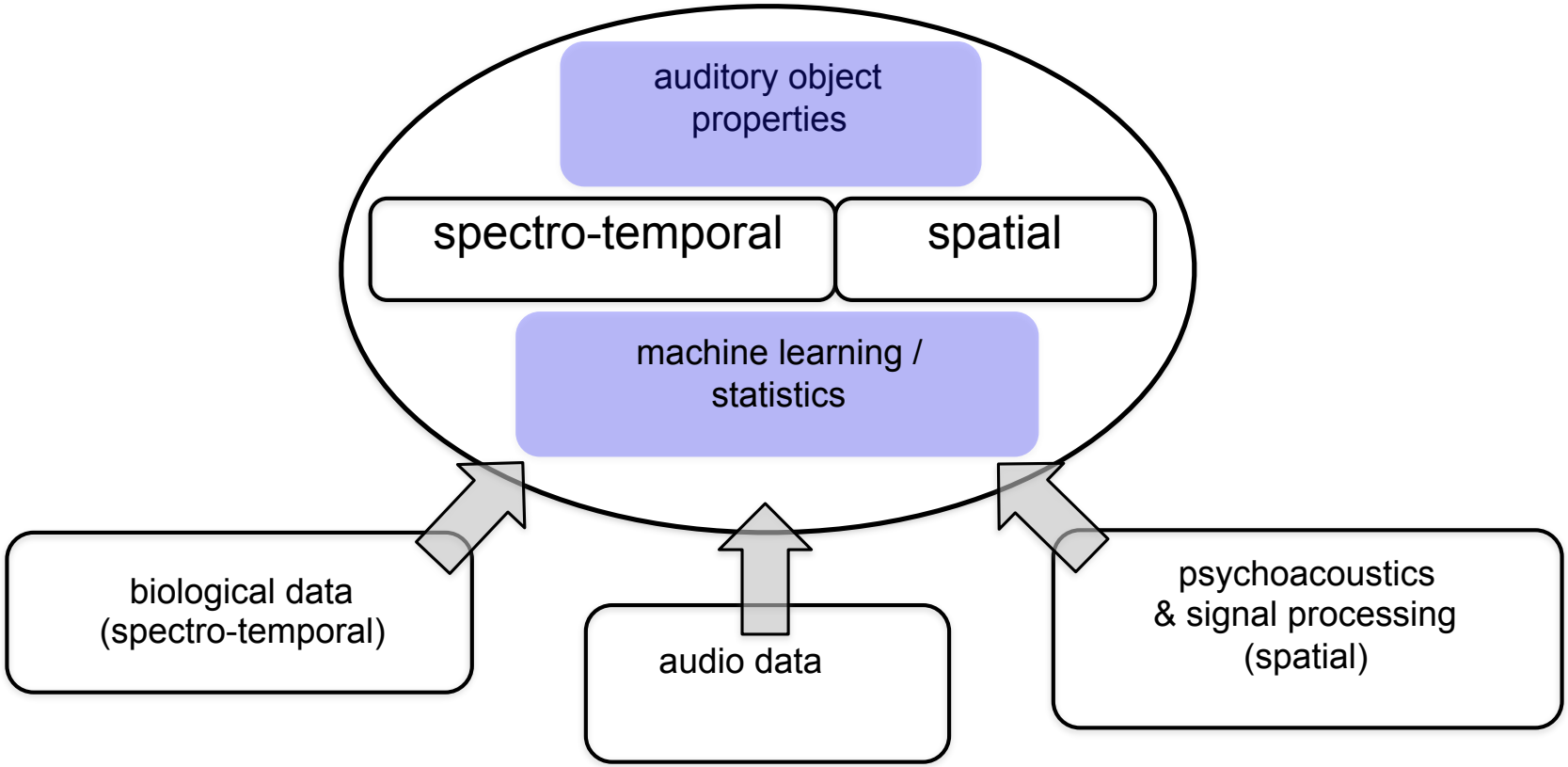
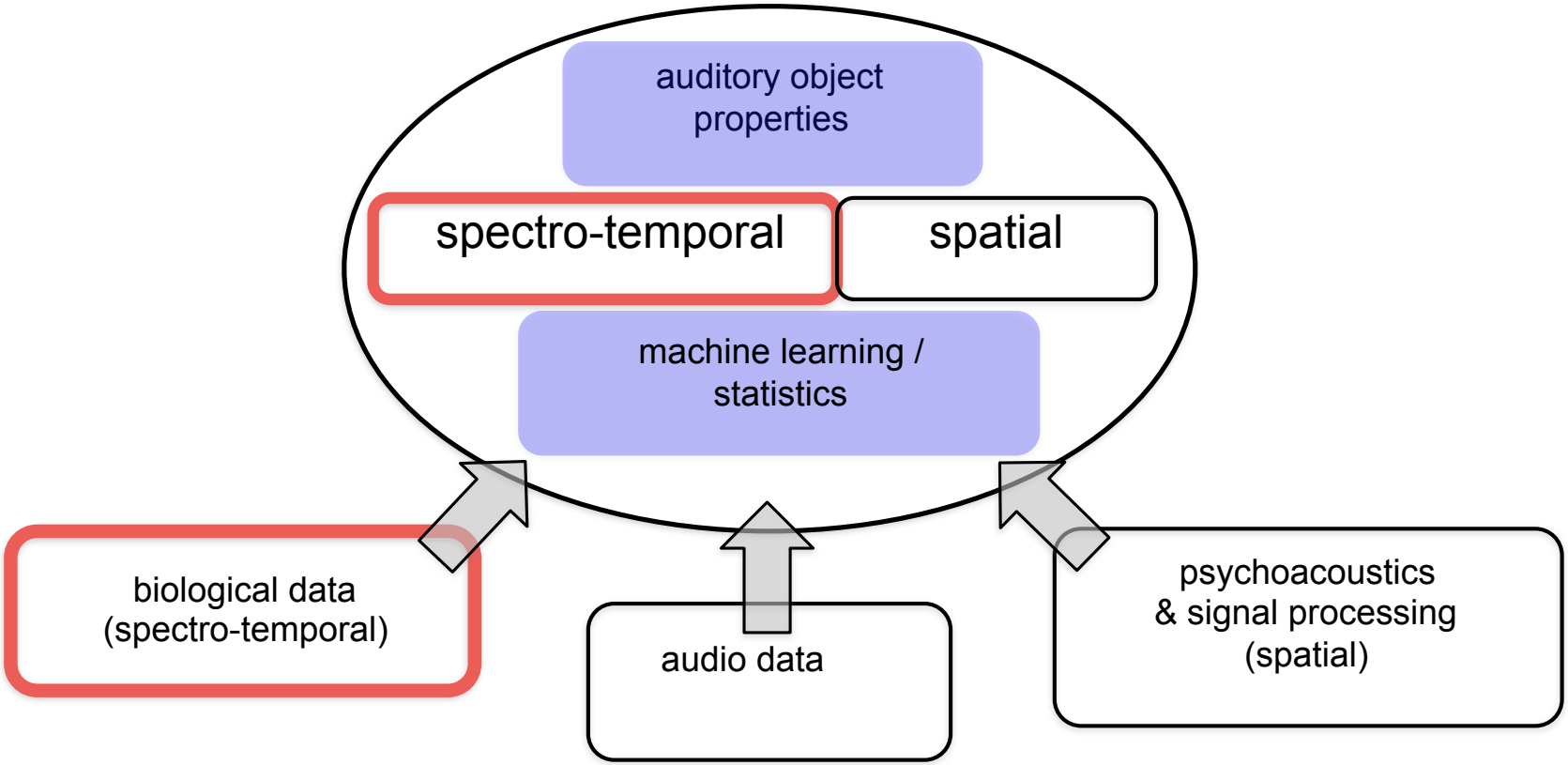
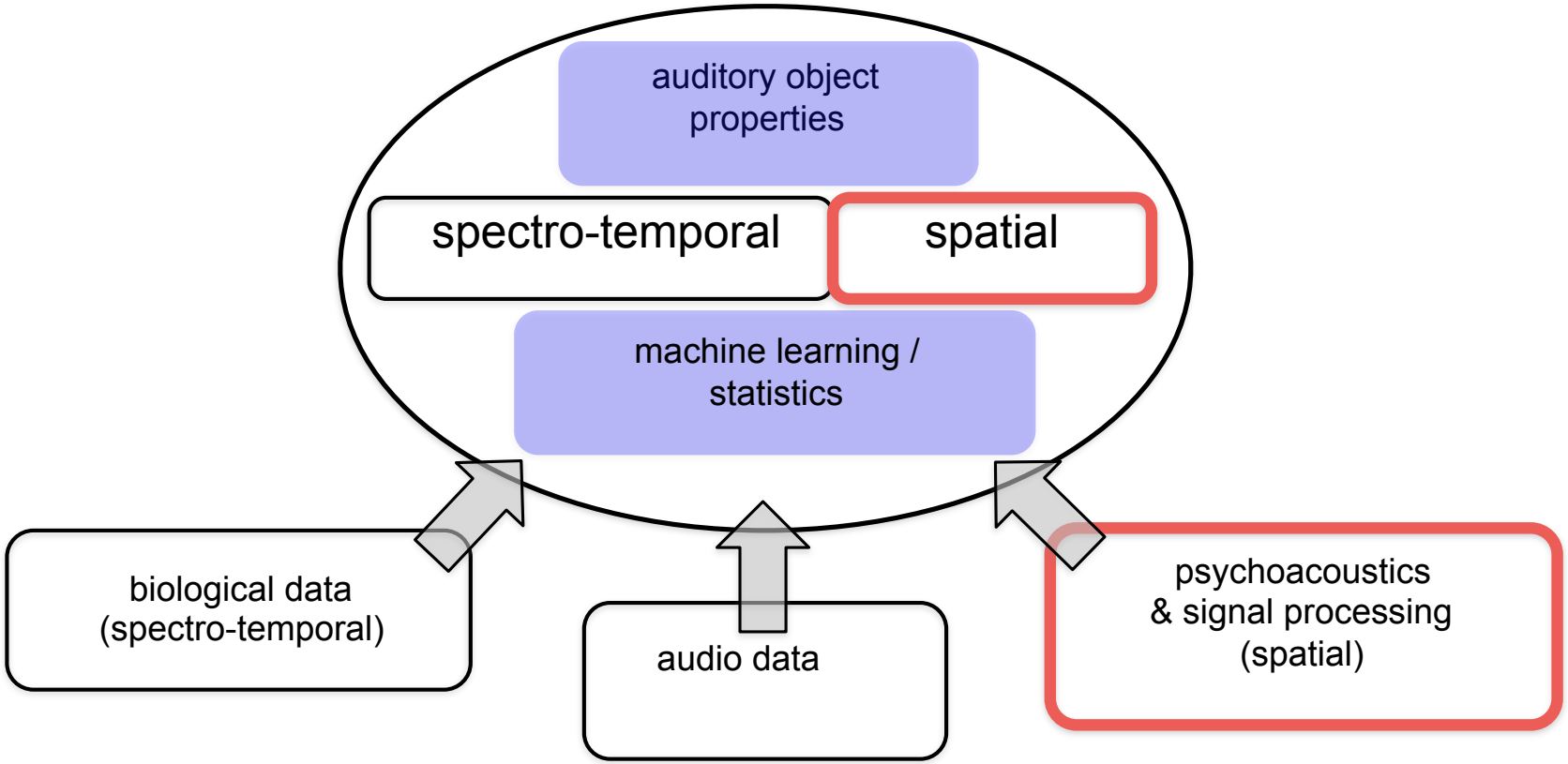
Sound source localization – from low-level sensor signals to mid-level representations, and back

Jörn Anemüller, Niko Moritz, Hendrik Kayser

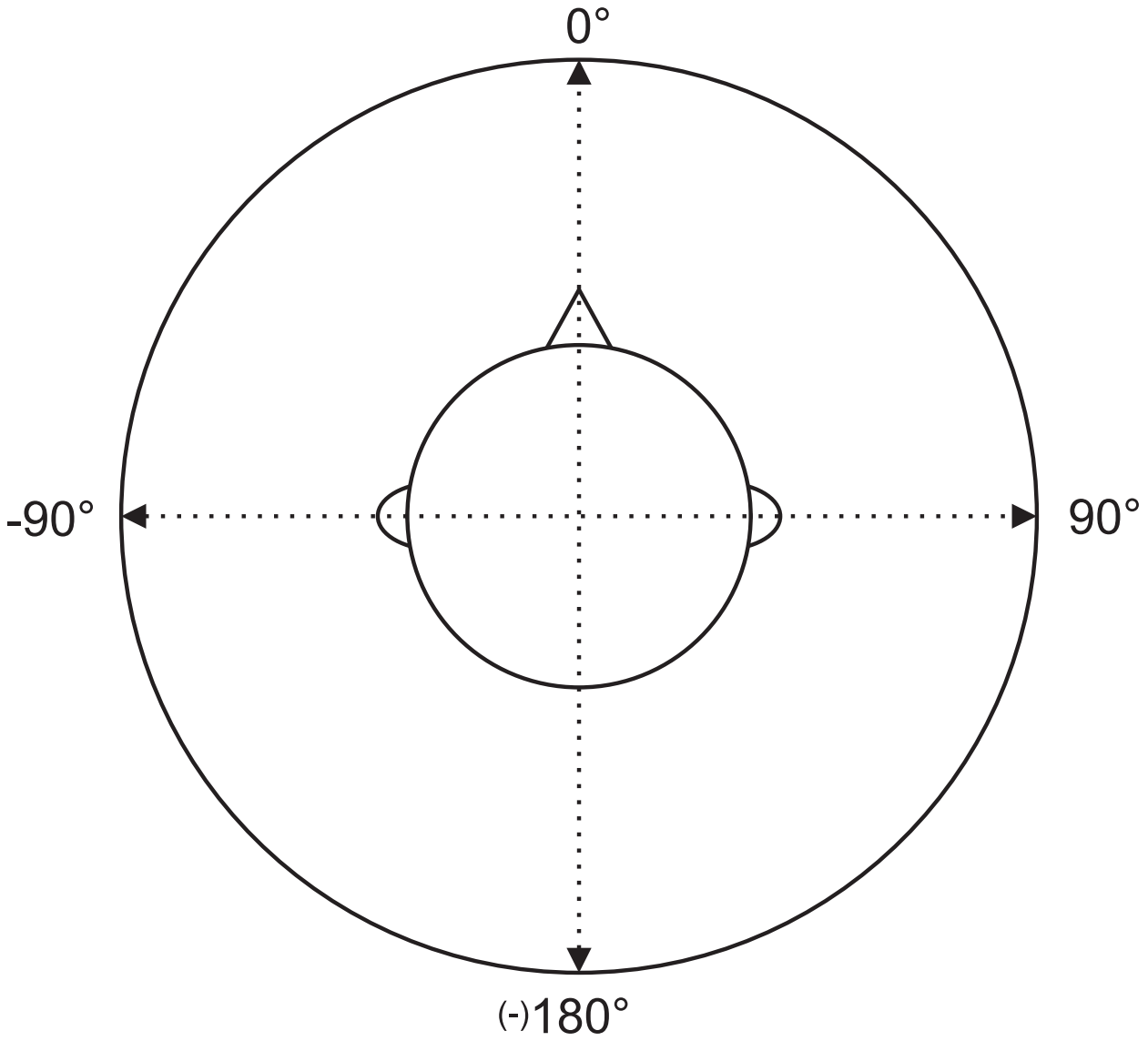
Computational Audition Group
Medical Physics Section and Cluster of Excellence Hearing4all
Carl von Ossietzky Universität Oldenburg


auditory object properties
spectro-temporal
spatial
machine learning / statistics
biological data (spectro-temporal)
audio data
psychoacoustics & signal processing (spatial)

