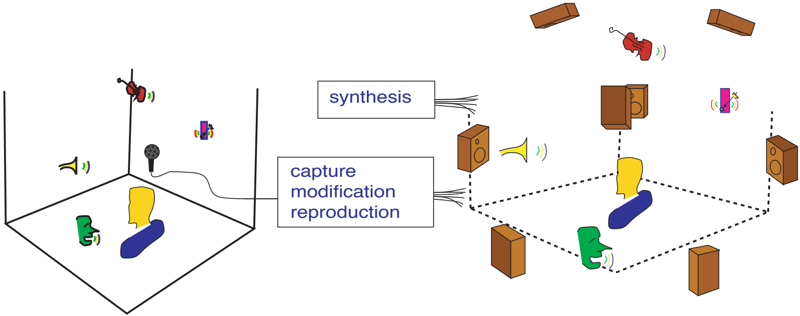


Reproduction of spatial sound



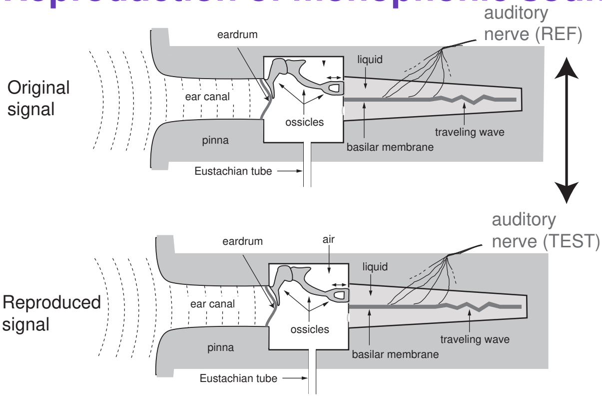
- Relay the perception

Reproduction of spatial sound



- Relay the perception
- Synthesize a desired perception

Reproduction of monophonic sound



- Auditory nerve outputs should match btw original and reproduced sound

Perceptual audio coding

- Reproduced signal does not have to be PCM samples of $x(t)$

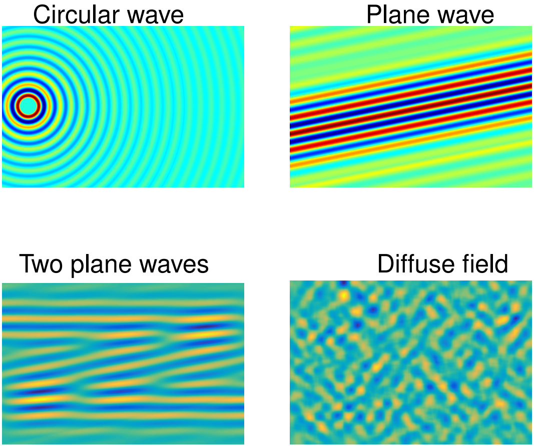
Perceptual audio coding

- Reproduced signal does not have to be PCM samples of $x(t)$
- In perceptual approaches the signal is a frequency-band representation with various masking effects taken into account

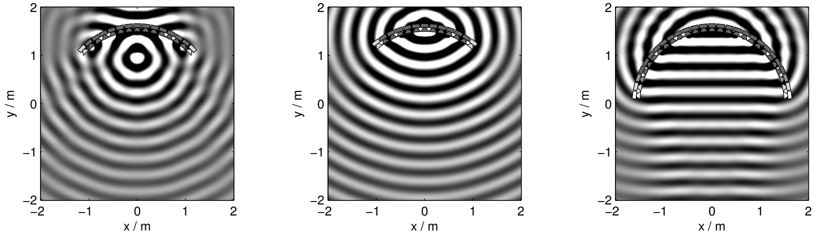
Perceptual audio coding

- Reproduced signal does not have to be PCM samples of $x(t)$
- In perceptual approaches the signal is a frequency-band representation with various masking effects taken into account
- What is the "signal" in spatial audio?

Examples of sound fields

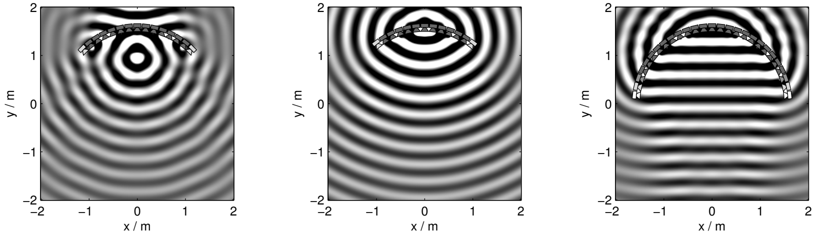


Physical approach: Wave field synthesis



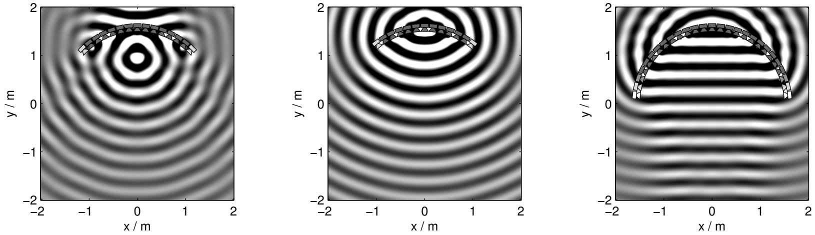
- If you reproduce the sound field totally, of course you will perceive the same space.

Physical approach: Wave field synthesis



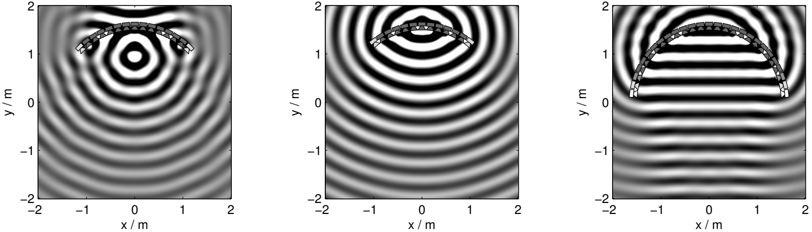
- If you reproduce the sound field totally, of course you will perceive the same space.
- 2D: Need of hundreds of loudspeakers, 3D: hundreds of thousands of loudspeakers

Physical approach: Wave field synthesis



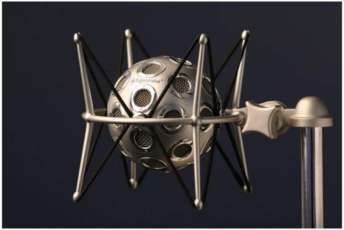
- If you reproduce the sound field totally, of course you will perceive the same space.
- 2D: Need of hundreds of loudspeakers, 3D: hundreds of thousands of loudspeakers
- Quality issues: spatial aliasing causes colorations, low frequency effects

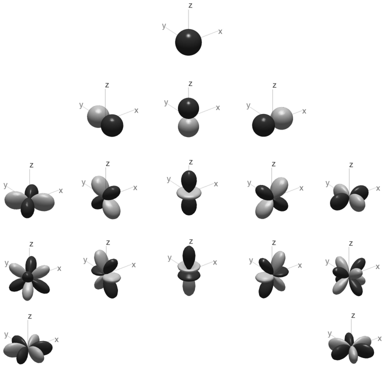
Physical approach: Wave field synthesis



- If you reproduce the sound field totally, of course you will perceive the same space.
- 2D: Need of hundreds of loudspeakers, 3D: hundreds of thousands of loudspeakers
- Quality issues: spatial aliasing causes colorations, low frequency effects
- Only synthesized content, no microphone techniques available

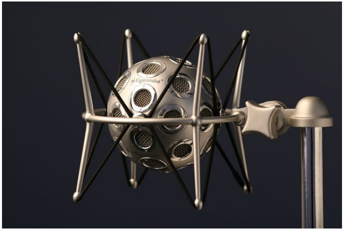
Physical approach: B-format recording

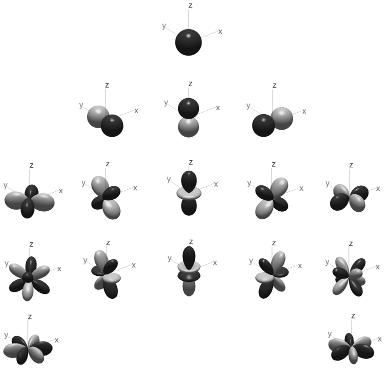




- Signals with directional patterns following to spherical harmonics

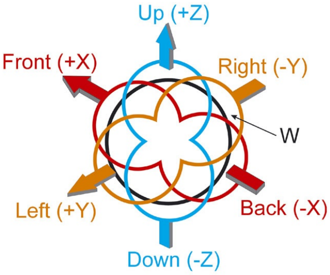
Physical approach: B-format recording





- Signals with directional patterns following to spherical harmonics
- Reproduce plane-wave expansion over loudspeakers

