DoA-estimation based on a complex Watson kernel method

Lukas Drude, Florian Jacob, Reinhold Haeb-Umbach

Department of Communications Engineering - University of Paderborn

September 01, 2015

Table of contents

Introduction

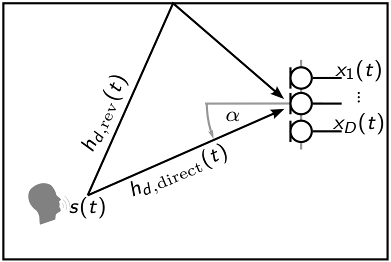
Motivation

- Direction of Arrival (DoA)-estimation is important for:
  - Speaker localization and tracking
  - Blind Source Separation (BSS)
  - Beamforming
  - Geometry calibration
  - ...

Task

- Estimate DoA from multi-channel audio signal
- Requirements:
  - Account for spatial aliasing
  - Exploit phase and level differences (microphone directivity relevant for beamforming [Gaubitch et al. 2014])
  - Derive statistical approach

Signal model



Time domain model

$$\begin{aligned} x_d(t) =& h_{d,\mathrm{direct}}(t) * s(t) \\ & + \underbrace{h_{d,\mathrm{rev}}(t) * s(t) + n_d(t)}_{\tilde{n}_d(t)} \end{aligned}$$

Assumption

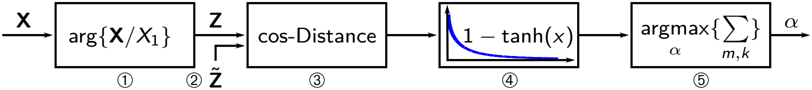
- Reverberation is uncorrelated from direct component

Short Time Fourier Transform (STFT) model

$$\mathbf{X}(m,k) = \mathbf{H}_{\mathrm{direct}}(k) S(m,k) + \tilde{\mathbf{N}}(m,k)$$

where m: frame index, k: frequency bin

[Loesch et al. 2010]



Procedure

1. Get phase relative to 1st channel
2. Compute expected phase for candidate direction: $\tilde{Z}_d = 2\pi f(k)\tau_d(\alpha)$
3. Calculate deviation: $c = 2D - 2\sum_{d=1}^{D} \cos(Z_d - \tilde{Z}_d)$
4. Reduce influence of outliers and increase spatial resolution
5. Sum up TF-slots and select maximum

Summary

+ Account for spatial aliasing
– Heuristic non-linearity

Proposed approach

Requirements

- Avoid heuristic non-linearity
- Additionally: Exploit level differences

Prepare microphone signal

- Source power influences length of $\mathbf{X}(m, k)$
- Absolute power contains no information $\Rightarrow$ Remove absolute power

$$\mathbf{Y}(m, k) = \mathbf{X}(m, k) / \|\mathbf{X}(m, k)\|$$

- Phase and level difference due to propagation conditions are maintained
- Observations $\mathbf{Y}(m, k)$ are located on complex unit hyper-sphere

Wanted

- Distribution to model the clusters related to the source position/direction

Observation model

Complex Watson probability density function

- 2 parameters: concentration κ and mode vector $\mathbf{W}$
- Observations $\mathbf{Y}$ are located on complex unit hyper-sphere

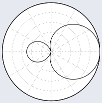
$$p(\mathbf{Y}; \kappa, \mathbf{W}) = \frac{1}{c_{\mathrm{W}}(\kappa)} \mathrm{e}^{\kappa \left| \mathbf{W}^{\mathrm{H}} \mathbf{Y} \right|^2}$$

Applications of complex Watson distribution

- Image recognition [Mardia et al. 1999]
- BSS [Ito et al 2013], [Tran Vu et al. 2013], [Jafari et al. 2014]