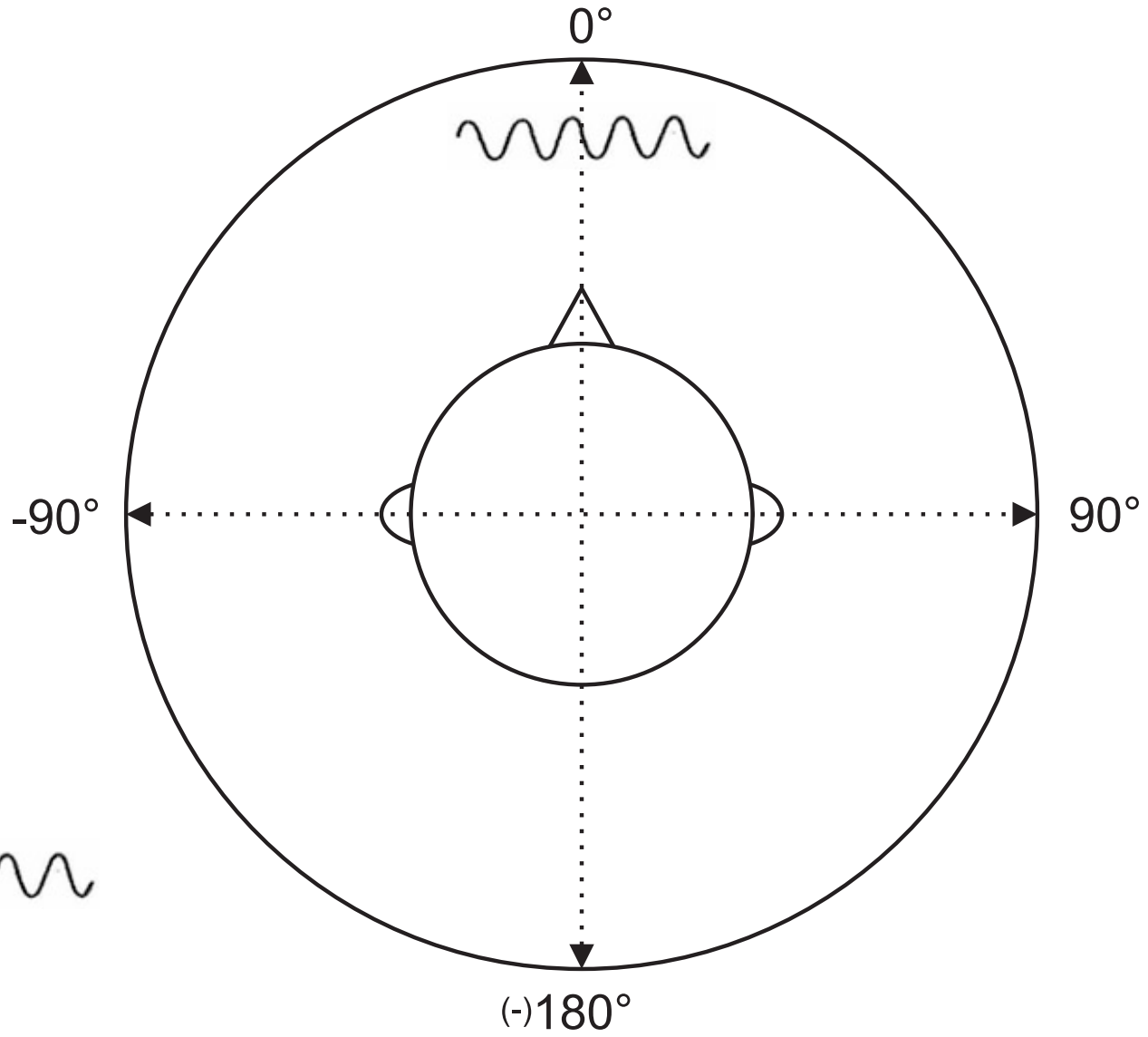

auditory object properties
spectro-temporal
spatial
machine learning / statistics
biological data (spectro-temporal)
audio data
psychoacoustics & signal processing (spatial)


auditory object properties
spectro-temporal
spatial
machine learning / statistics
biological data (spectro-temporal)
audio data
psychoacoustics & signal processing (spatial)