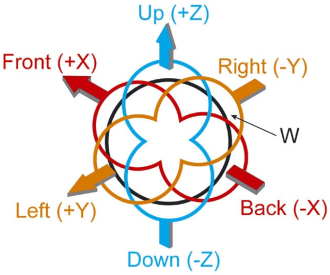
First-order B-format recording



www.soundfield.com

- Captures signals with zeroth-order and first-order spherical harmonics

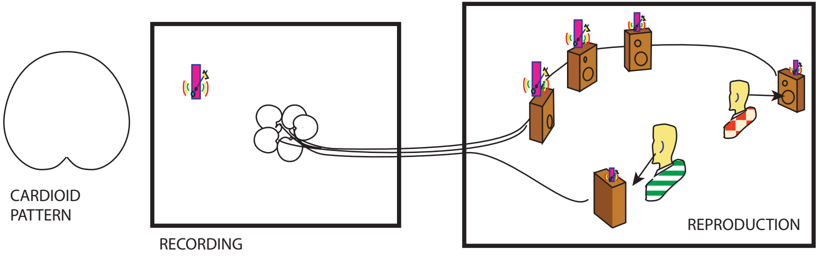
First-order B-format recording



www.soundfield.com

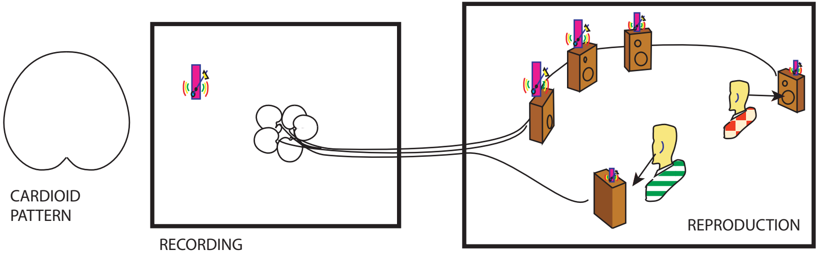
- Captures signals with zeroth-order and first-order spherical harmonics
- Pressure signal W. 3D velocity signals XYZ.

First-order Ambisonics



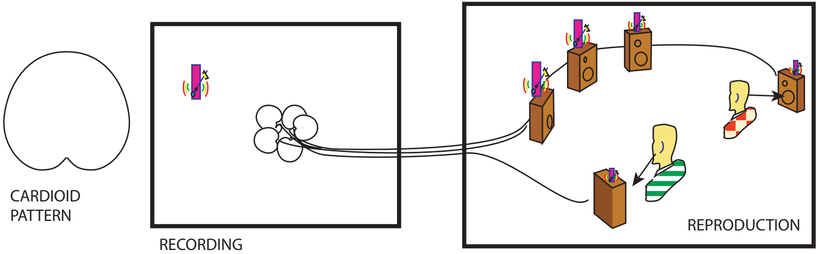
- Weighted sum of WXYZ signals (mixing, matrixing)

First-order Ambisonics



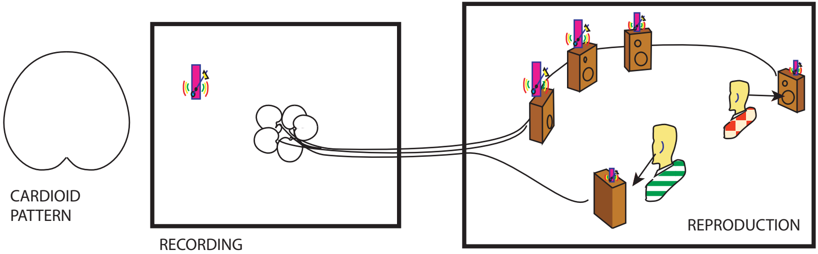
- Weighted sum of WXYZ signals (mixing, matrixing)
- High coherence between loudspeaker signals

First-order Ambisonics



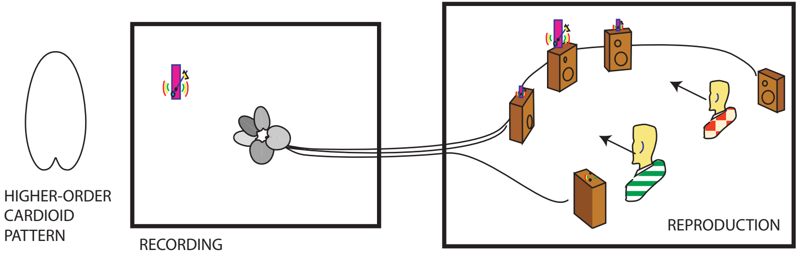
- Weighted sum of WXYZ signals (mixing, matrixing)
- High coherence between loudspeaker signals
- Spectral and spatial issues, very small listening area

First-order Ambisonics



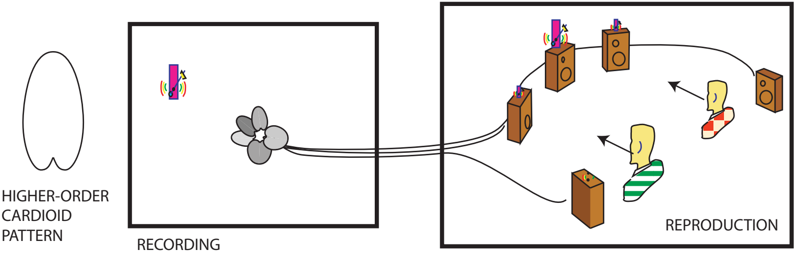
- Weighted sum of WXYZ signals (mixing, matrixing)
- High coherence between loudspeaker signals
- Spectral and spatial issues, very small listening area
- Moderate issues with low-frequency noise and spatial aliasing

Higher-order Ambisonics



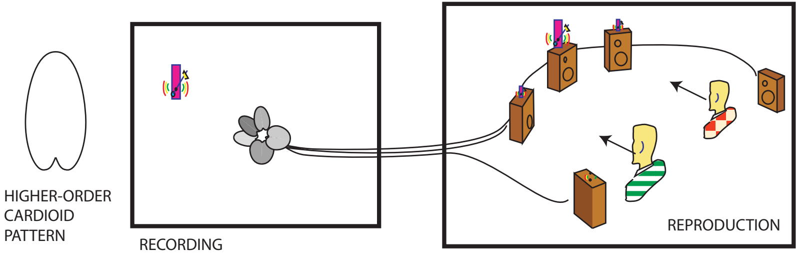
- More spherical harmonics captured

Higher-order Ambisonics



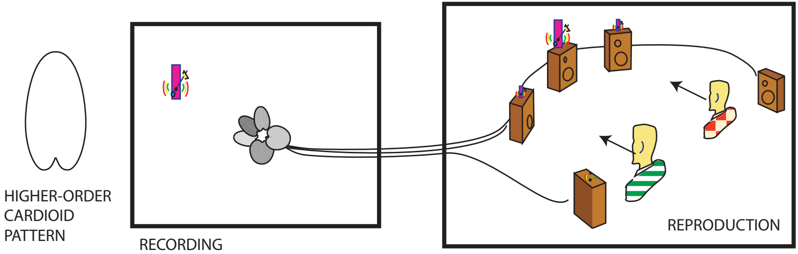
- More spherical harmonics captured
- Better resolution, more expensive devices

Higher-order Ambisonics



- More spherical harmonics captured
- Better resolution, more expensive devices
- Good quality in limited frequency window

Higher-order Ambisonics



- More spherical harmonics captured
- Better resolution, more expensive devices
- Good quality in limited frequency window
- Emphasized problems with low-frequency noise and high-frequency aliasing

Spatial sound perception

- Physical methods for spatial sound have shortcomings

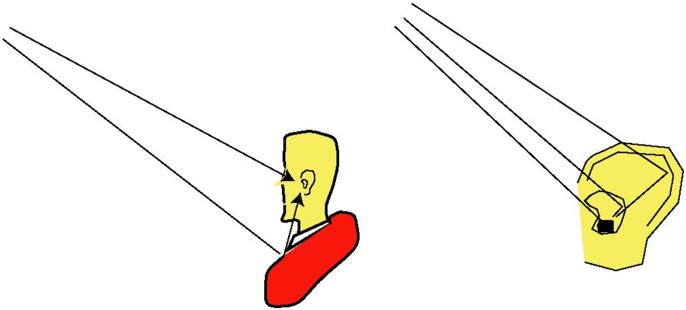
Spatial sound perception

- Physical methods for spatial sound have shortcomings
- Lets have a look on human spatial hearing

Spatial sound perception

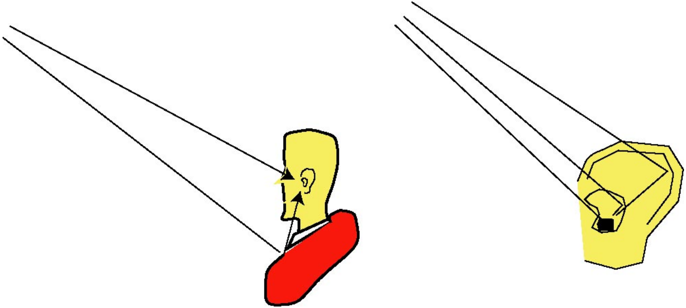
- Physical methods for spatial sound have shortcomings
- Lets have a look on human spatial hearing
- Could we bypass the problems somehow