Model

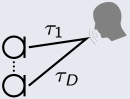
Anechoic model



ideal directivity pattern

$$\widetilde{\mathbf{W}}(k,\alpha) = \Big[\underbrace{A_d(\alpha)}_{\text{Attenuation}}\underbrace{\mathrm{e}^{-2\pi \mathrm{j} f(k)\tau_d(\alpha)}}_{\text{Phase}}\Big]_{1\leq d\leq D}$$

$$\mathbf{W} = \widetilde{\mathbf{W}}/\|\widetilde{\mathbf{W}}\|$$



- Get impinging direction α related to the direct path

$\Rightarrow$ Find best match with anechoic model

- Candidate model can be 3D or positions

Summary

$$\mathbf{X} \rightarrow \boxed{\frac{\mathbf{X}(m,k)}{\|\mathbf{X}(m,k)\|}} \xrightarrow{\mathbf{Y}} \boxed{p(\mathbf{Y};\kappa,\mathbf{W})} \rightarrow \boxed{\underset{\alpha}{\operatorname{argmax}}\Big\{\sum_{m,k}\Big\}} \xrightarrow{\alpha}$$

$\mathbf{W}$

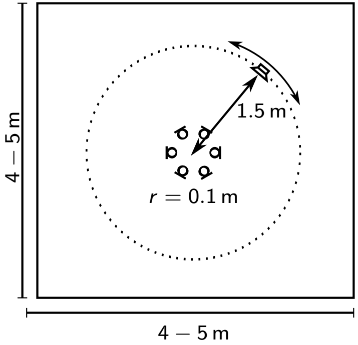
$+$ Exploit phase <u>and</u> level differences

$+$ Only one parameter instead of heuristic function

Experimental evaluation

Setup

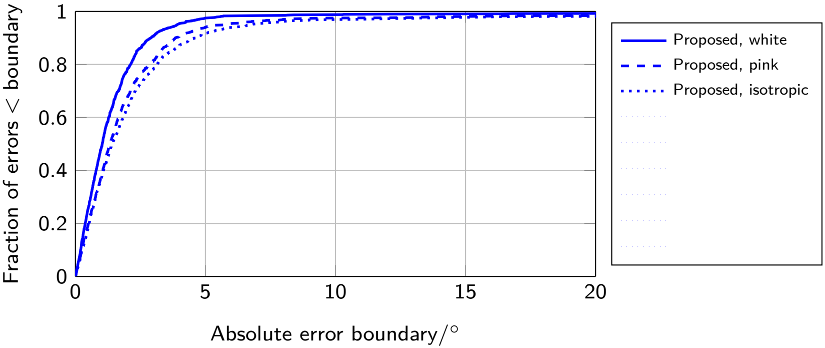
- 1000 sampled setups
- Speech samples from TIMIT database (Duration 1 sec)
- $T_{60} = 400\,\mathrm{ms}$ (image method)
- Noise types (SNR $= 10\,\mathrm{dB}$):
  - white (spectral and spatial)
  - pink (spatial white)
  - isotropic (spectral white)



Algorithms

- [Loesch et al. 2010]
- Proposed: Waston Kernel
- Steered Response Power with Phase Transform (SRP-PHAT) [DSB01]

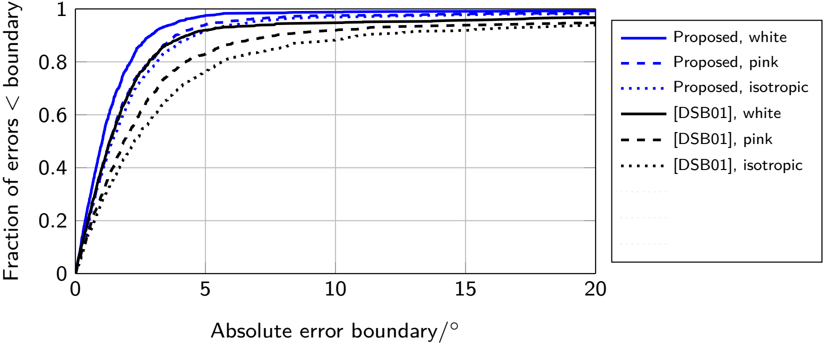
Performance for omnidirectional microphones



Conclusion

- Proposed algorithm achieved best performance in all conditions
- Smallest degradation