Spatial sound perception

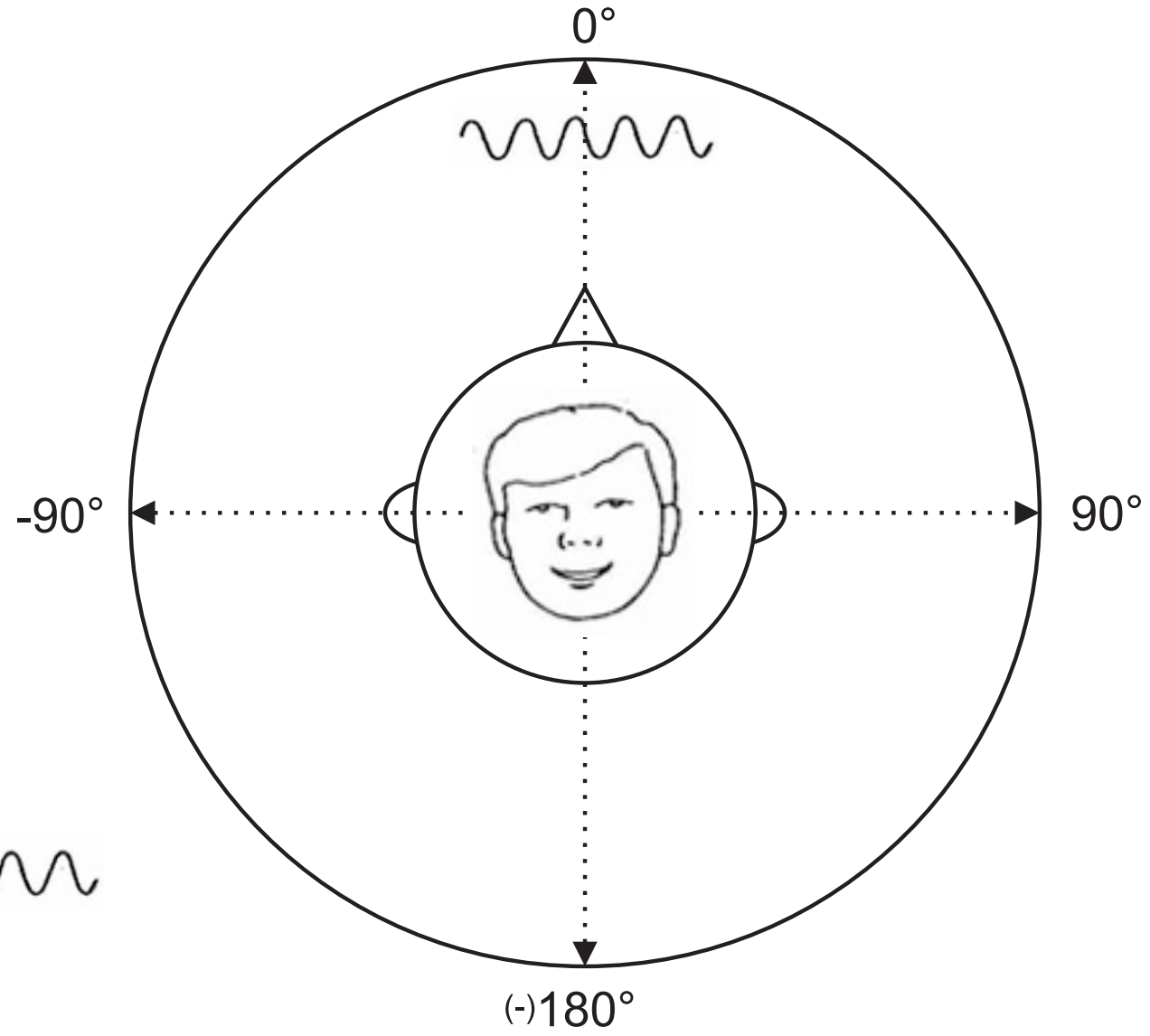


Spatial sound perception



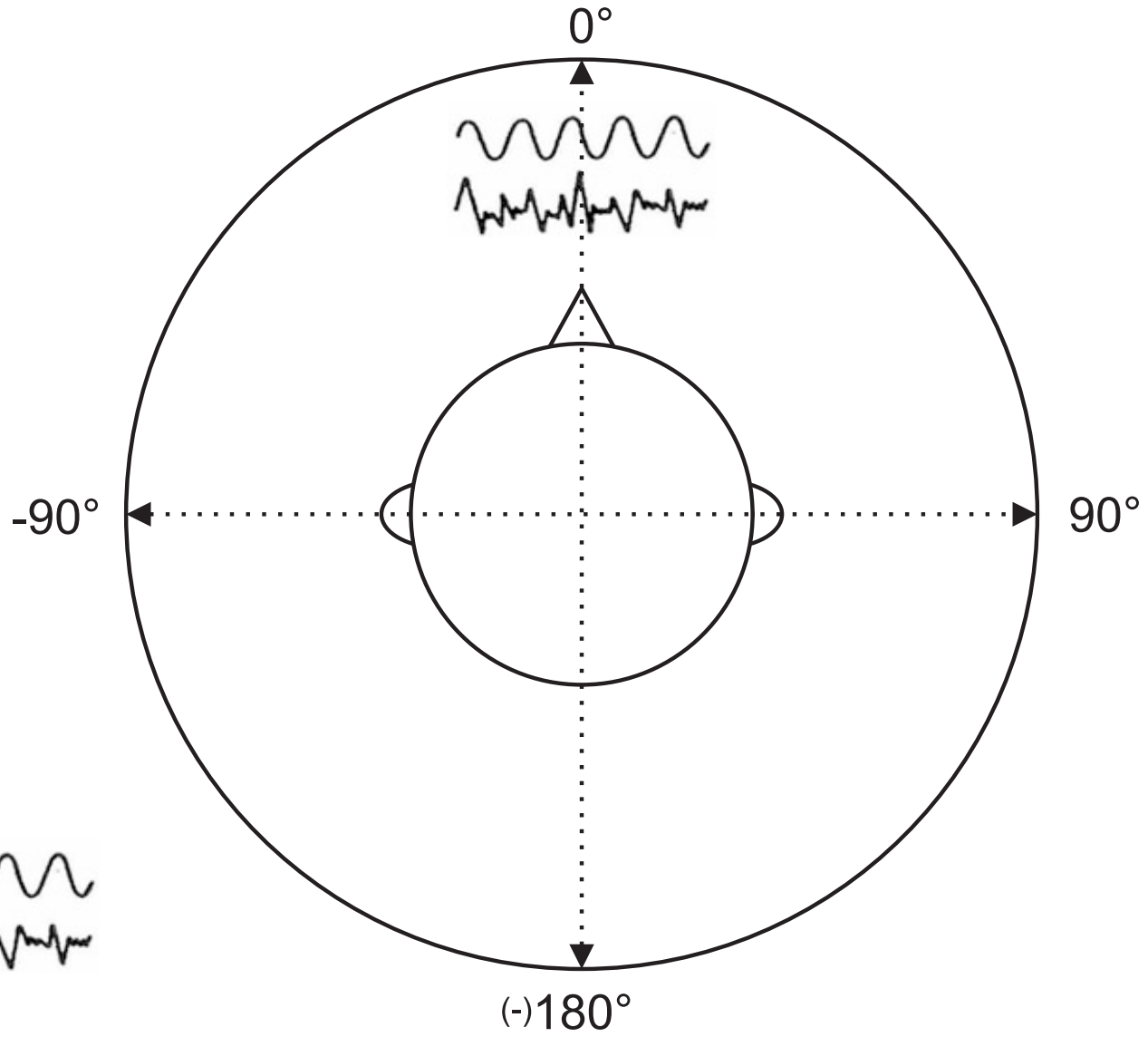
signal 

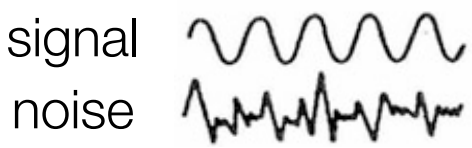
Spatial sound perception



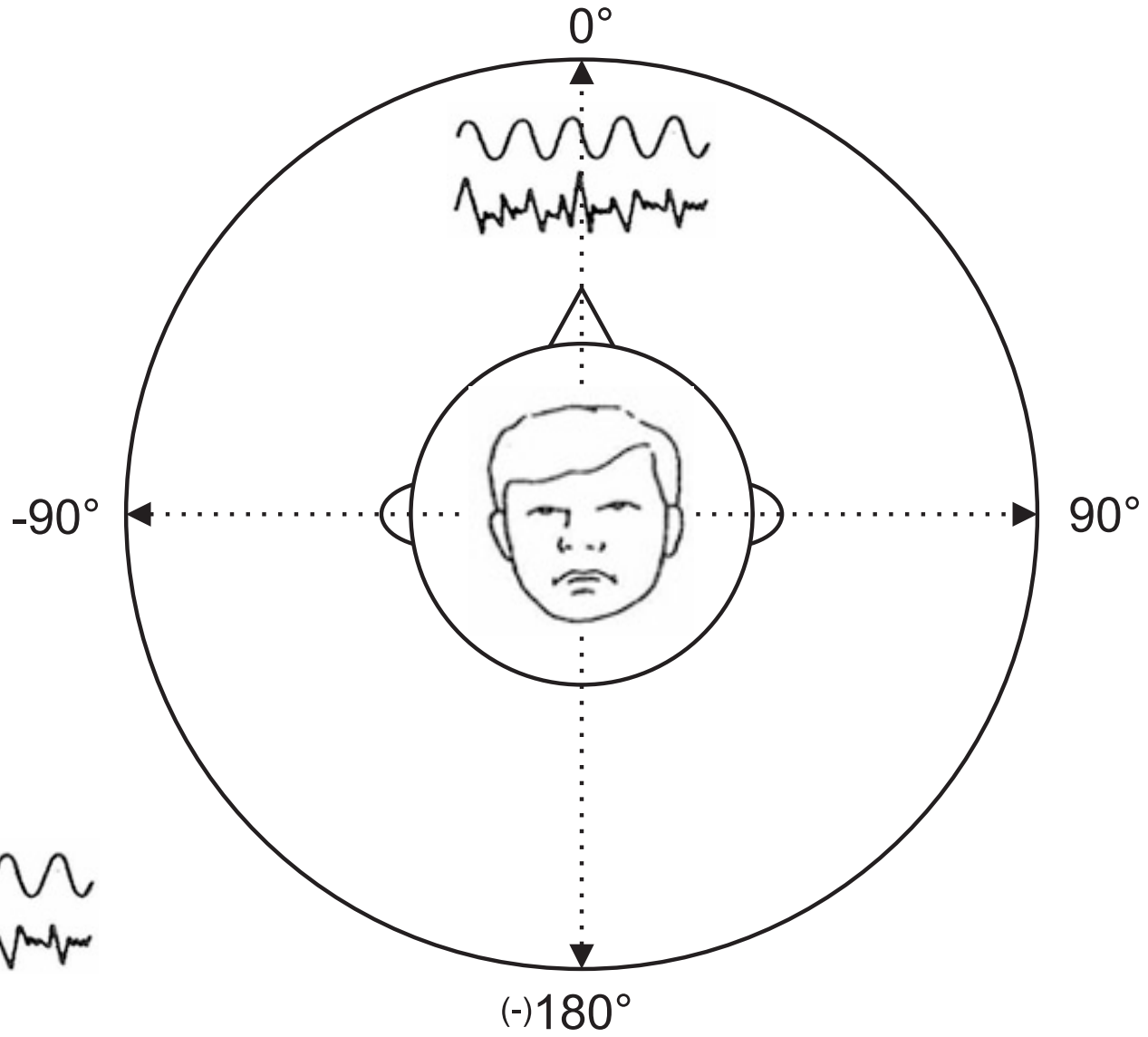
signal

Spatial sound perception



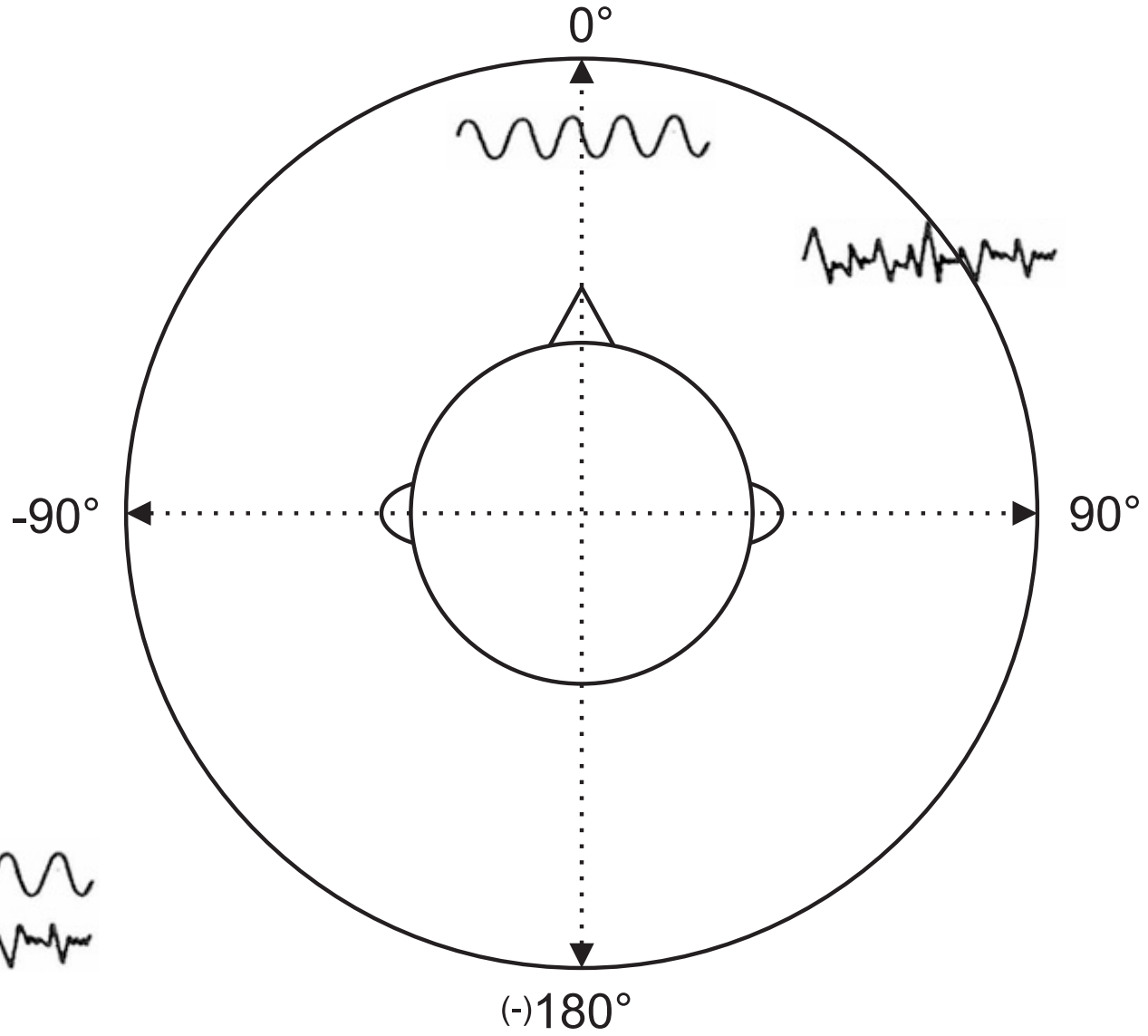


Spatial sound perception



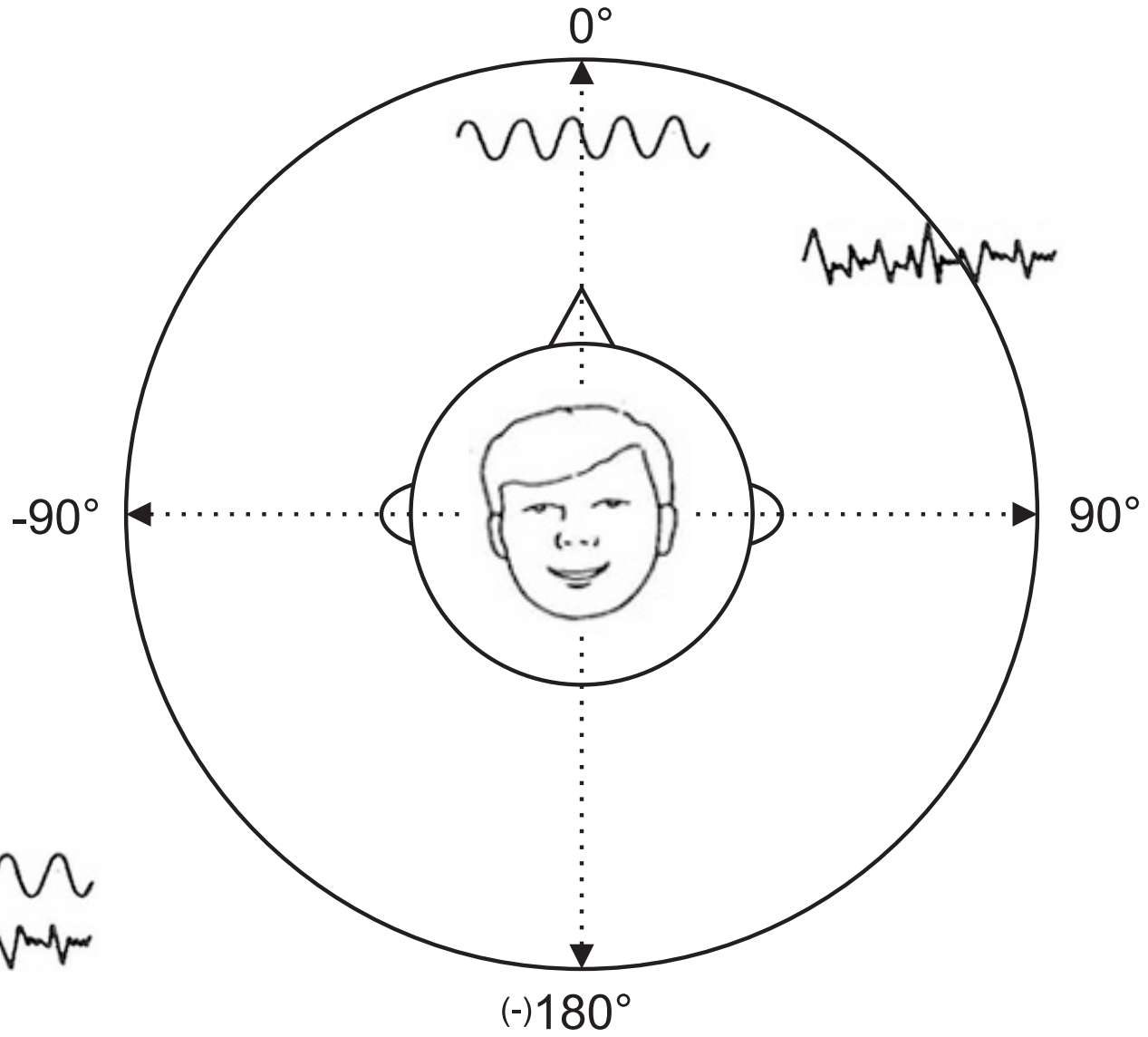
signal
noise

Spatial sound perception



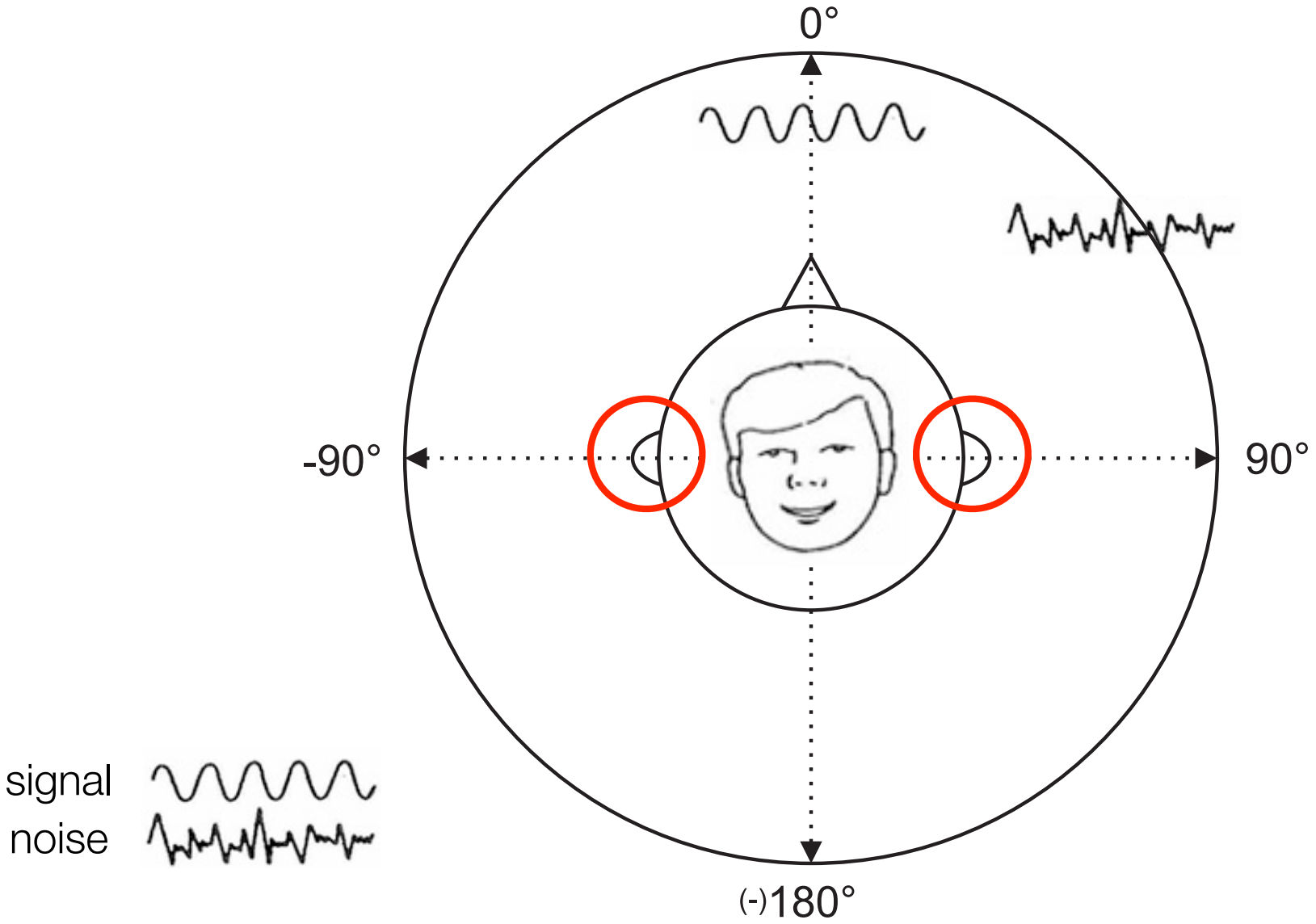
signal
noise

Spatial sound perception

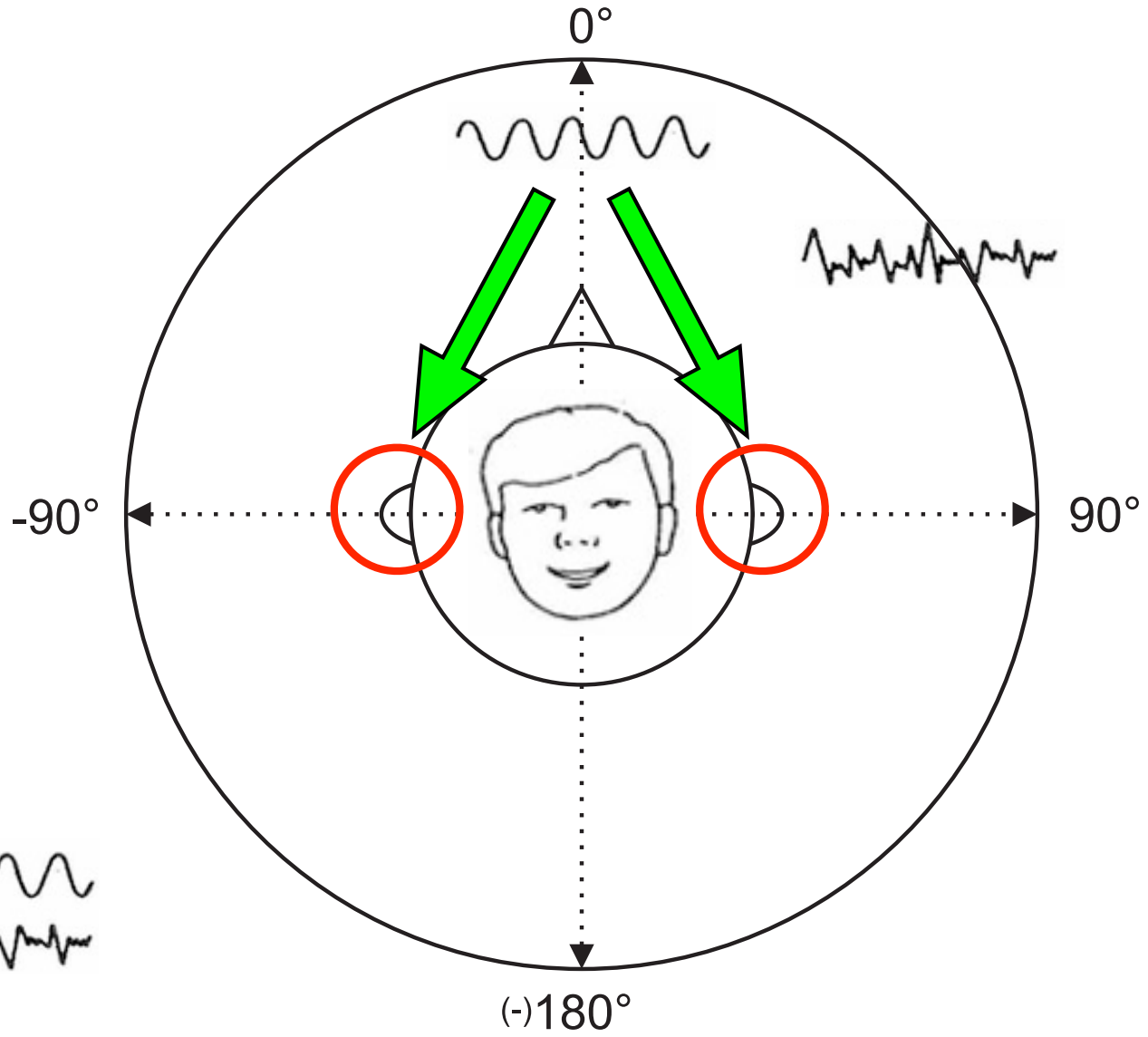


signal
noise

Spatial sound perception

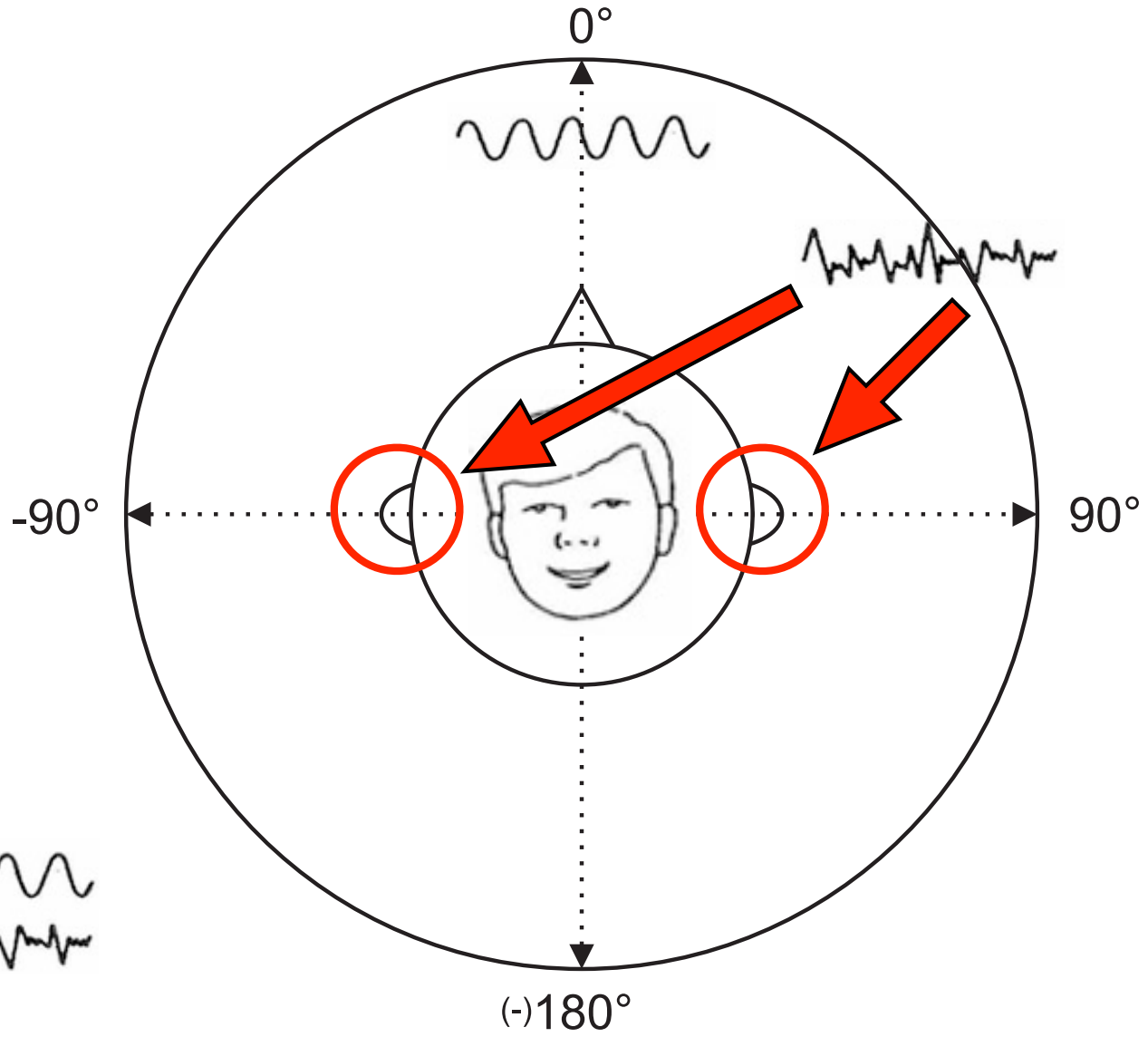


Spatial sound perception



signal
noise

Spatial sound perception



signal
noise

“Hard-wired” and “learned” acoustic features for sound localization

Which features permit robust localization?

- partly “hard-wired”
  - mechanisms evolved over long time
- partly “long-time learned”
  - stimulus statistics important
- partly “adapted on-the-fly”
  - new environment, keep learned info

“Stop! Stop! What’s that sound? What’s that sound?”

“Hard-wired” and “learned” acoustic features for sound localization

Which features permit robust localization?

partly “hard-wired”

mechanisms evolved over long time