Human spatial hearing



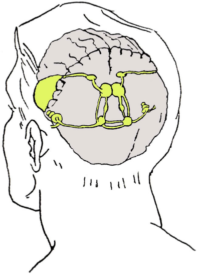
- One ear alone knows quite little of direction

Human spatial hearing



- One ear alone knows quite little of direction
- Response to very large range of wavelengths (2cm–30m)

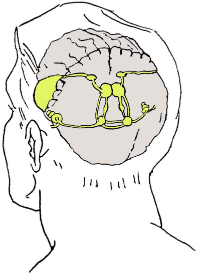
Human spatial hearing



ⓒJ. Blauert

- The perception of space is formed in signal analysis by the brains

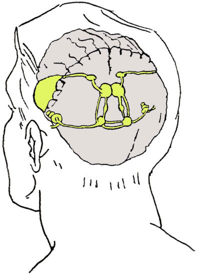
Human spatial hearing



©J. Blauert

- The perception of space is formed in signal analysis by the brains
- Signal characteristics in one ear / Signal differences between two ears

Human spatial hearing



©J. Blauert

- The perception of space is formed in signal analysis by the brains
- Signal characteristics in one ear / Signal differences between two ears
- Hearing mechanisms estimate the location of the source(s) and also the properties of the room

Monaural and binaural cues that carry spatial information

- Binaural differences depending on frequency and time
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)

Monaural and binaural cues that carry spatial information

- Binaural differences depending on frequency and time
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
- Head-related spectral cues, depending on time

Monaural and binaural cues that carry spatial information

- Binaural differences depending on frequency and time
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
- Head-related spectral cues, depending on time
- Dynamic change of cues depending on head movements

Monaural and binaural cues that carry spatial information

- Binaural differences depending on frequency and time
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
- Head-related spectral cues, depending on time
- Dynamic change of cues depending on head movements

Could we reproduce these cues somehow?

Reproducing spatial cues?

- Binaural recording $\rightarrow$ binaural playback, yes, but...

Reproducing spatial cues?

- Binaural recording $\rightarrow$ binaural playback, yes, but...
  - complicated to reproduce dynamic cues and individual HRTF characteristics

Reproducing spatial cues?

- Binaural recording $\rightarrow$ binaural playback, yes, but...
  - complicated to reproduce dynamic cues and individual HRTF characteristics
- How could we reproduce spatial auditory cues, if
  - input comes from a B-format microphone, and

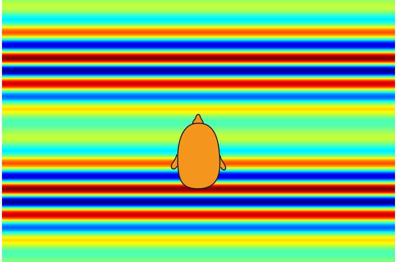
Reproducing spatial cues?

- Binaural recording $\rightarrow$ binaural playback, yes, but...
  - complicated to reproduce dynamic cues and individual HRTF characteristics
- How could we reproduce spatial auditory cues, if
  - input comes from a B-format microphone, and
  - output goes either to loudspeakers or headphones?

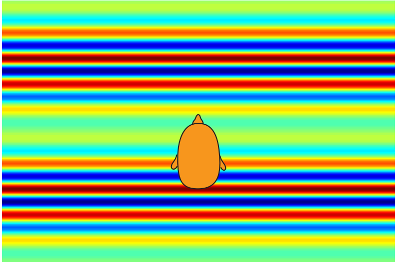
Reproducing spatial cues?

- Binaural recording → binaural playback, yes, but...
  - complicated to reproduce dynamic cues and individual HRTF characteristics
- How could we reproduce spatial auditory cues, if
  - input comes from a B-format microphone, and
  - output goes either to loudspeakers or headphones?
- Lets have a look at characteristic cases of sound fields

Plane wave

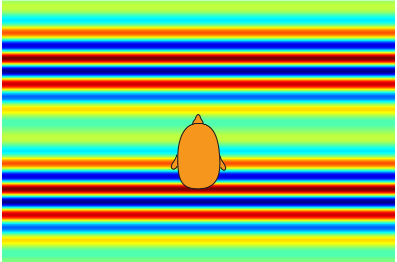


Plane wave



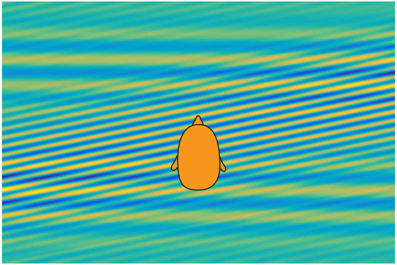
- Real field:
  - Consistent ITD, ILD and spectral cues
  - Accurate localization in most cases

Plane wave



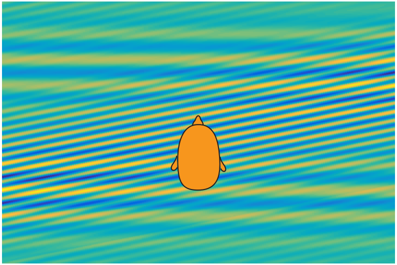
- Real field:
  - Consistent ITD, ILD and spectral cues
  - Accurate localization in most cases
- Reproduced field:
  - Should be reproduced as a plane wave preserving the spectral content

Several plane waves separated in frequency



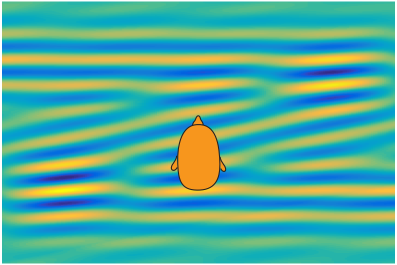
- Real field
  - Perceived as individual auditory objects

Several plane waves separated in frequency



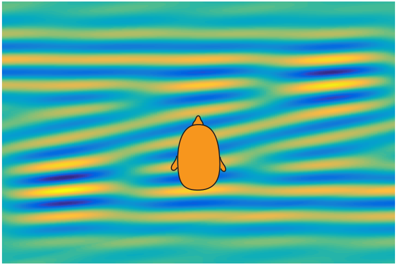
- Real field
  - Perceived as individual auditory objects
- Reproduced field
  - Spatial characteristics should be preserved

Several plane waves overlapping in frequency



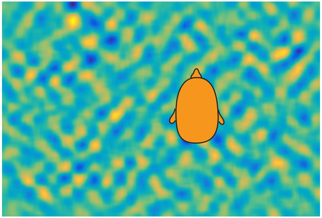
- Real field:
  - Localization may be erronious
  - Large difference in DOAs $\rightarrow$ blurred auditory image

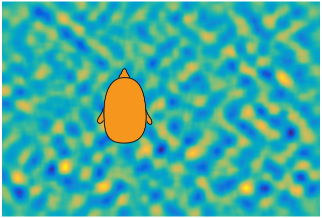
Several plane waves overlapping in frequency



- Real field:
  - Localization may be erronious
  - Large difference in DOAs $\rightarrow$ blurred auditory image
- Reproduced field
  - Reproduce spectral content right
  - *Spatial reconstruction does not have to be accurate!*

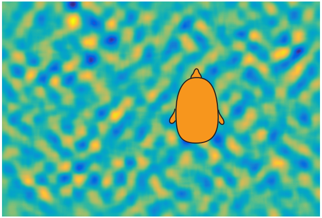
Diffuse field

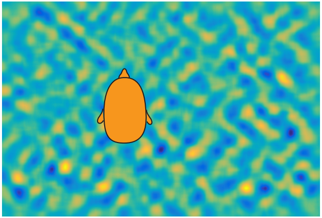




- Real field:
  - Often perceived surrounding the listener
  - Not sensitive to the instantaneous spatial fine structure of wave field

Diffuse field

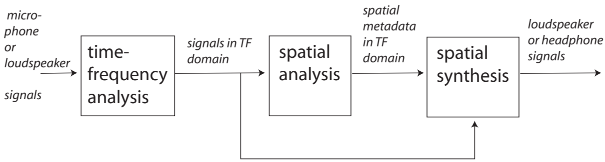




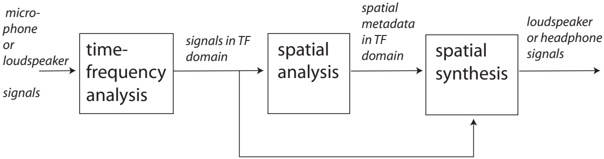
- Real field:
  - Often perceived surrounding the listener
  - Not sensitive to the instantaneous spatial fine structure of wave field
- Reproduced field:
  - Reproduce spectral content right
  - *Spatial reconstruction does not have to be accurate!*