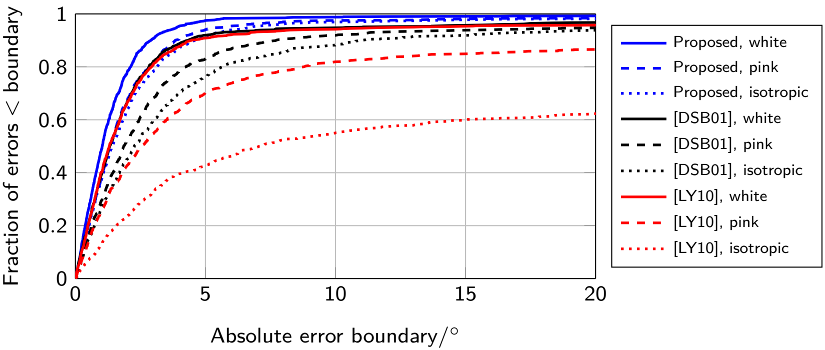
Performance for omnidirectional microphones



Conclusion

- Proposed algorithm achieved best performance in all conditions
- Smallest degradation

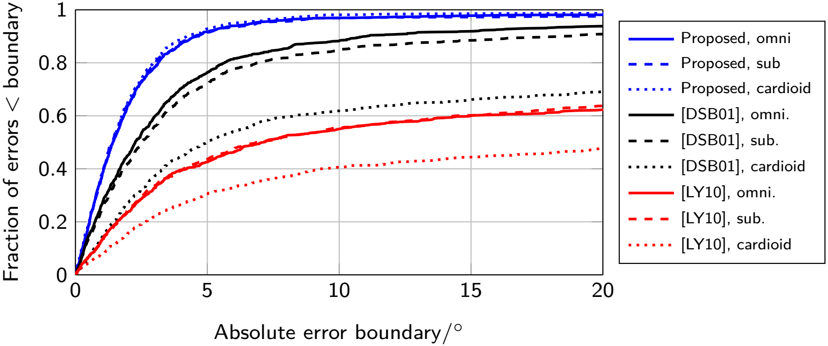
Performance for omnidirectional microphones

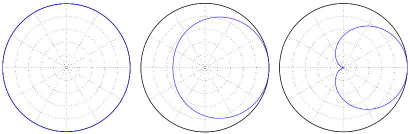


Conclusion

- Proposed algorithm achieved best performance in all conditions
- Smallest degradation

Performance with given sensor directivity - isotropic noise





omnidirectional subcardioid cardioid

Conclusion

- SRP-PHAT and [LY10]: unconsidered directivity pattern $\Rightarrow$ large degradation
- Proposed algorithm: no significant changes

Summary

Summary

- DoA (or source position) estimators:
  - Loesch
  - Watson Kernel (proposed)
- Statistically motivated (Complex Watson distribution)
- Exploits phase and level differences

  $\Rightarrow$ Avoid degradation due to pattern mismatch
- Even better performance for omnidirectional microphones
- Can be easily combined with speech presence probability (SPP) estimation

References

DiBiase, J. H. ; Silverman, H. F. ; Brandstein, M. S.:
Robust localization in reverberant rooms.
In: *Microphone Arrays*.
Springer, 2001, S. 157–180

Gaubitch, N. D. ; Martinez, J. ; Kleijn, W. B. ; Heusdens, Richard.:
On near-field beamforming with smartphone based ad-hoc microphone arrays.
In: *IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, S. 95–99

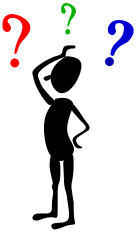
Ito, N. ; Araki, S. ; Nakatani, T.:
Permutation-free convolutive blind source separation via full-band clustering based on frequency-independent source presence priors.
In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, 3238–3242

Jafari, I. ; Togneri, R. ; Nordholm, S.:
On the use of the Watson mixture model for clustering-based under-determined blind source separation.
In: *Fifteenth Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2014, S. 988–992

Loesch, B. ; Yang, B.:
Blind source separation based on time-frequency sparseness in the presence of spatial aliasing.
In: *Latent Variable Analysis and Signal Separation*.
Springer, 2010, S. 1–8

Mardia, K. V. ; Dryden, I. L.:
The complex Watson distribution and shape analysis.
In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (1999), Nr. 4, S. 913–926

Tran Vu, D. H. ; Haeb-Umbach, R.:
Blind speech separation employing directional statistics in an expectation maximization framework.
In: *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, S. 241–244



Thank you for your attention!

Questions?

This work has been supported by Deutsche Forschungsgemeinschaft (DFG) under contract no. Ha3455/7-2.

Florian Jacob

University of Paderborn
Department of Communications Engineering

jacob@nt.uni-paderborn.de
nt.uni-paderborn.de