partly “long-time learned”

stimulus statistics important

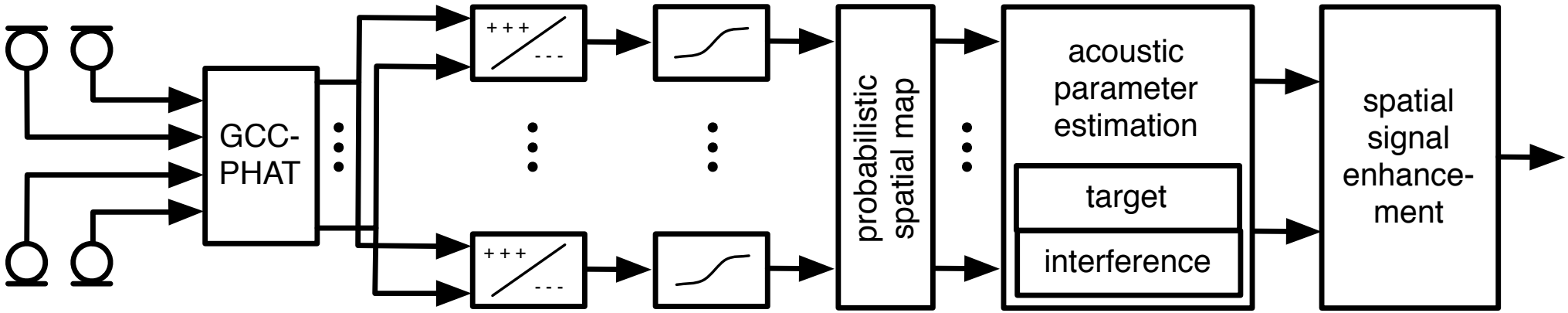
partly “adapted on-the-fly”

new environment, keep learned info

“Stop! Stop! What’s that sound? What’s that sound?”

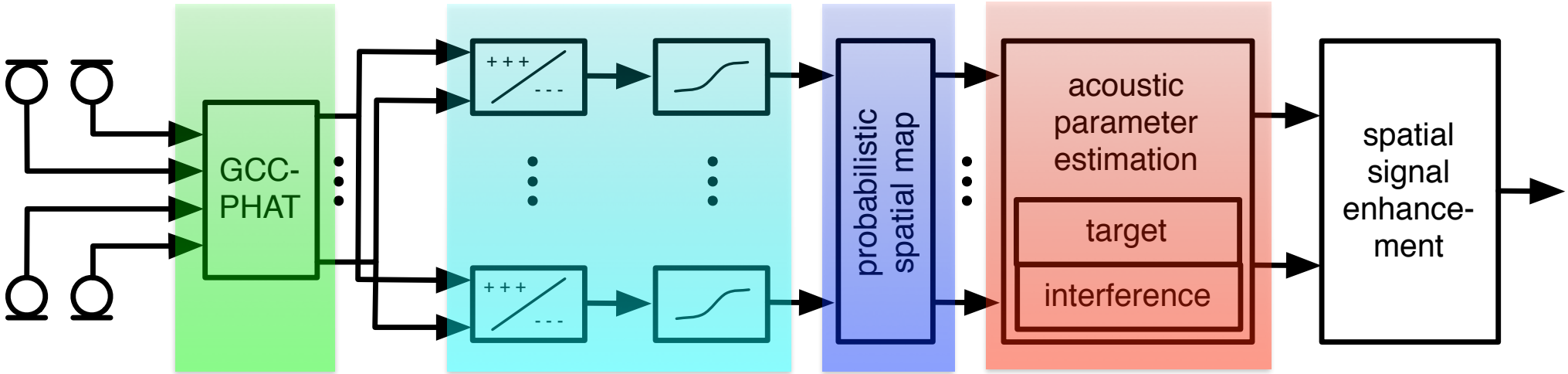
System overview: Localization and signal enhancement

Goal: Spatial source localization and enhancement with robust performance and fast computation



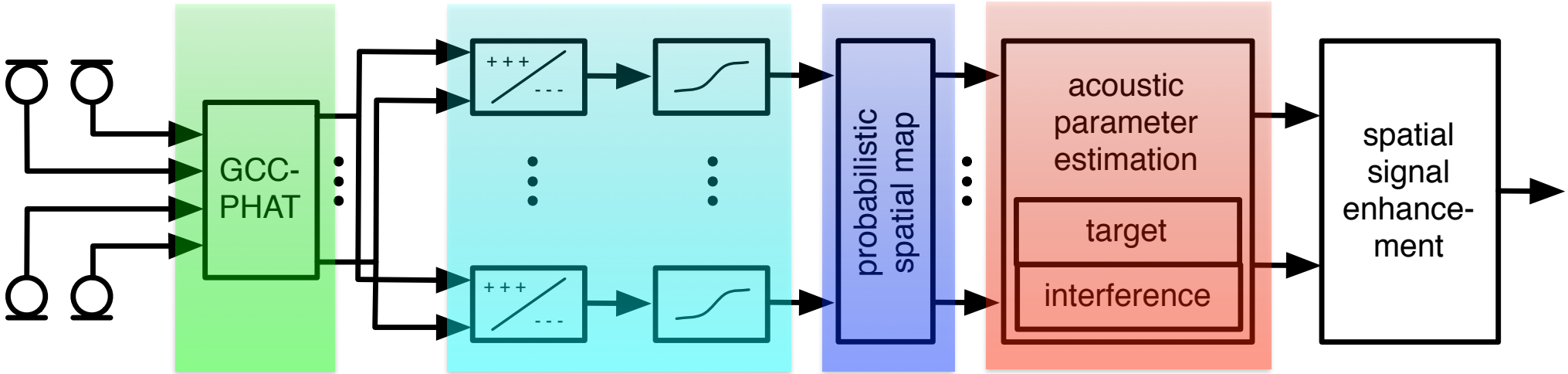
System overview: Localization and signal enhancement

Goal: Spatial source localization and enhancement with robust performance and fast computation



System overview: Localization and signal enhancement

Goal: Spatial source localization and enhancement with robust performance and fast computation



Hard-coded features

Generalised cross-correlation

Supervised learning (long-term)

Bank of SVMs

Sparse representation

Probabilistic localisation map

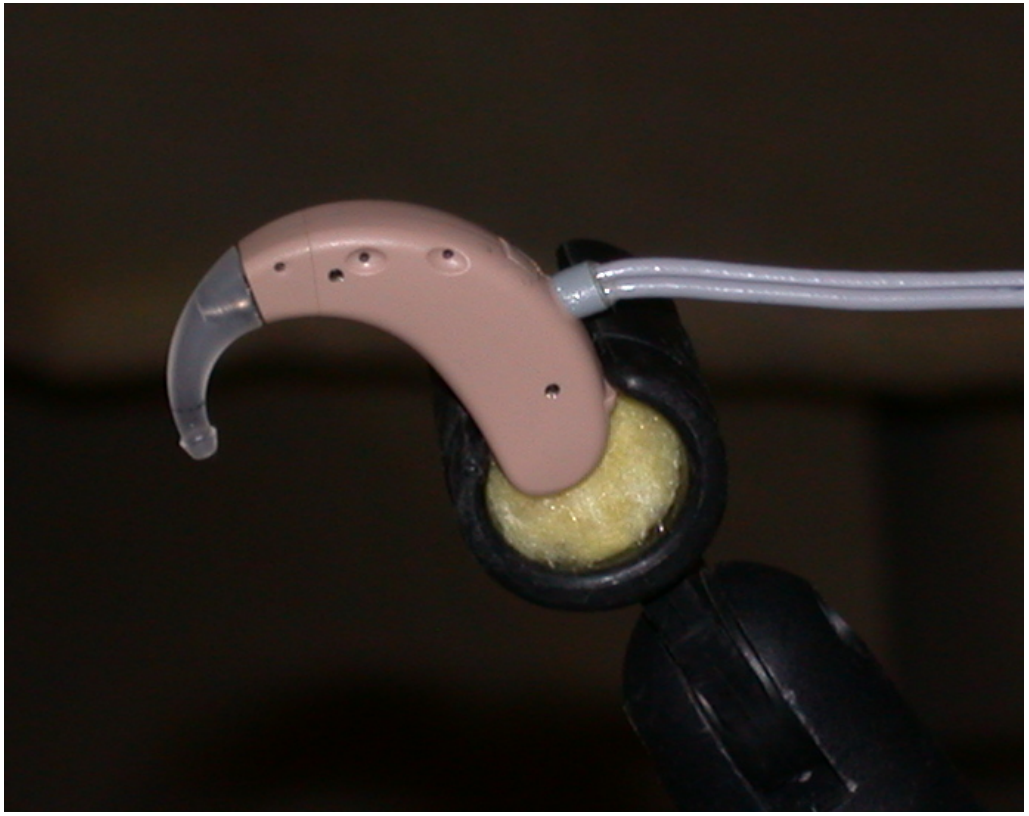
Acoustic parameter (re-)estimation (fast adaptation)

Anemüller, Kayser (2017)

Microphone Geometry


TEAC
TEAC

Microphone Geometry



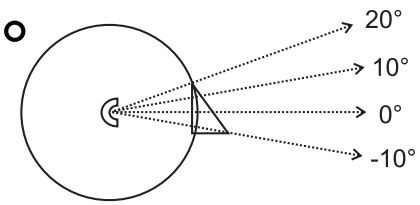
- 6-channel hearing aid microphone array:
  - 3-channels on each ear
- Mounted on a head an torso simulator