Parametric time-frequency-domain spatial audio

Parametric time-frequency-domain spatial audio

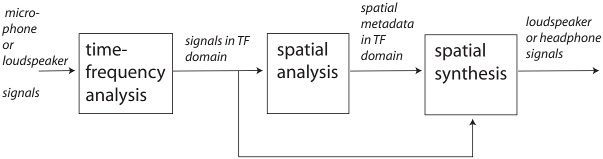


Parametric time-frequency-domain spatial audio



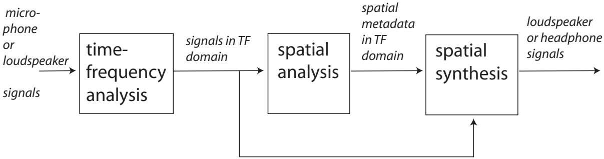
- Separation of wave field to plane waves and residual

Parametric time-frequency-domain spatial audio



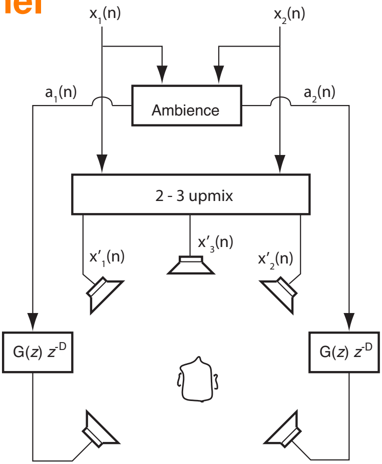
- Separation of wave field to plane waves and residual
- Reproduce plane waves with point-like virtual sources

Parametric time-frequency-domain spatial audio



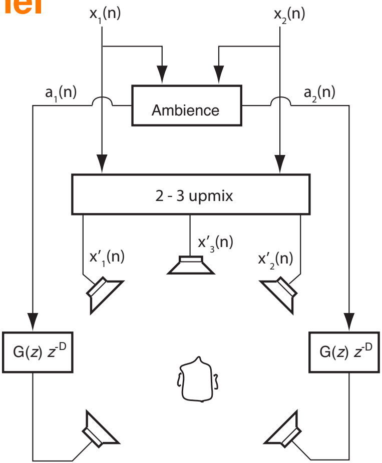
- Separation of wave field to plane waves and residual
- Reproduce plane waves with point-like virtual sources
- Reproduce the residual with surrounding method

First approach: upmixing of stereo to multichannel



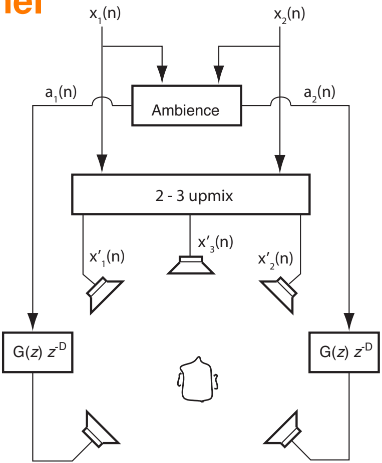
- Avendano & Jot [JAES, 2004] (Demo 2002 in AES 22nd Conf).

First approach: upmixing of stereo to multichannel



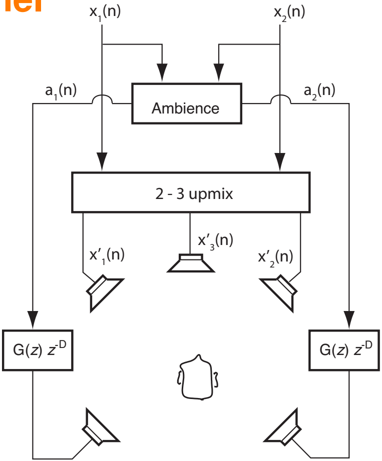
- Avendano & Jot [JAES, 2004] (Demo 2002 in AES 22nd Conf).
- Extract panned and ambient components from two-channel stereophonic input

First approach: upmixing of stereo to multichannel



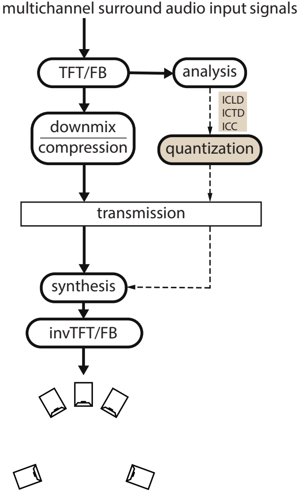
- Avendano & Jot [JAES, 2004] (Demo 2002 in AES 22nd Conf).
- Extract panned and ambient components from two-channel stereophonic input
- Use amplitude panning to reproduce "panned" components

First approach: upmixing of stereo to multichannel



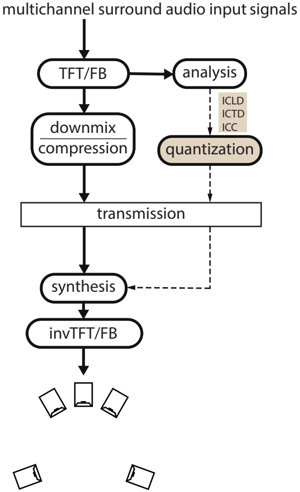
- Avendano & Jot [JAES, 2004] (Demo 2002 in AES 22nd Conf).
- Extract panned and ambient components from two-channel stereophonic input
- Use amplitude panning to reproduce "panned" components
- Reproduce ambient component with surround loudspeakers

Coding of multichannel audio



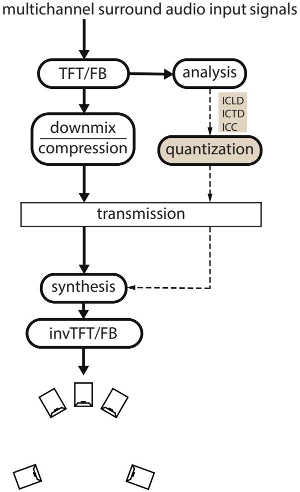
- Faller & Baumgarte [IEEE Trans. on Speech and Audio Processing, 2003]

Coding of multichannel audio



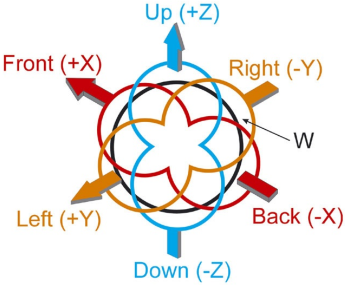
- Faller & Baumgarte [IEEE Trans. on Speech and Audio Processing, 2003]
- Inter-channel differences are used as metadata

Coding of multichannel audio



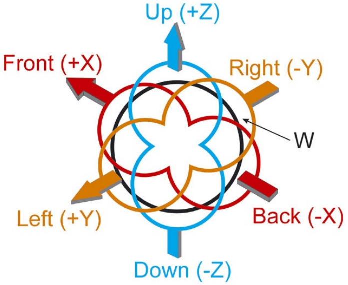
- Faller & Baumgarte [IEEE Trans. on Speech and Audio Processing, 2003]
- Inter-channel differences are used as metadata
- Huge savings in data rate

Reproduction of recorded spatial sound



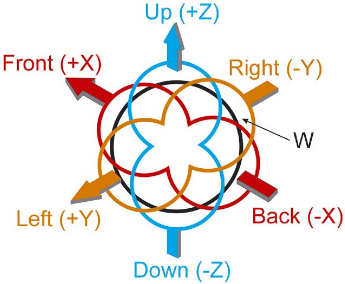
- 1999-2000: First-order B-format has some kind of Cartesian coordinate system, why does it not work?

Reproduction of recorded spatial sound



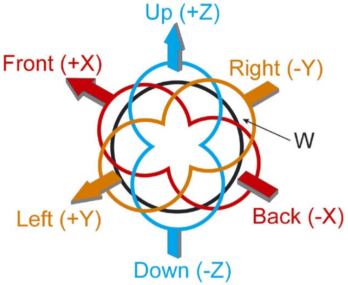
- 1999-2000: First-order B-format has some kind of Cartesian coordinate system, why does it not work?
- 2000: first idea of steering sound according to analyzed direction

Reproduction of recorded spatial sound



- 1999-2000: First-order B-format has some kind of Cartesian coordinate system, why does it not work?
- 2000: first idea of steering sound according to analyzed direction
- Collaboration with Juha Merimaa 2001–

Reproduction of recorded spatial sound



- 1999-2000: First-order B-format has some kind of Cartesian coordinate system, why does it not work?
- 2000: first idea of steering sound according to analyzed direction
- Collaboration with Juha Merimaa 2001–
- Spatial impulse response rendering (SIRR), Merimaa & Pulkki [WASPAA 2013]
- Directional audio coding (DirAC), Pulkki [JAES 2007]

Assumptions in DirAC

- At one time-frequency-position a listener

Assumptions in DirAC

- At one time-frequency-position a listener
  - perceives a direction if the ear canal signals are coherent

Assumptions in DirAC

- At one time-frequency-position a listener
  - perceives a direction if the ear canal signals are coherent
  - perceives a spatially broad auditory component if the ear canal signals are not coherent

Assumptions in DirAC

- At one time-frequency-position a listener
  - perceives a direction if the ear canal signals are coherent
  - perceives a spatially broad auditory component if the ear canal signals are not coherent
- Good reproduction quality is obtained, if we reproduce correctly the spectrum of sound and
  - diffuseness,