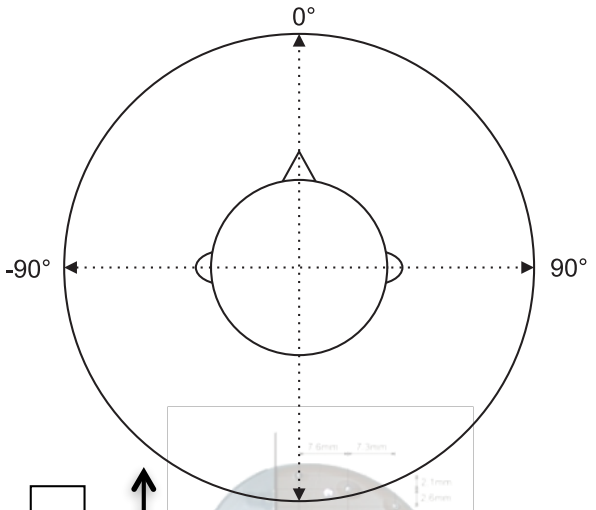
Anechoic Environment

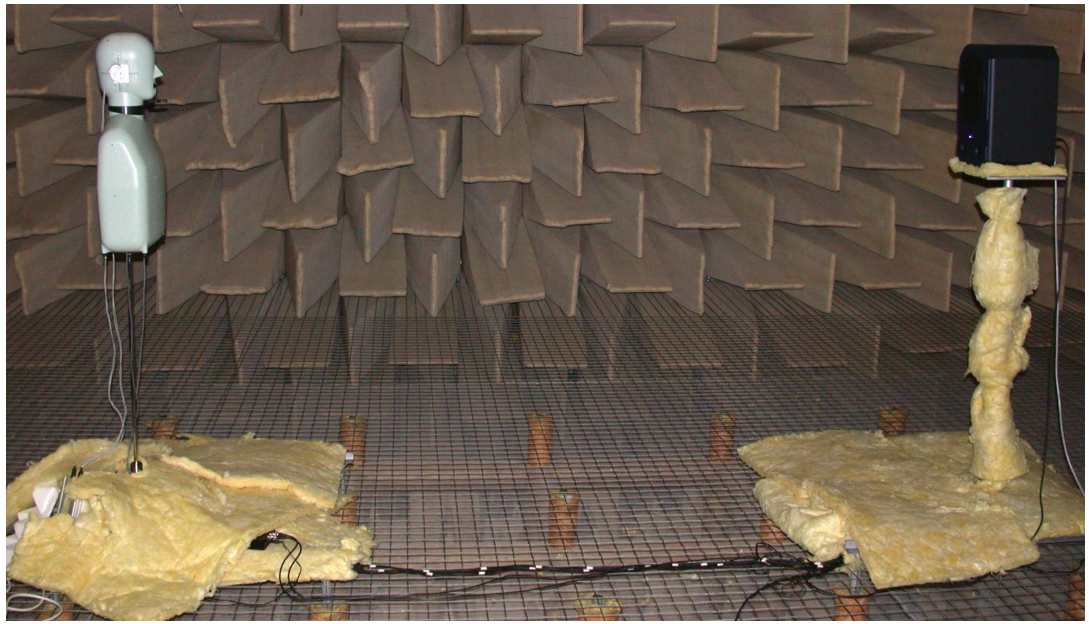
Azimuth: 0°, 5°, ... , 180°

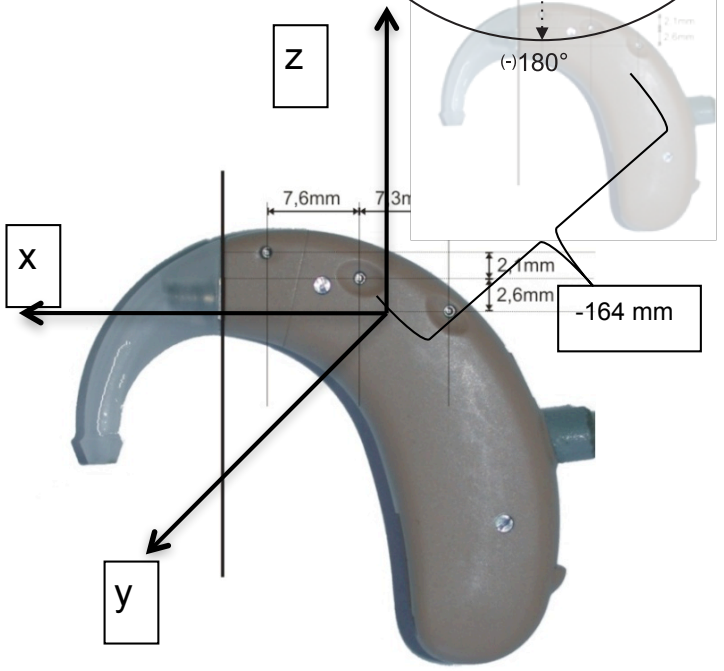
Elevation: -10°, 0°, 10°, 20°

Distance: 0.8m, 3m





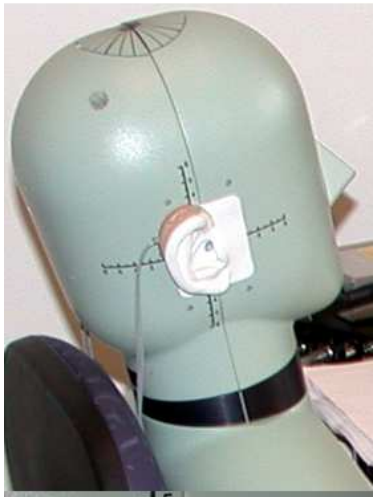


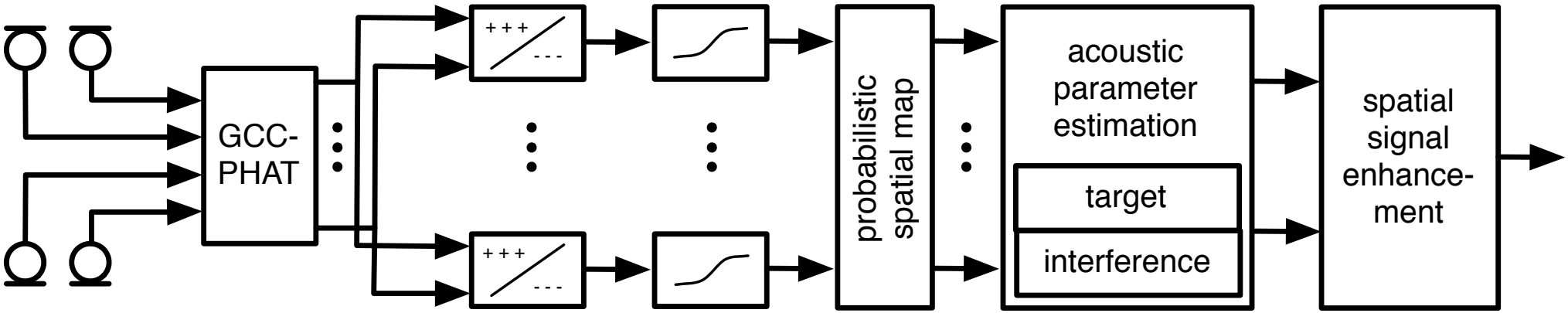


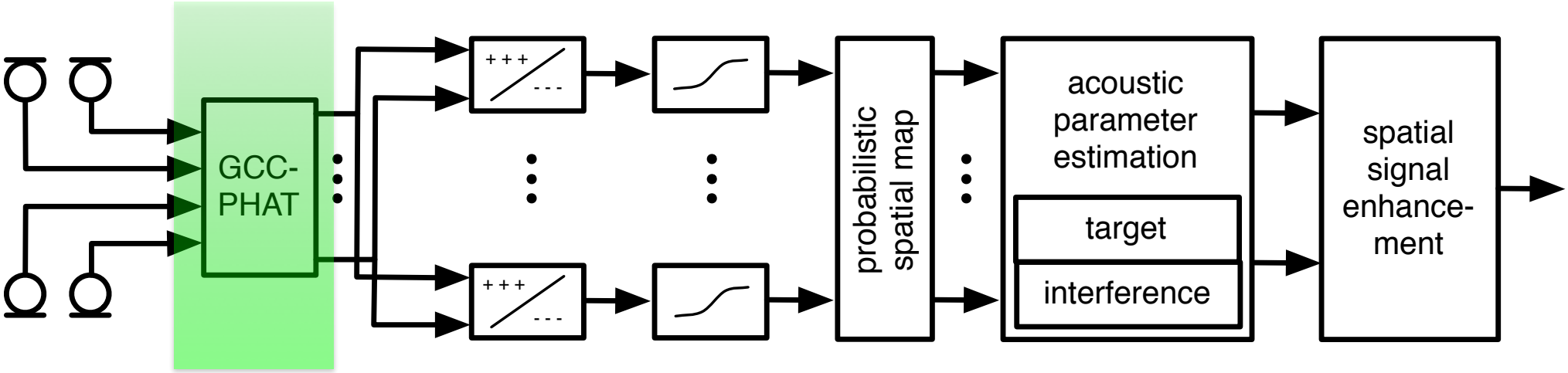
Kayser, Ewert, Anemüller, Rohdenburg, Hohmann, Kollmeier (2009)

Echoic Environments

- Office
  - Azimuth: 0°, 5°, ... , 180°
  - Distance: 1m
- Several settings indoors and outdoors
  - Office, courtyard, cafeteria




GCC-
PHAT
+ + +
- - -
+ + +
- - -
probabilistic
spatial map
acoustic
parameter
estimation
target
interference
spatial
signal
enhance-
ment



Hard-coded features

Generalised cross-correlation

GCC-PHAT correlation features

Generalized cross-correlation with phase-transform

- “Hard-wired”
- Only phase-differences accounted for
- Weighting towards higher frequency (low energy) spectral bands

$$\rho_{ij}(t,\tau) = \mathcal{IFFT}\left\{ \frac{X_i(t,f)}{|X_i(t,f)|} \frac{X_j^*(t,f)}{|X_j(t,f)|} \right\}$$

GCC-PHAT correlation features

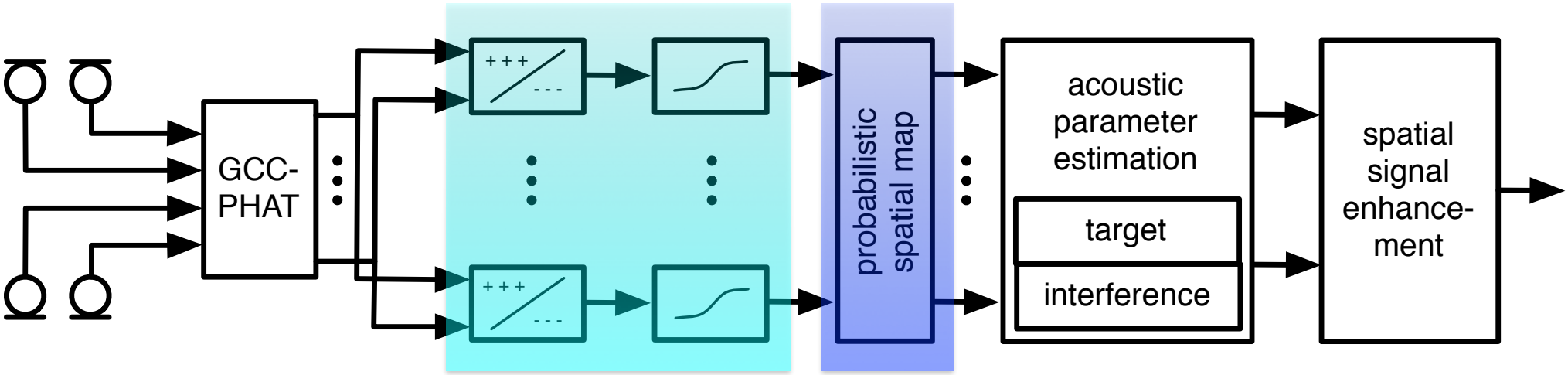
Generalized cross-correlation with phase-transform

"Hard-wired"

Only phase-differences accounted for

Weighting towards higher frequency (low energy) spectral bands

$$\rho_{ij}(t,\tau) = \mathcal{IFFT}\left\{ \frac{X_i(t,f)}{|X_i(t,f)|} \frac{X_j^*(t,f)}{|X_j(t,f)|} \right\}$$



Supervised learning (long-term)

Bank of SVMs

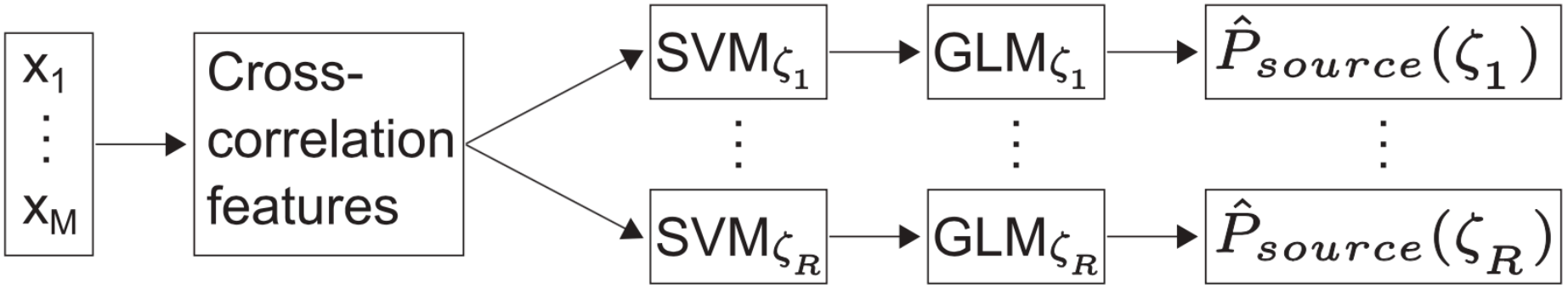
Sparse representation

Probabilistic localisation map

Learning-based approach to acoustic source localisation

Train one SVM per possible source position

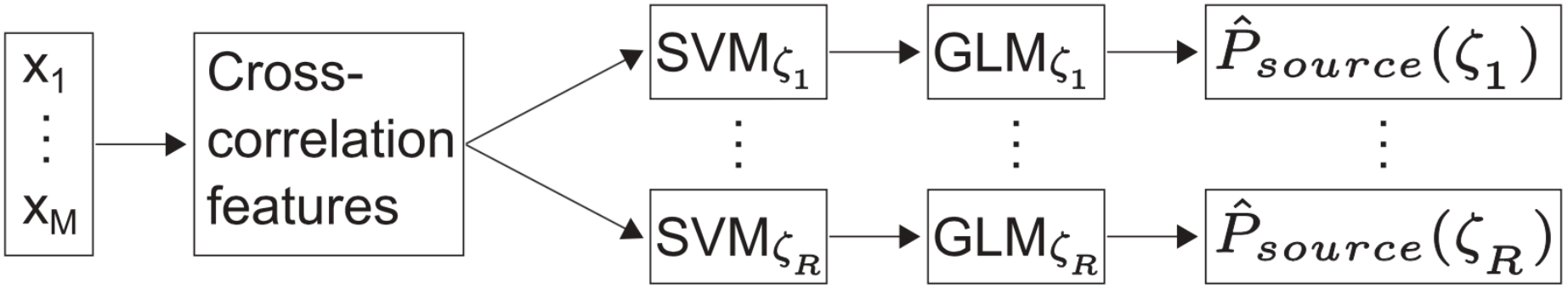
"long-term learned"



Learning-based approach to acoustic source localisation

Train one SVM per possible source position

"long-term learned"



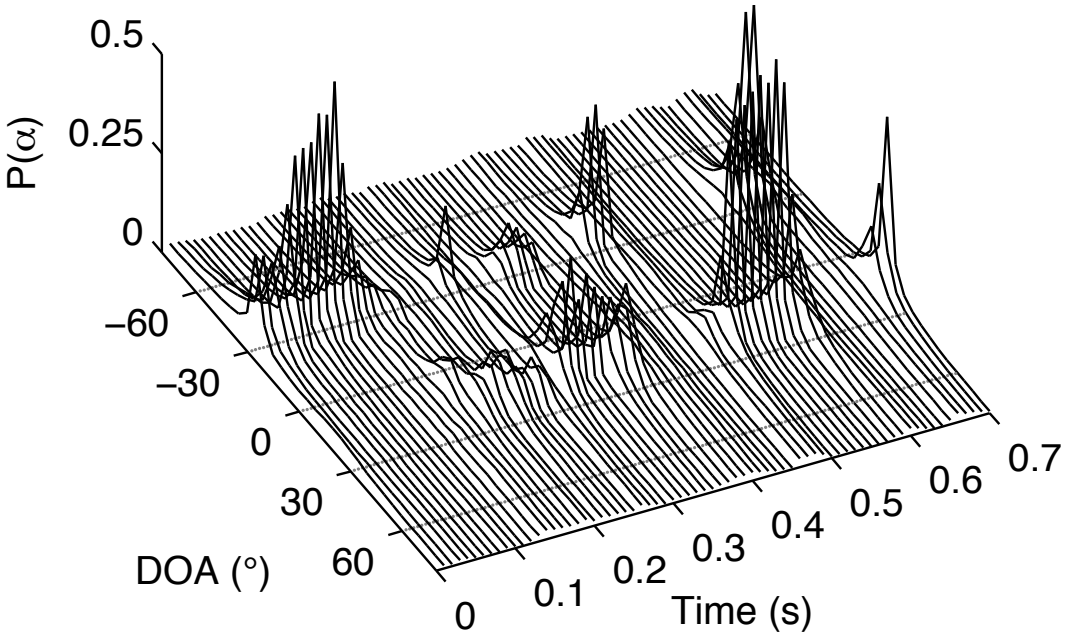
SVM Class.: $$d(\zeta_r) = \langle \boldsymbol{w}(\zeta_r), \boldsymbol{\phi} \rangle + b(\zeta_r)$$

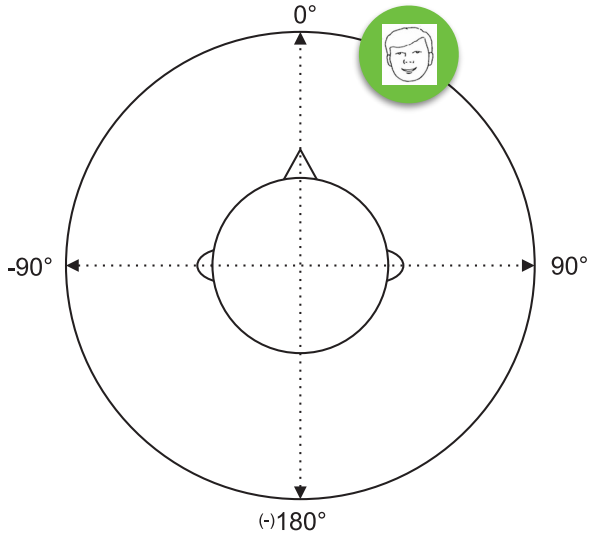
GLM prob.: $$\hat{P}_{source}(\zeta_r) = \hat{P}(d(\zeta_r)) = \frac{1}{1 + e^{-(\beta_1(\zeta_r) + \beta_2(\zeta_r) d(\zeta_r))}}$$

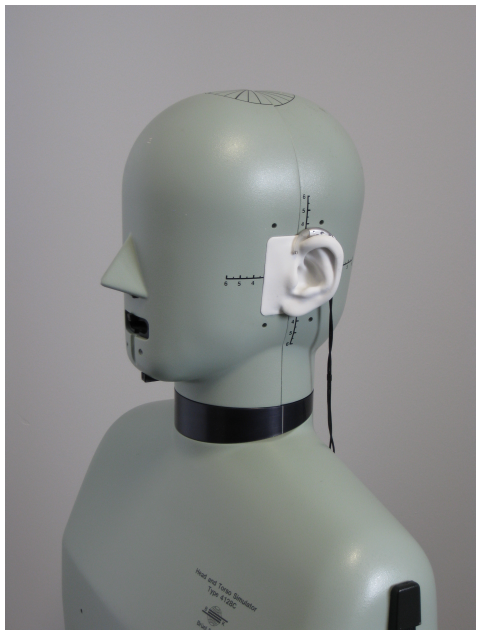
Max. direction: $$\hat{\zeta} = \underset{\zeta_r}{\operatorname{argmax}} \left[\hat{P}_{source}(\zeta_r) \right]$$

Probabilistic spatial map with HRTF setup

Probabilistic spatial localization map







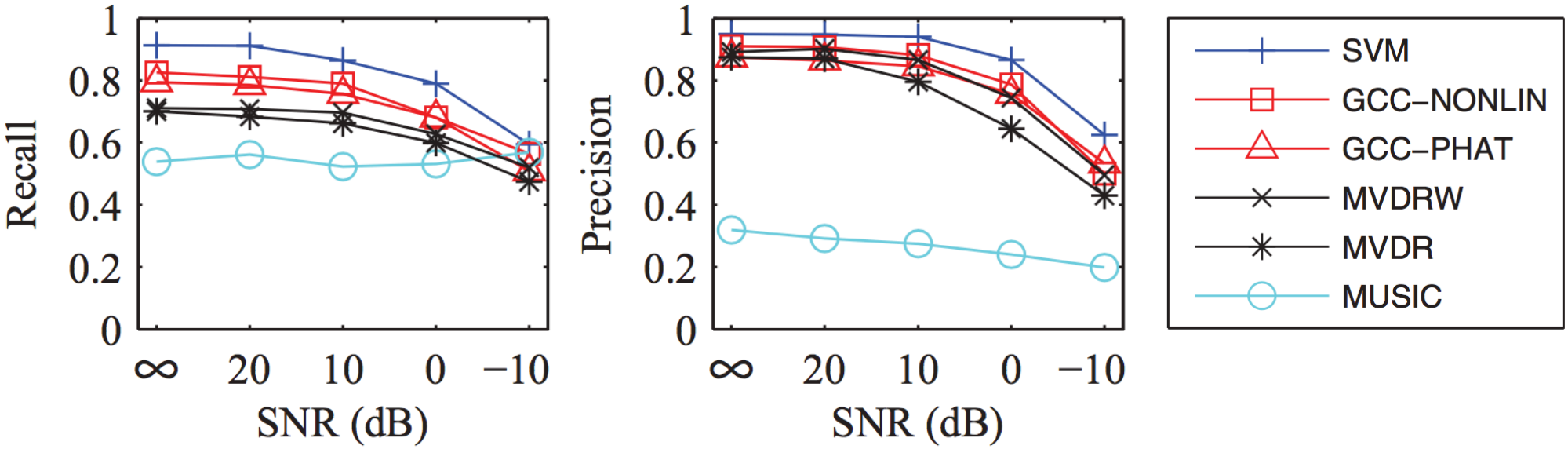
A-posteriori speech probabilities

Highly kurtotic sparse representation

Results: Localization

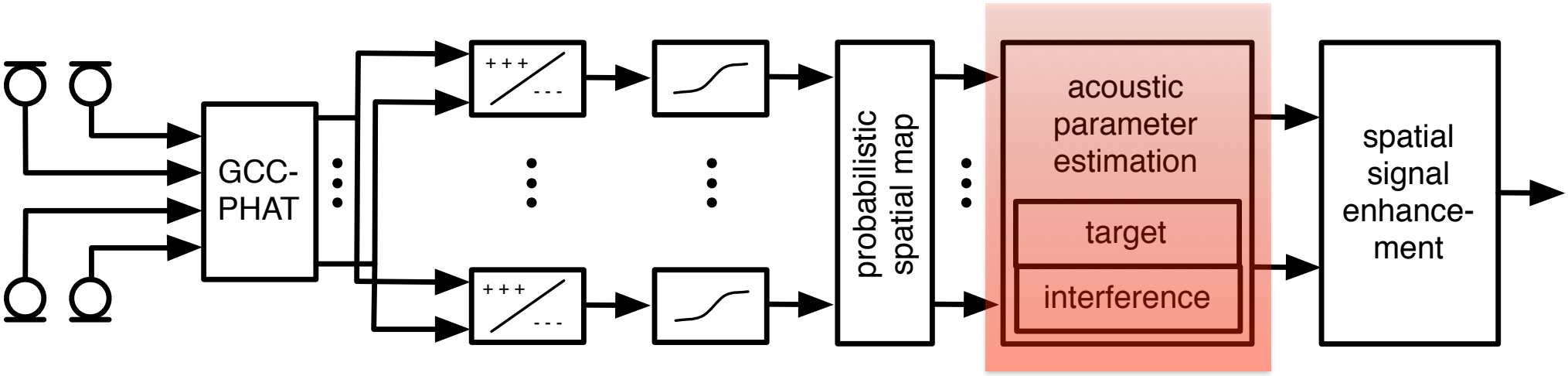
	Training data	Test data
Environment	anechoic	office, courtyard, cafeteria
Reverberation	< 50 ms	300 ms, 900 ms, 1300 ms
# Sources	1	2, 3 ,4
Noise	speech-shaped, diffuse	on-site recording
SNR (dB)	{-10, 0, 10, 20 }	-10, 0, 10, 20

Results: Localization



	Training data	Test data
Environment	anechoic	office, courtyard, cafeteria
Reverberation	< 50 ms	300 ms, 900 ms, 1300 ms
# Sources	1	2, 3 ,4
Noise	speech-shaped, diffuse	on-site recording
SNR (dB)	{-10, 0, 10, 20 }	-10, 0, 10, 20

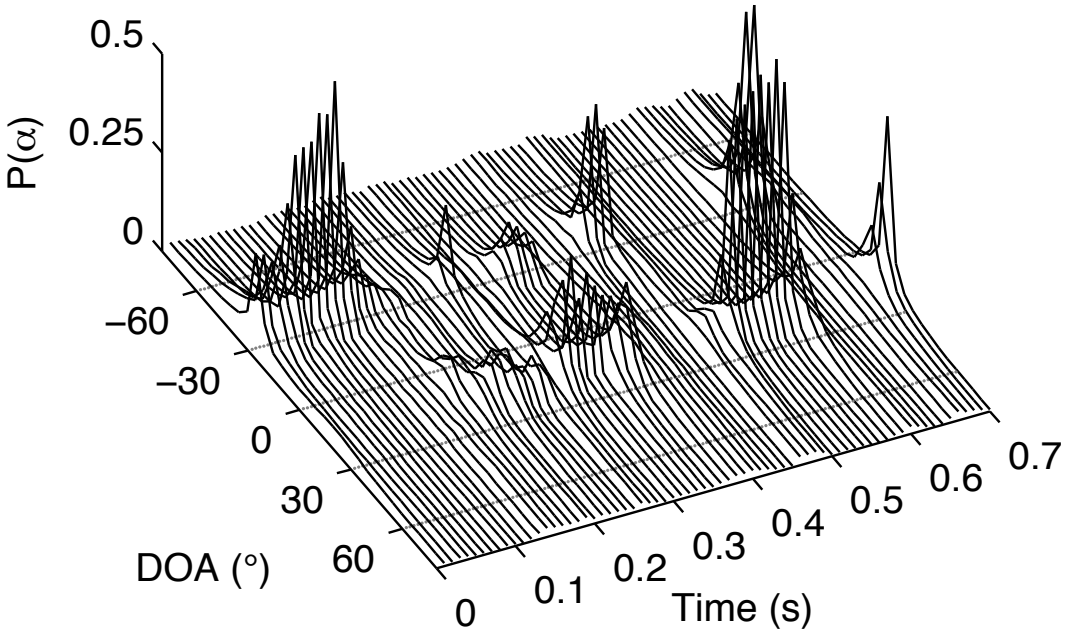
[Kayer, Anemüller, IWAENC 2014]

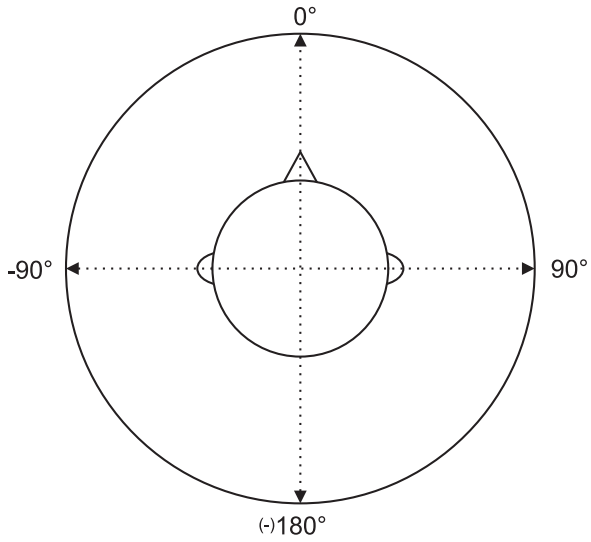


Acoustic parameter (re-)estimation (fast adaptation)