Assumptions in DirAC

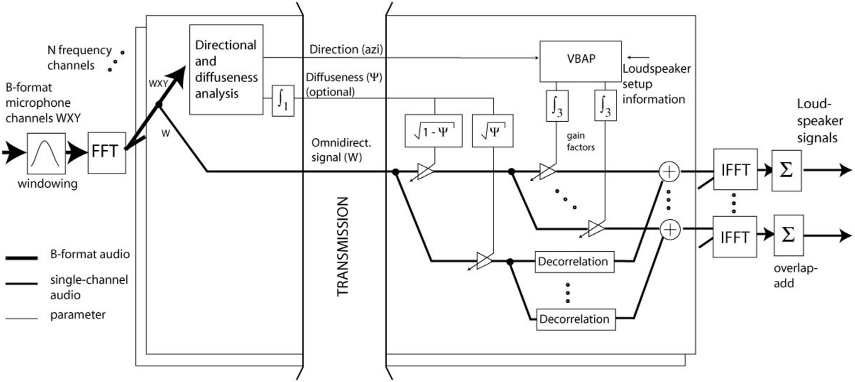
- At one time-frequency-position a listener
  - perceives a direction if the ear canal signals are coherent
  - perceives a spatially broad auditory component if the ear canal signals are not coherent
- Good reproduction quality is obtained, if we reproduce correctly the spectrum of sound and
  - diffuseness,
  - and if diffuseness is low: also direction

Assumptions in DirAC

- At one time-frequency-position a listener
  - perceives a direction if the ear canal signals are coherent
  - perceives a spatially broad auditory component if the ear canal signals are not coherent
- Good reproduction quality is obtained, if we reproduce correctly the spectrum of sound and
  - diffuseness,
  - and if diffuseness is low: also direction

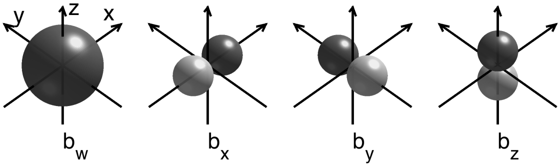
We thus indirectly assume thus that the sound field consists of single plane wave and diffuse component independently at each frequency band

"Teleconference" implementation



[Pulkki, AES Conv 2006]

Acoustical quantities measured with first-order B-format microphone



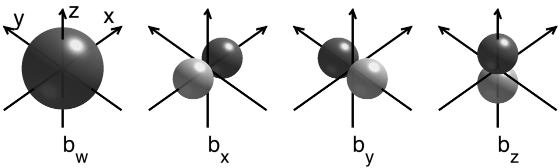
Pressure: $p(n) = (1/s)b_w(n)$

3D velocity vector: $\mathbf{u}(n) = -1/(s\rho_0 c\sqrt{2}) \begin{bmatrix} b_x(n) \\ b_y(n) \\ b_z(n) \end{bmatrix}$

3D intensity vector: $\mathbf{i}(n) = p(n)\mathbf{u}(n)$

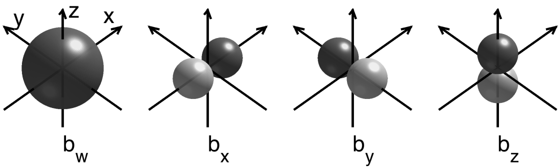
Instantaneous energy: $e = \frac{\rho_0}{2}||\mathbf{u}||^2 + |p|^2/2\rho_0 c^2$

Analysis of spatial parameters



Direction:	intensity vector	$\mathrm{DOA} = \angle \mathrm{E}[-\mathbf{i}]$
Diffuseness	net flow / total energy	$\psi = 1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{c\mathrm{E}[e]}$
	temporal fluctuation of **i**	$\psi = \sqrt{1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{\mathrm{E}[\|\|\mathbf{i}\|\|]}}$

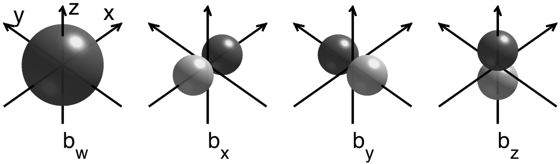
Analysis of spatial parameters



Direction:	intensity vector	$\mathrm{DOA} = \angle\, \mathrm{E}[-\mathbf{i}]$
Diffuseness	net flow / total energy	$\psi = 1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{c\mathrm{E}[e]}$
	temporal fluctuation of **i**	$\psi = \sqrt{1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{\mathrm{E}[\|\|\mathbf{i}\|\|]}}$

- Robust to small deviations from ideal microphone characteristics

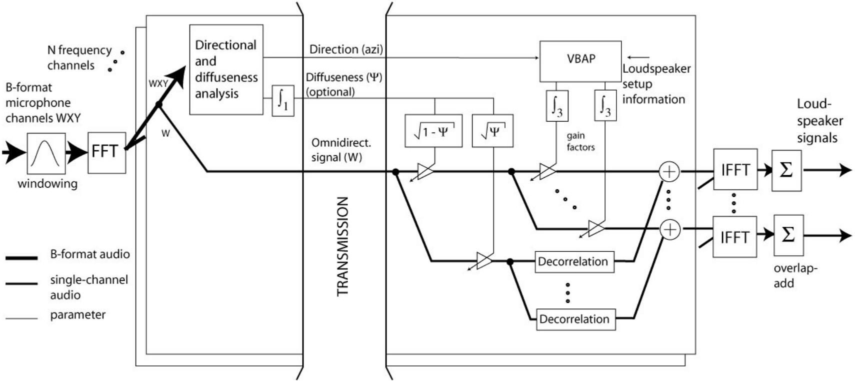
Analysis of spatial parameters



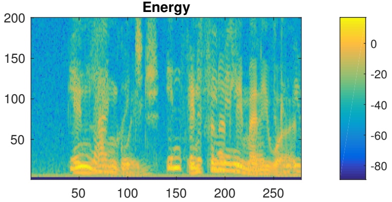
Direction:	intensity vector	$\mathrm{DOA} = \angle \mathrm{E}[-\mathbf{i}]$
Diffuseness	net flow / total energy	$\psi = 1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{c\mathrm{E}[e]}$
	temporal fluctuation of **i**	$\psi = \sqrt{1 - \frac{\|\|\mathrm{E}[\mathbf{i}]\|\|}{\mathrm{E}[\|\|\mathbf{i}\|\|]}}$

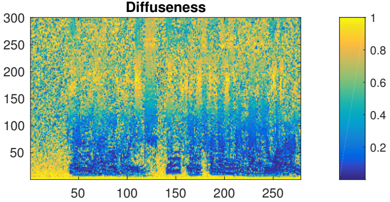
- Robust to small deviations from ideal microphone characteristics
- Measures energetic properties of sound field within given frequency band, useful?

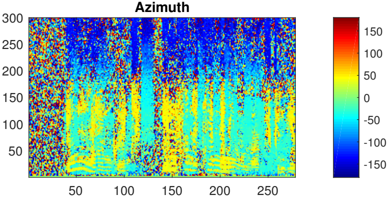
"Teleconference" implementation

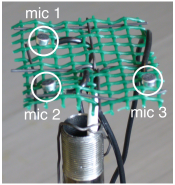


Example with low-end 3-microphone 2D array





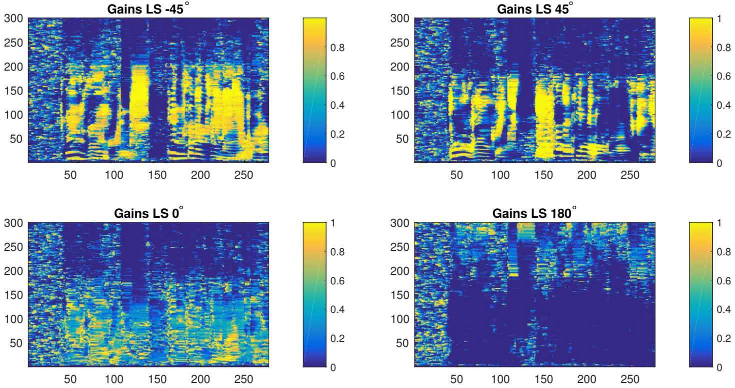




Two speech sources in $\pm 45^{\circ}$, anechoic chamber

▷ Sound captured with one of the microphones

Examples of soft masks for non-diffuse stream



Rendering to 8-channel octagonal loudspeaker setup

▷ Mono ▷ ND2ch ▷ D2ch ▷ ND+D2ch ▷ ALL CHAN

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts
- Sources on opposite sides of the microphone with spectral overlap

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts
- Sources on opposite sides of the microphone with spectral overlap
  - Severe model mismatch

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts
- Sources on opposite sides of the microphone with spectral overlap
  - Severe model mismatch
  - Timbral artifacts: "added room effect", "smearing of transients"

Properties of teleconference-DirAC

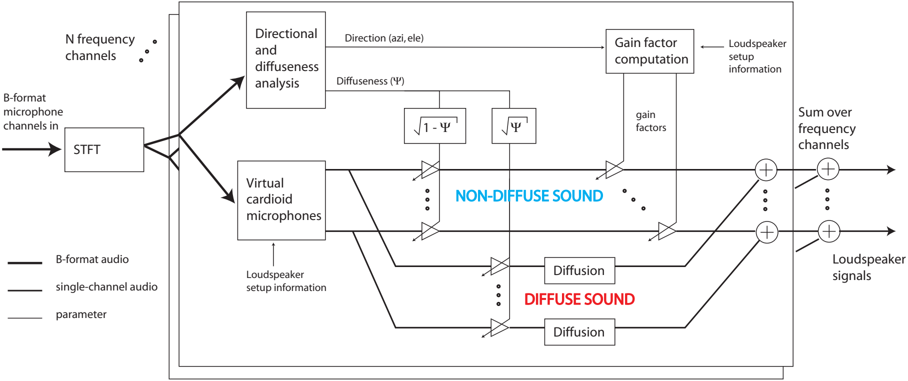
- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts
- Sources on opposite sides of the microphone with spectral overlap
    - Severe model mismatch
    - Timbral artifacts: "added room effect", "smearing of transients"
    - Spatial artifacts: "sources pull each other"

Properties of teleconference-DirAC

- Very good quality with spectrally non-overlapping sources in free field
- Diffuse reverberation subject to spatial and timbral artifacts
- Sources on opposite sides of the microphone with spectral overlap
  - Severe model mismatch
  - Timbral artifacts: "added room effect", "smearing of transients"
  - Spatial artifacts: "sources pull each other"
  - Sources near noise masking threshold: impossible to localize

Why these artifacts?

"HQ" implementation



This works better, but dont exactly know why (2007).

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely
- Some artifacts still present
- Challenging acoustical conditions

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely
- Some artifacts still present
- Challenging acoustical conditions
  - Surrounding applause signals

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely
- Some artifacts still present
- Challenging acoustical conditions
  - Surrounding applause signals
  - Small rooms with strong early reflections

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely
- Some artifacts still present
- Challenging acoustical conditions
  - Surrounding applause signals
  - Small rooms with strong early reflections
  - Several sources overlapping in spectrum

Properties of HQ-DirAC

- All teleconference-DirAC artifacts are mitigated largely
- Some artifacts still present
- Challenging acoustical conditions
  - Surrounding applause signals
  - Small rooms with strong early reflections
  - Several sources overlapping in spectrum
- Potential loss of energy

Limitations of differential arrays

- Low-frequency noise

Limitations of differential arrays

- Low-frequency noise
  - Instable direction and diffuseness estimation
  - Mitigated with temporal integration

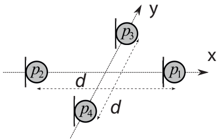
Limitations of differential arrays

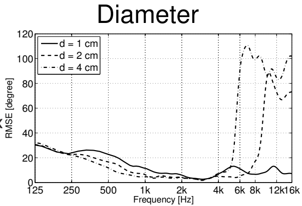
- Low-frequency noise
  - Instable direction and diffuseness estimation
  - Mitigated with temporal integration
- Spatial aliasing

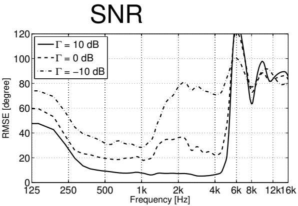
Limitations of differential arrays

- Low-frequency noise
  - Instable direction and diffuseness estimation
  - Mitigated with temporal integration
- Spatial aliasing
  - Highly biased directional values
  - In some cases can be mitigated

Square array

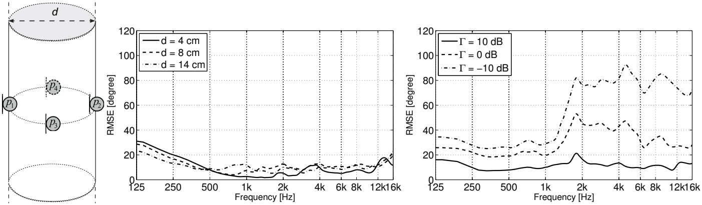






- Pressure gradient
- Square arrays of omni microphones, B-format microphones
- LF noise, HF aliasing
- Ahonen, del Galdo et al [JAES 2012]

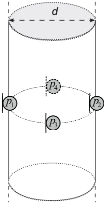
Arrays with shadowing

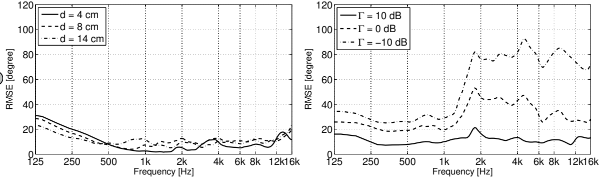


$$
\begin{aligned}
\widetilde{i}_x(n,k) &= |p_1(n,k)|^2 - |p_2(n,k)|^2, \\
\widetilde{i}_y(n,k) &= |p_3(n,k)|^2 - |p_4(n,k)|^2,
\end{aligned}
$$

- When capsule signals are available, and some shadowing takes place

Arrays with shadowing

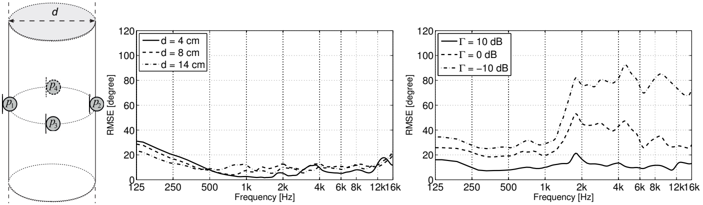




$$\tilde{i}_x(n,k) = |p_1(n,k)|^2 - |p_2(n,k)|^2,$$
$$\tilde{i}_y(n,k) = |p_3(n,k)|^2 - |p_4(n,k)|^2,$$

- When capsule signals are available, and some shadowing takes place
- LF+MF: pressure gradient
- HF: energy gradient

Arrays with shadowing



$$\widetilde{i}_x(n,k) = |p_1(n,k)|^2 - |p_2(n,k)|^2,$$
$$\widetilde{i}_y(n,k) = |p_3(n,k)|^2 - |p_4(n,k)|^2,$$

- When capsule signals are available, and some shadowing takes place
- LF+MF: pressure gradient
- HF: energy gradient
- A-format microphones, cylinder arrays and spherical arrays

Other microphone arrays with DirAC

- Multiple microphones in array: ESPRIT by Thiergart, Kratschmer et al, [AES Conv 2011]

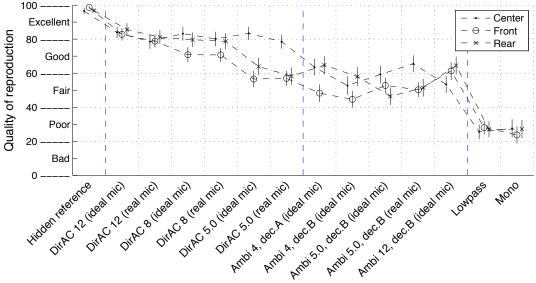
Other microphone arrays with DirAC

- Multiple microphones in array: ESPRIT by Thiergart, Kratschmer et al, [AES Conv 2011]
- Two microphones, cross-correlation: Kratschmer, Thiergart et al [AES Conv 2012]

Other microphone arrays with DirAC

- Multiple microphones in array: ESPRIT by Thiergart, Kratschmer et al, [AES Conv 2011]
- Two microphones, cross-correlation: Kratschmer, Thiergart et al [AES Conv 2012]
- Basically any DOA analysis method can be applied

HQ-DirAC subjective tests



- HQ-DirAC, comparison to reference scenario with 24 loudspeakers
- Largest issues with spatially complex scenarios audible as small timbral artifacts
- Vilkamo [JAES 2009]

Sources of timbral artifacts

- Diffuse sound leaks into non-diffuse stream
  - This has not found to be a problem

Sources of timbral artifacts

- Diffuse sound leaks into non-diffuse stream
  - This has not found to be a problem
- Non-diffuse sound leaks into diffuse stream
  - Major problem
  - Transients are decorrelated, causing annoying smearing
  - Direct sound is decorrelated, "added room effect", or "sources are perceived too far" issues

Sources of timbral artifacts

- Diffuse sound leaks into non-diffuse stream
  - This has not found to be a problem
- Non-diffuse sound leaks into diffuse stream
  - Major problem
  - Transients are decorrelated, causing annoying smearing
  - Direct sound is decorrelated, "added room effect", or "sources are perceived too far" issues
- Target for development: minimize decorrelated energy!