Estimation of *source* covariance matrix

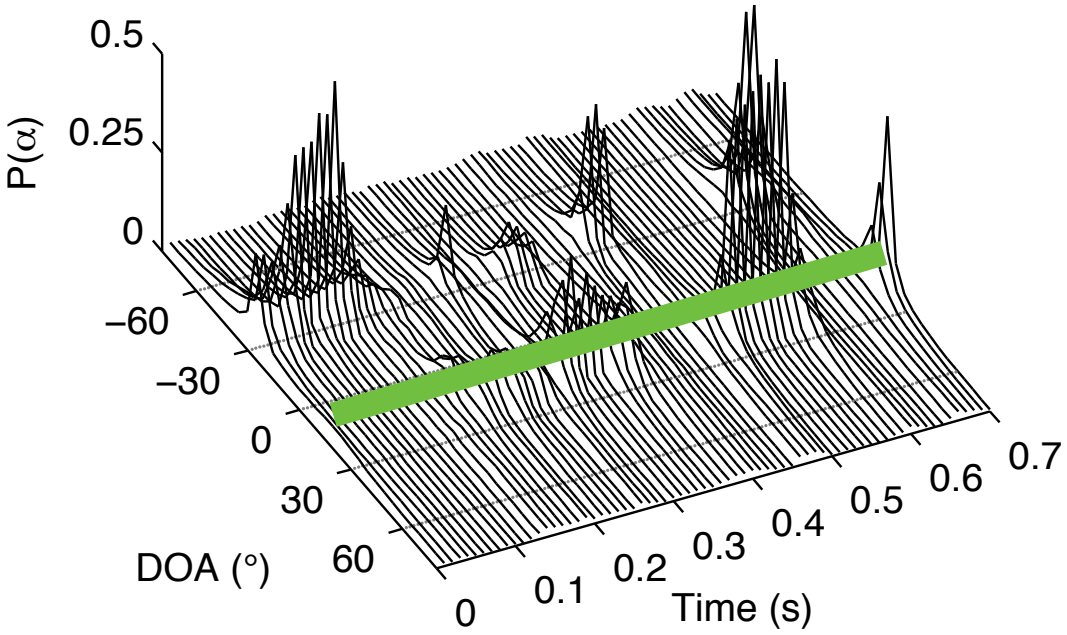
Probabilistic spatial localization map



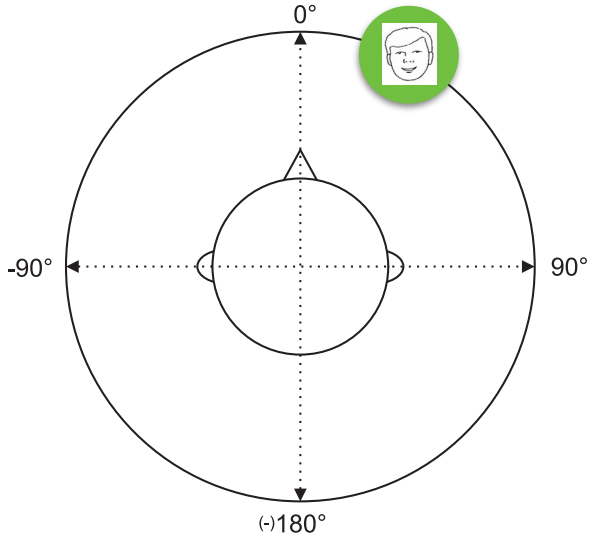


Estimation of *source* covariance matrix

Probabilistic spatial localization map

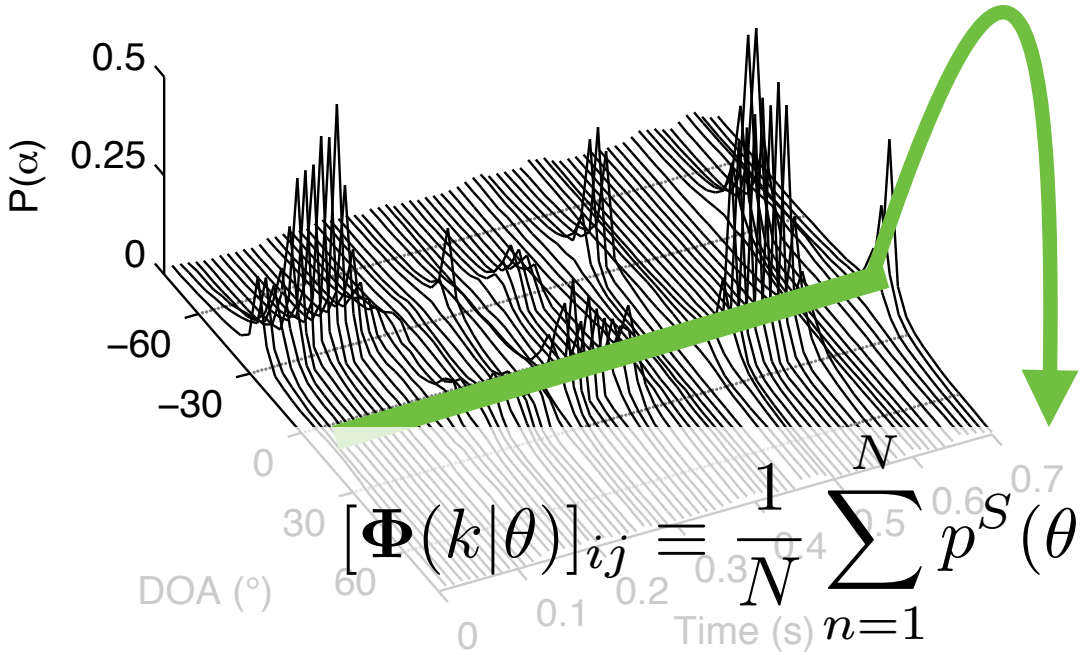


1. Decide for target source direction



Estimation of *source* covariance matrix

Probabilistic spatial localization map

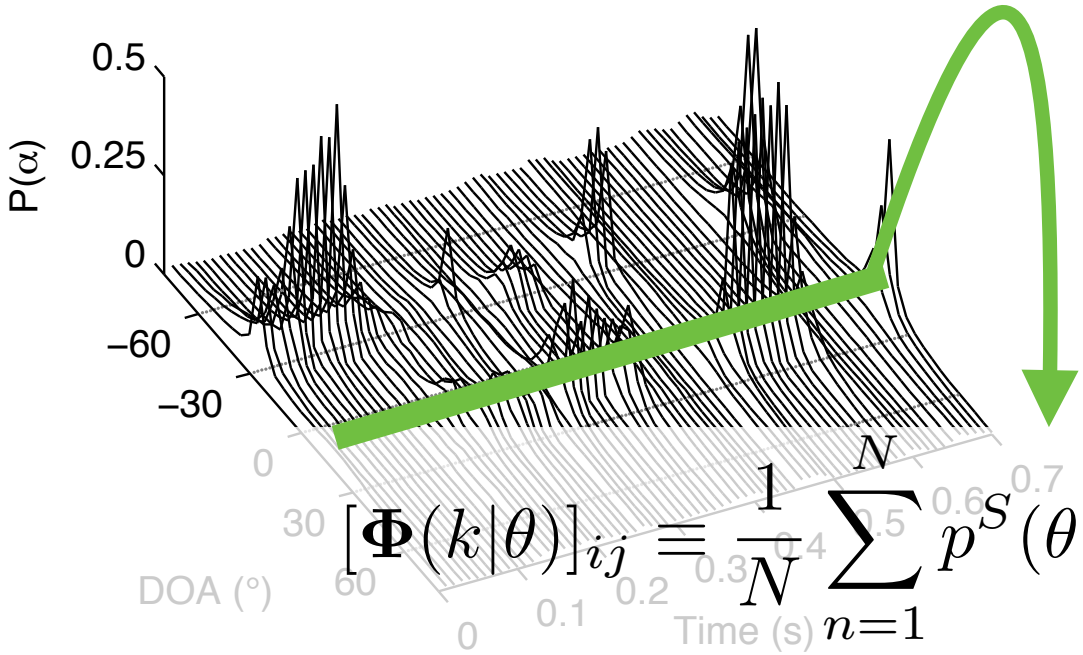


1. Decide for target source direction
2. Compute source-probability weighted microphone covariance matrix

$$[\mathbf{\Phi}(k|\theta)]_{ij} \equiv \frac{1}{N}\sum_{n=1}^{N} p^S(\theta,n)\, c_{ij}(n,k)^{-1}\, x_i^*(n,k)\, x_j(n,k)$$

Estimation of *source* covariance matrix

Probabilistic spatial localization map



1. Decide for target source direction
2. Compute source-probability weighted microphone covariance matrix

$$[\mathbf{\Phi}(k|\theta)]_{ij} \equiv \frac{1}{N}\sum_{n=1}^{N} p^S(\theta,n)\, c_{ij}(n,k)^{-1}\, x_i^*(n,k)\, x_j(n,k)$$

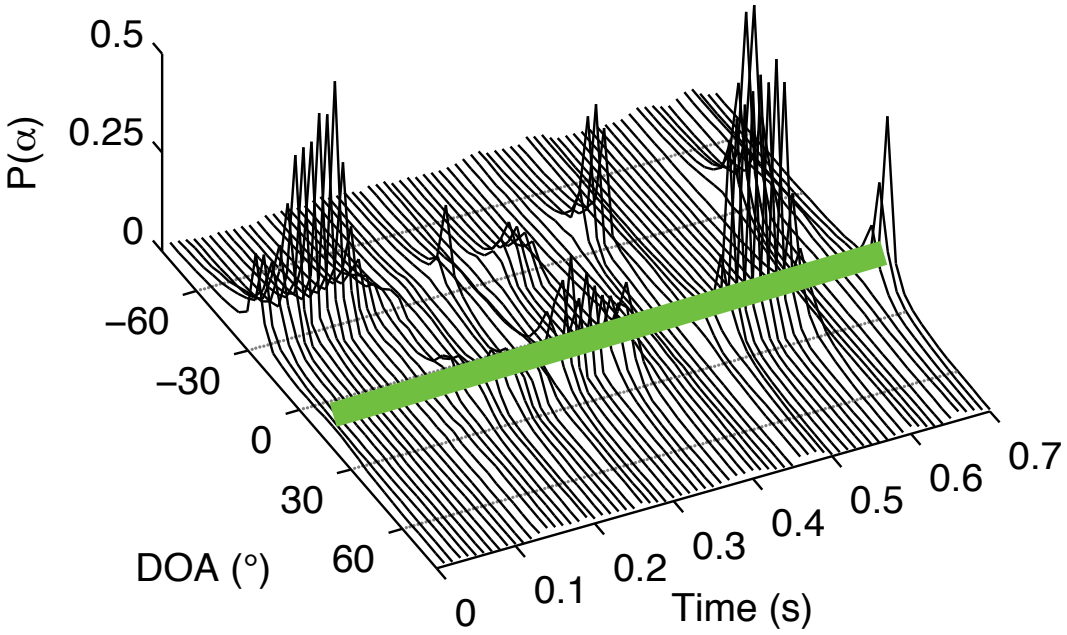
3. Normalization to unit gain, i.e., only phase retained

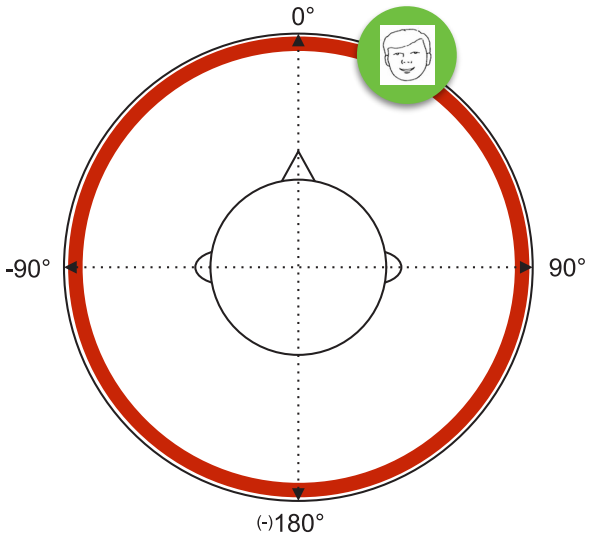
$$c_{ij}(n,k) = |x_i(n,k)|\,|x_j(n,k)|$$

$$d_j(k|\theta) = [\mathbf{\Phi}(k|\theta)]_{i^*j} \,/\, |[\mathbf{\Phi}(k|\theta)]_{i^*j}|$$