How to avoid decorrelation

- Processing of transients separately

How to avoid decorrelation

- Processing of transients separately
  - Recognize transients
  - Use better time resolution / bypass decorrelation [Laitinen, Kuech et al. 2011]

How to avoid decorrelation

- Processing of transients separately
  - Recognize transients
  - Use better time resolution / bypass decorrelation [Laitinen, Kuech et al. 2011]
- Covariance-domain processing
  - minimize decorrelated energy

How to avoid decorrelation

- Processing of transients separately
  - Recognize transients
  - Use better time resolution / bypass decorrelation [Laitinen, Kuech et al. 2011]
- Covariance-domain processing
  - minimize decorrelated energy
- Divide sound field into sectors from higher-order recording
  - perform separate analysis for each sector

How to avoid decorrelation

- Processing of transients separately
  - Recognize transients
  - Use better time resolution / bypass decorrelation [Laitinen, Kuech et al. 2011]
- Covariance-domain processing
  - minimize decorrelated energy
- Divide sound field into sectors from higher-order recording
  - perform separate analysis for each sector
- Perform more elaborate analysis to sound field (e.g., multiple DOA values), Thiergart [IEEE TASLP, 2014]

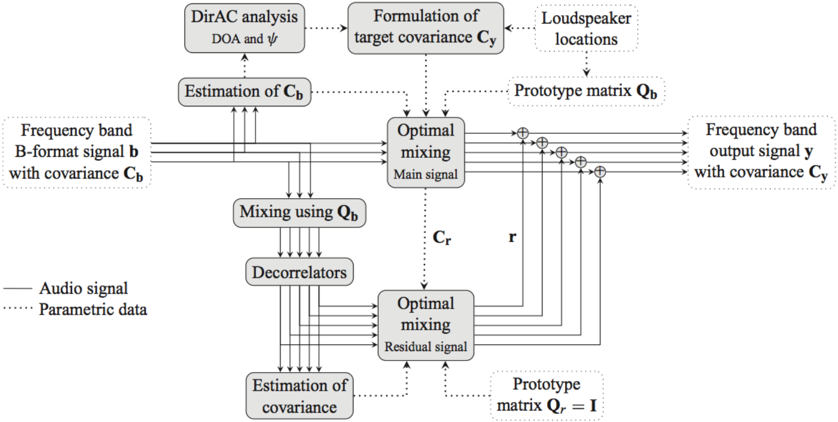
Covariance-domain processing

Least-squares optimized solution for synthesis

- the covariance matrix of output is dictated by directional parameters
- optimized mixing solution leads to minimization of decorrelated energy

[Vilkamo, Bäckström, Kuntz: JAES 2013]

Covariance-domain processing



Solutions with different model of the sound field

- Higher number of microphones gives more information about sound field

Solutions with different model of the sound field

- Higher number of microphones gives more information about sound field
- How to use that information in sound reproduction?

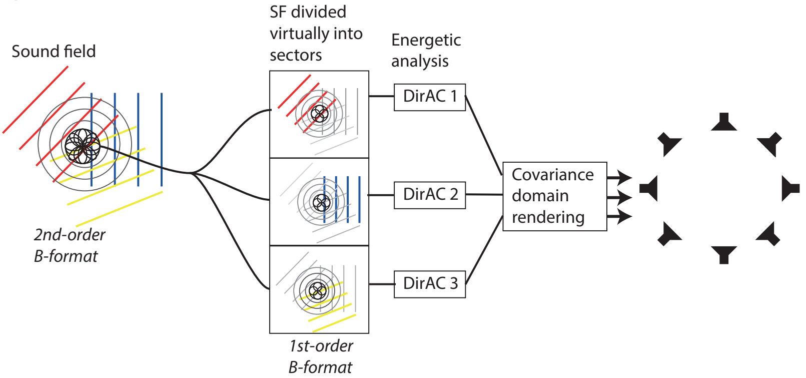
Solutions with different model of the sound field

- Higher number of microphones gives more information about sound field
- How to use that information in sound reproduction?
- Divide sound field into sectors (Pulkki, Politis), perform lower-order reproduction for each

Solutions with different model of the sound field

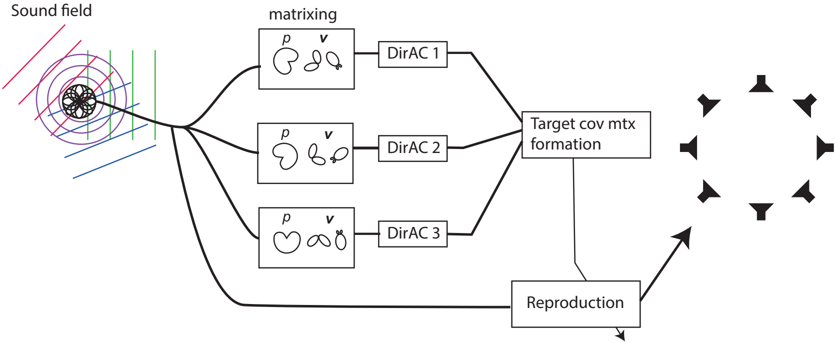
- Higher number of microphones gives more information about sound field
- How to use that information in sound reproduction?
- Divide sound field into sectors (Pulkki, Politis), perform lower-order reproduction for each
- Analyze multiple DOAs, and then reproduce (Thiergart & Habets, Mouchtaris group, Berge)

Sector-based parametric spatial sound reproduction



[Politis et al: IEEE J. Selected Topics Sig Proc 9.5 (2015)]

Sector-based parametric spatial sound reproduction



Sector-based parametric spatial sound reproduction

"Higher-order DirAC"

Sector-based parametric spatial sound reproduction

"Higher-order DirAC"

- Challenging acoustical conditions occur rarely within sectors

Sector-based parametric spatial sound reproduction

"Higher-order DirAC"

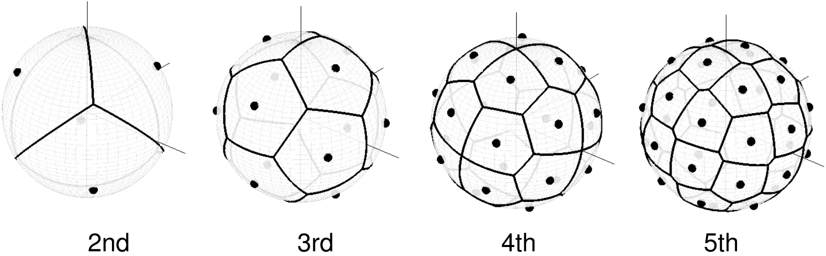
- Challenging acoustical conditions occur rarely within sectors
- Parameters computed with N:th -order input
- Audio signals used in synthesis obtained with (N-1):th -order input
- Self-noise issue of higher-order microphones are also avoided

Sector-based parametric spatial sound reproduction

"Higher-order DirAC"

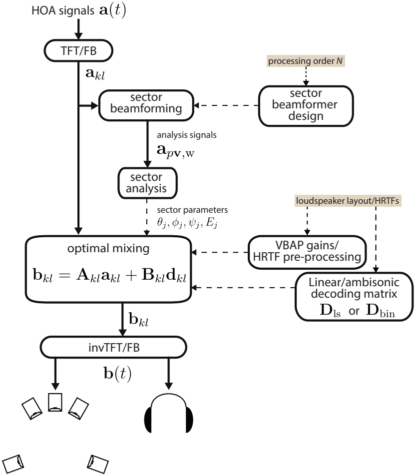
- Challenging acoustical conditions occur rarely within sectors
- Parameters computed with N:th -order input
- Audio signals used in synthesis obtained with (N-1):th -order input
- Self-noise issue of higher-order microphones are also avoided
- System does not lose acoustic energy in any case

Sectors for HO-microphones

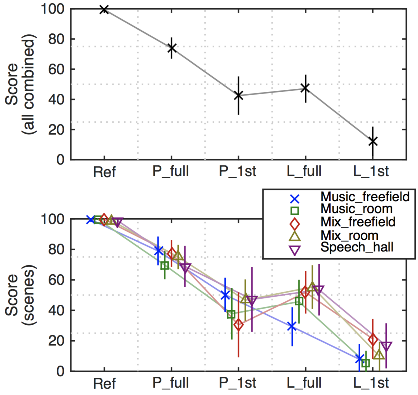


Different frequency bands utilize different number of sectors

Processing


HOA signals $\mathbf{a}(t)$
TFT/FB
$\mathbf{a}_{kl}$
processing order N
sector beamforming
sector beamformer design
analysis signals
$\mathbf{a}_{p\mathbf{v},\mathrm{w}}$
sector analysis
sector parameters
$\theta_j, \phi_j, \psi_j, E_j$
loudspeaker layout/HRTFs
optimal mixing
$\mathbf{b}_{kl} = \mathbf{A}_{kl}\mathbf{a}_{kl} + \mathbf{B}_{kl}\mathbf{d}_{kl}$
VBAP gains/
HRTF pre-processing
Linear/ambisonic decoding matrix
$\mathbf{D}_{\mathrm{ls}}$ or $\mathbf{D}_{\mathrm{bin}}$
$\mathbf{b}_{kl}$
invTFT/FB
$\mathbf{b}(t)$

Subjective evaluation



- [Politis & Vilkamo & Pulkki IEEE J. Selected Topics Sig Proc 9.5 (2015)]
- Reference: 28 loudspeakers in anechoic chamber, very challenging 3D sound environments
- Test: Eigenmic recording, playback over HO-DirAC, 1st-order DirAC, 4th-order Ambisonics, 1st-order Ambisonics

Why does HO-DirAC provide better results?

- Spatially separated plane waves sharing the same frequency are processed in different sectors

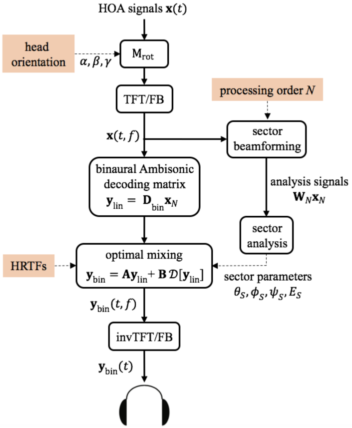
Why does HO-DirAC provide better results?

- Spatially separated plane waves sharing the same frequency are processed in different sectors
- Global diffuse field is not diffuse in individual sectors, nevertheless, the combined output is again diffuse

Why does HO-DirAC provide better results?

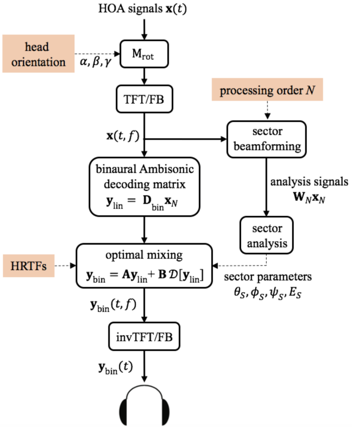
- Spatially separated plane waves sharing the same frequency are processed in different sectors
- Global diffuse field is not diffuse in individual sectors, nevertheless, the combined output is again diffuse
- Avoidance of decorrelation!

HO-DirAC for head-tracked headphones



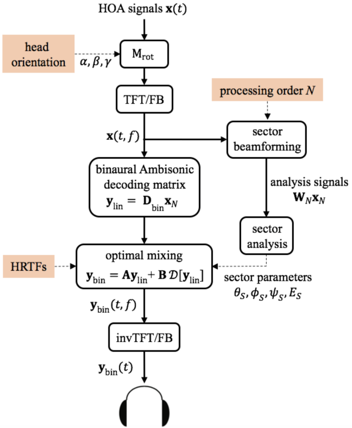
- Processing optimized for dynamic rendering

HO-DirAC for head-tracked headphones



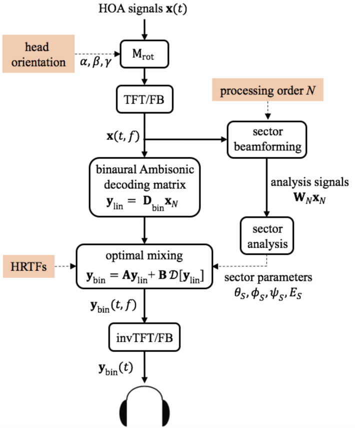
- Processing optimized for dynamic rendering
- Sector-based computation is used to derive covariance matrices

HO-DirAC for head-tracked headphones



- Processing optimized for dynamic rendering
- Sector-based computation is used to derive covariance matrices
- WASPAA 2017: Wed 10:30–12:30 *Enhancement of ambisonic binaural reproduction...* Politis, McCormack, Pulkki

HO-DirAC for head-tracked headphones



- Processing optimized for dynamic rendering
- Sector-based computation is used to derive covariance matrices
- WASPAA 2017: Wed 10:30–12:30 *Enhancement of ambisonic binaural reproduction...* Politis, McCormack, Pulkki
- **DEMOS AVAILABLE**

Time-frequency spatial audio in general

- You can lay down the assumptions differently

Time-frequency spatial audio in general

- You can lay down the assumptions differently
- Different applications exist

Time-frequency spatial audio in general

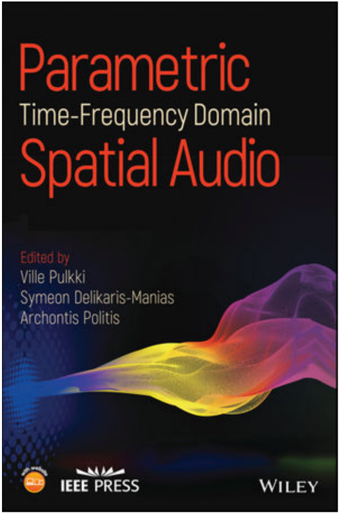
- You can lay down the assumptions differently
- Different applications exist
- A number of methods for different tasks in spatial audio have resulted in

Time-frequency spatial audio in general

- You can lay down the assumptions differently
- Different applications exist
- A number of methods for different tasks in spatial audio have resulted in
- ... and we have edited a book about that. :-)

Book

- 15 chapters, 416 pages
- Matlab code
- Available in Dec 2017



Analysis and synthesis

- Time-frequency processing – methods and tools
  J. Vilkamo, T. Bäckström

Analysis and synthesis

- Time-frequency processing – methods and tools
  J. Vilkamo, T. Bäckström
- Spatial decomposition by spherical array processing
  D. L. Alon and B. Rafaely

Analysis and synthesis

- Time-frequency processing – methods and tools
  J. Vilkamo, T. Bäckström
- Spatial decomposition by spherical array processing
  D. L. Alon and B. Rafaely
- Sound Field Analysis using Sparse Recovery
  C. T. Jin, N. Epain and T. Noohi

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki
- First-order directional audio coding (DirAC)
  V. Pulkki, M.-V. Laitinen, J. Vilkamo, and J. Ahonen

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki
- First-order directional audio coding (DirAC)
  V. Pulkki, M.-V. Laitinen, J. Vilkamo, and J. Ahonen
- Higher-order Directional Audio Coding
  A. Politis and V. Pulkki

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki
- First-order directional audio coding (DirAC)
  V. Pulkki, M.-V. Laitinen, J. Vilkamo, and J. Ahonen
- Higher-order Directional Audio Coding
  A. Politis and V. Pulkki
- Multi-channel sound acquisition using a multi-wave sound field model
  O. Thiergart and E. Habets

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki
- First-order directional audio coding (DirAC)
  V. Pulkki, M.-V. Laitinen, J. Vilkamo, and J. Ahonen
- Higher-order Directional Audio Coding
  A. Politis and V. Pulkki
- Multi-channel sound acquisition using a multi-wave sound field model
  O. Thiergart and E. Habets
- Adaptive mixing of noisy and robust beamformers for reproduction of spatial sound
  S. Delikaris-Manias and J. Vilkamo

Reproduction of spatial sound

- Overview to time-frequency-domain parametric spatial audio techniques
  A. Politis, S. Delikaris-Manias and V. Pulkki
- First-order directional audio coding (DirAC)
  V. Pulkki, M.-V. Laitinen, J. Vilkamo, and J. Ahonen
- Higher-order Directional Audio Coding
  A. Politis and V. Pulkki
- Multi-channel sound acquisition using a multi-wave sound field model
  O. Thiergart and E. Habets
- Adaptive mixing of noisy and robust beamformers for reproduction of spatial sound
  S. Delikaris-Manias and J. Vilkamo
- Source Separation and Reconstruction of Spatial Audio Using Spectrogram Factorization
  J. Nikunen and T. Virtanen

Signal-dependent spatial filtering

- Time-frequency-domain spatial audio enhancement
  S. Delikaris-Manias and P. Pertila

Signal-dependent spatial filtering

- Time-frequency-domain spatial audio enhancement
  S. Delikaris-Manias and P. Pertila
- Cross-spectrum-based post filter utilizing noisy and robust beamformers
  S. Delikaris-Manias, V. Pulkki

Signal-dependent spatial filtering

- Time-frequency-domain spatial audio enhancement
  S. Delikaris-Manias and P. Pertila
- Cross-spectrum-based post filter utilizing noisy and robust beamformers
  S. Delikaris-Manias, V. Pulkki
- Microphone array-based speech enhancement using neural networks
  P. Pertila

Applications

- Upmixing and Beamforming in Professional Audio
  C. Faller

Applications

- Upmixing and Beamforming in Professional Audio
  C. Faller
- Spatial sound scene synthesis and manipulation for virtual reality and audio effects
  V. Pulkki, A. Politis, T. Pihlajamaki and M-V. Laitinen

Applications

- Upmixing and Beamforming in Professional Audio
  C. Faller
- Spatial sound scene synthesis and manipulation for virtual reality and audio effects
  V. Pulkki, A. Politis, T. Pihlajamaki and M-V. Laitinen
- Parametric spatial audio techniques in teleconferencing and remote presence
  A. Alexandridis, D. Pavlidi, N. Stefanakis, and A. Mouchtaris

Conclusions

- Parametric time-frequency-domain spatial audio – treating also spatial auditory cues as signal in reproduction

Conclusions

- Parametric time-frequency-domain spatial audio – treating also spatial auditory cues as signal in reproduction
- Task-specific signal-dependent and spatial-condition-dependent non-linear DSP

Conclusions

- Parametric time-frequency-domain spatial audio – treating also spatial auditory cues as signal in reproduction
- Task-specific signal-dependent and spatial-condition-dependent non-linear DSP
- Enhancement of quality of spatial sound when compared with linear methods