Estimation of *noise* covariance matrix

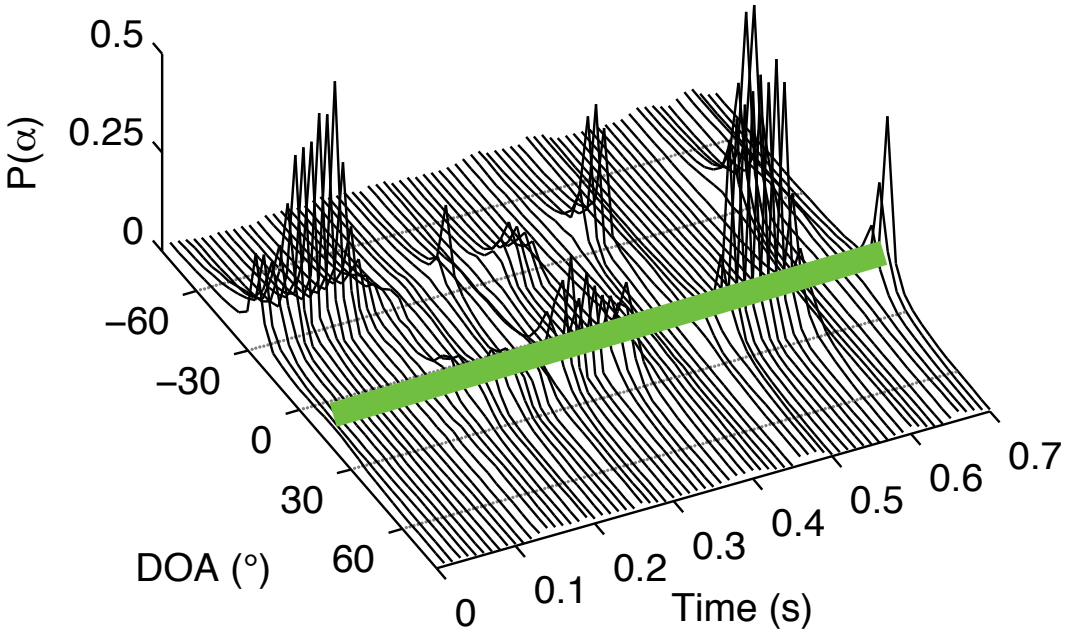
Probabilistic spatial localization map

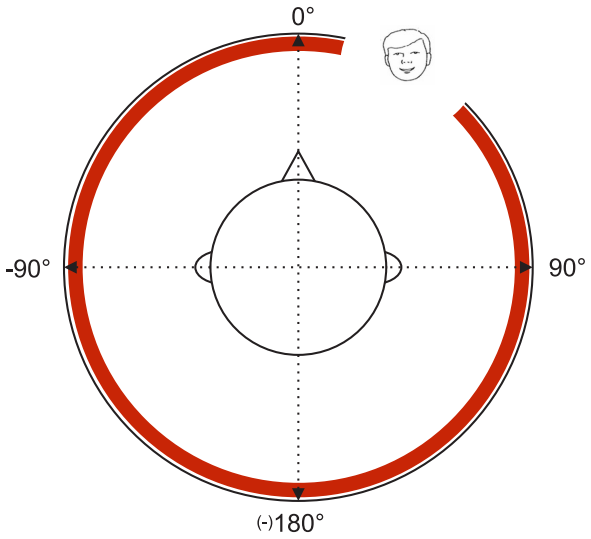




Estimation of *noise* covariance matrix

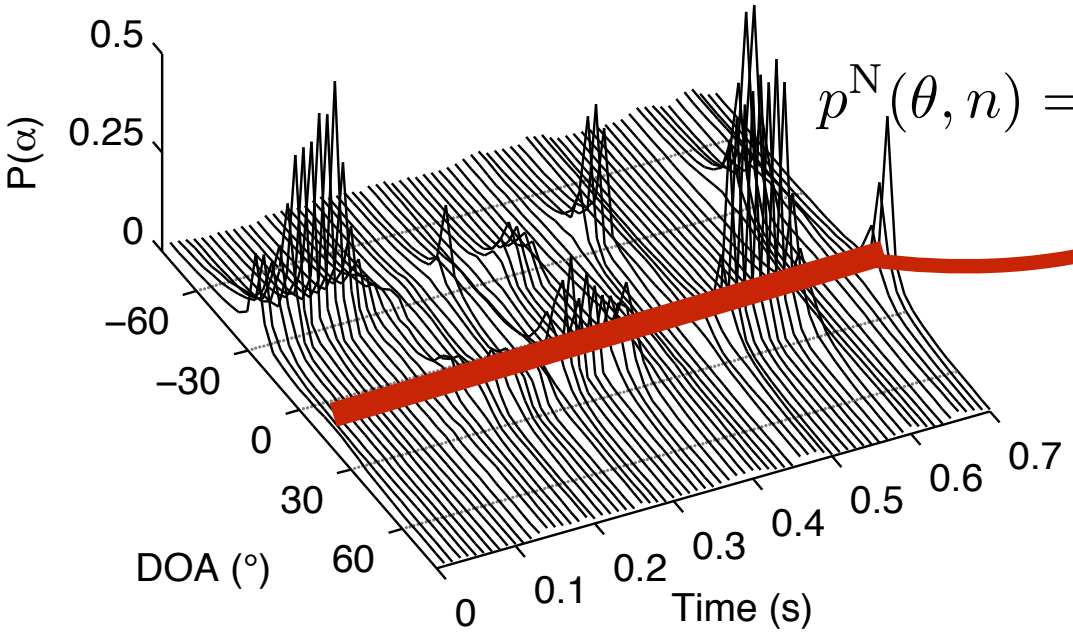
Probabilistic spatial localization map





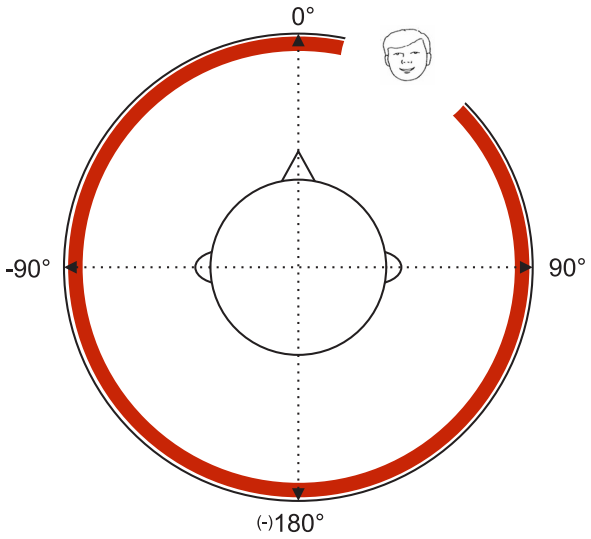
Estimation of *noise* covariance matrix

Probabilistic spatial localization map



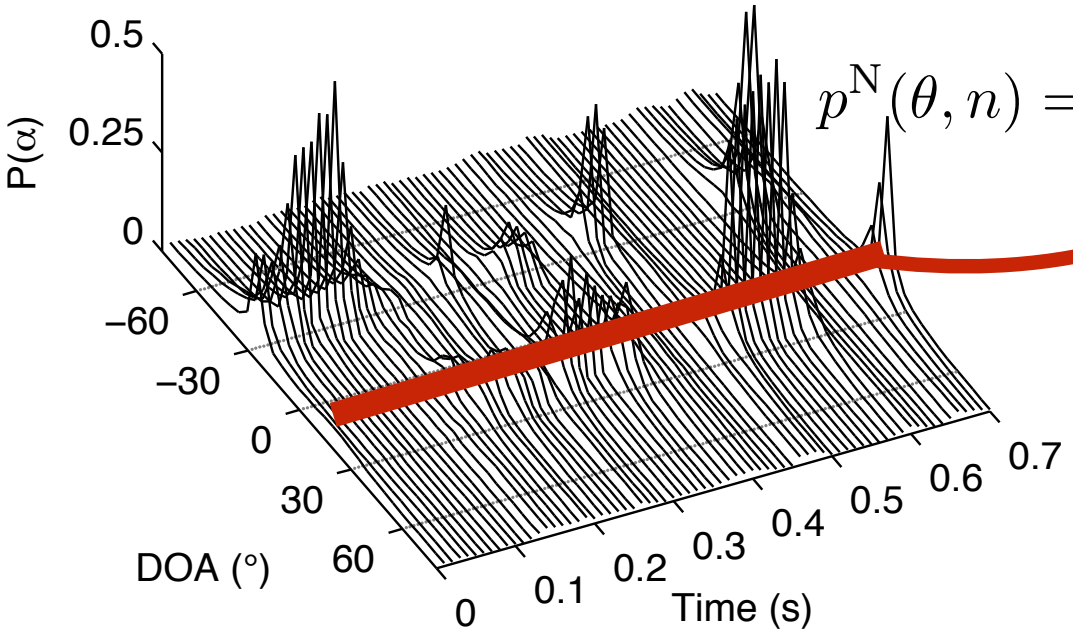
1. Estimated noise probability time-course

$$p^{\mathrm{N}}(\theta, n) = \begin{cases} \gamma\,(1 - p^{S}(\theta, n)), & p^{S}(\theta, n) < p_0 \\ 0, & p^{S}(\theta, n) \geq p_0 \end{cases}$$



Estimation of *noise* covariance matrix

Probabilistic spatial localization map



1. Estimated noise probability time-course

$$p^{\mathrm{N}}(\theta, n) = \begin{cases} \gamma\,(1 - p^{S}(\theta, n)), & p^{S}(\theta, n) < p_0 \\ 0, & p^{S}(\theta, n) \geq p_0 \end{cases}$$

2. Compute noise-probability weighted microphone covariance matrix

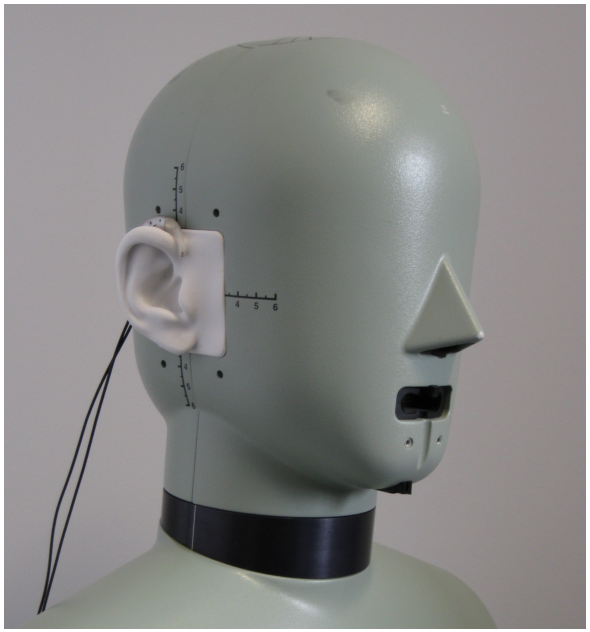
$$[\mathbf{R}(k|\theta)]_{ij} = \frac{1}{N} \sum_{n=1}^{N} p^{\mathrm{N}}(\theta, n)\, x_i^*(n, k)\, x_j(n, k)$$

Signal enhancement with MVDR

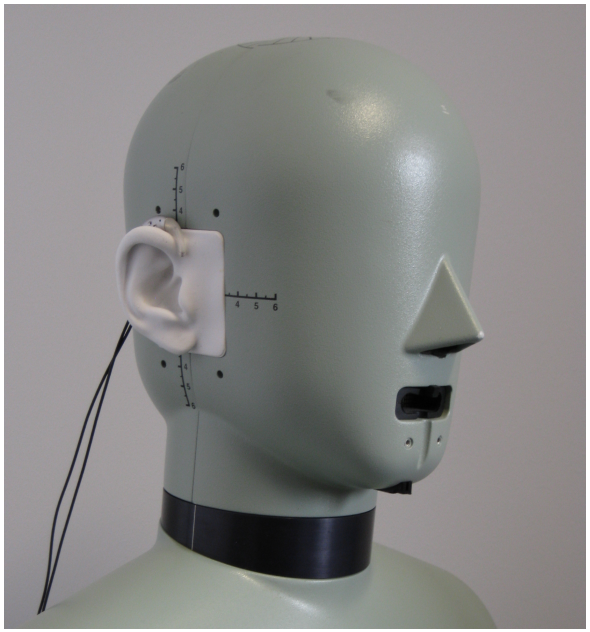
Use estimated source and noise covariances to form MVDR projection vector

$$\mathbf{w}(k|\theta) = \frac{\mathbf{R}^{-1}(k|\theta)\,\mathbf{d}(k|\theta)}{\mathbf{d}^{\mathsf{H}}(k|\theta)\,\mathbf{R}^{-1}(k|\theta)\,\mathbf{d}(k|\theta)}$$

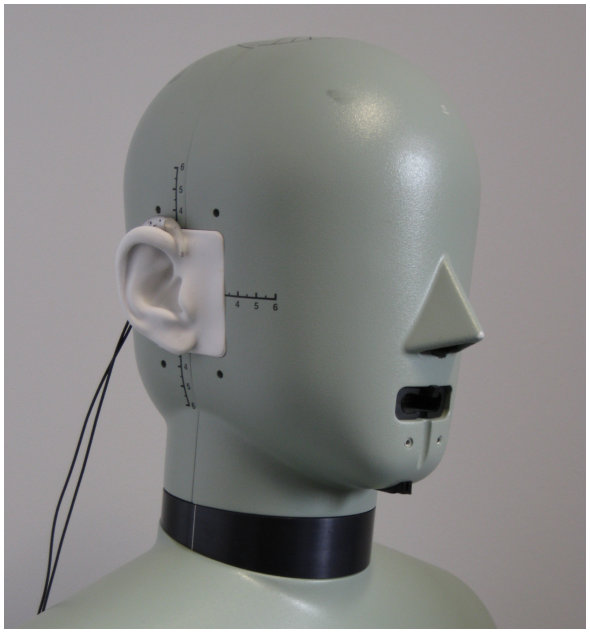
Evaluation: Setup

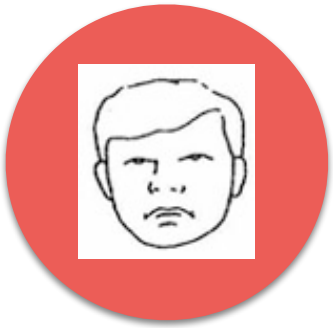


Evaluation: Setup

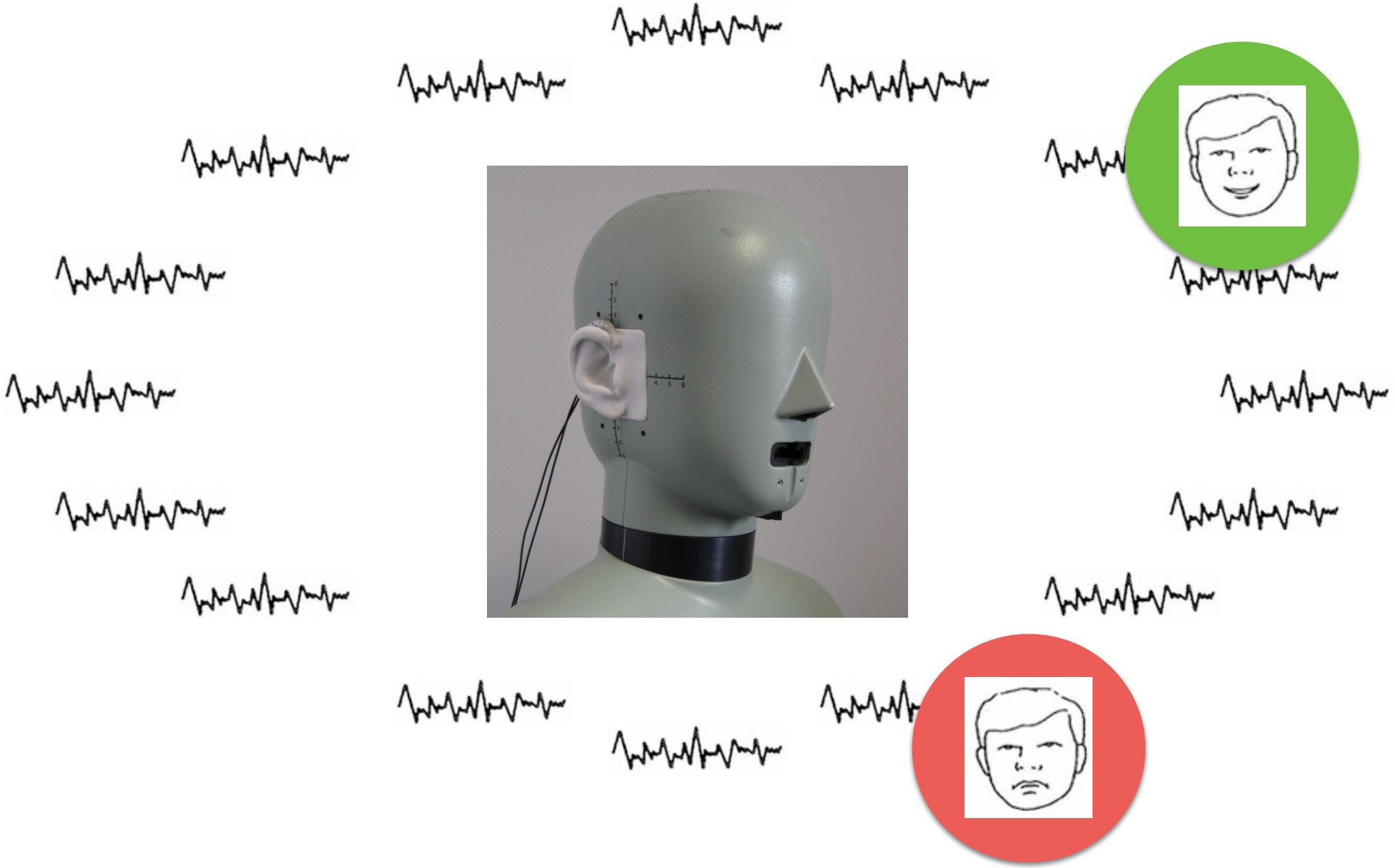


Evaluation: Setup





Evaluation: Setup



Evaluation: Data

6-channel bilateral **hearing aid** setup

Head-related impulse responses for **anechoic and reverberant** (office) environment (database [Kayser et al., 2009])

Target speech: **TIMIT** utterances

Interfering speaker: **TIMIT** utterances, different spatial position

SIR: -10dB, 0dB, 10dB and ∞dB

Noise: head-related **isotropic noise field**, speech shaped spectrum

SNR: -10dB, 0dB, 10dB and ∞dB

Target and interferer **positions**:

6832 position combinations in **anechoic** environment
3472 in **office** environment

Evaluation: Acoustic parameter models

Comparison of:
proposed **probabilistic estimation** of speech and noise covariance with
free-field target model and **isotropic** noise model

	Speech covariance	Noise covariance
PrS+PrN	prob. model	prob. model
FfS+PrN	free-field HRTF model	prob. model
PrS+IsoN	prob. model	isotr. model
FfS+IsoN	free-field HRTF model	isotr. model

Evaluation: Acoustic parameter models

Comparison of:
proposed **probabilistic estimation** of speech and noise covariance with
free-field target model and **isotropic** noise model

	Speech covariance	Noise covariance
PrS+PrN	prob. model	prob. model
FfS+PrN	free-field HRTF model	prob. model
PrS+IsoN	prob. model	isotr. model
FfS+IsoN	free-field HRTF model	isotr. model

Anemüller, Kayser (2017)

Results: Anechoic test conditions

Anechoic environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	3.0	**9.6**	-1.0	6.9
-10	0	7.7	**15.1**	-1.5	8.9
-10	10	12.9	**20.8**	-0.8	10.0
-10	∞	18.6	**26.3**	0.8	10.2
0	-10	1.7	**7.8**	1.4	6.1
0	0	2.6	**9.1**	2.2	6.9
0	10	7.0	**13.4**	2.6	8.8
0	∞	16.3	**20.8**	3.7	10.2
10	-10	1.7	**7.6**	1.7	6.1
10	0	1.5	**7.3**	3.5	6.2
10	10	2.7	**8.1**	4.6	6.9
10	∞	12.9	**14.7**	5.8	10.2
∞	-10	1.9	**7.6**	1.7	6.1
∞	0	0.9	**7.1**	3.5	6.1
∞	10	2.2	**6.3**	4.6	6.1

Results: Anechoic test conditions

Anechoic environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	3.0	**9.6**	-1.0	6.9
-10	0	7.7	**15.1**	-1.5	8.9
-10	10	12.9	**20.8**	-0.8	10.0
-10	∞	18.6	**26.3**	0.8	10.2
0	-10	1.7	**7.8**	1.4	6.1
0	0	2.6	**9.1**	2.2	6.9
0	10	7.0	**13.4**	2.6	8.8
0	∞	16.3	**20.8**	3.7	10.2
10	-10	1.7	**7.6**	1.7	6.1
10	0	1.5	**7.3**	3.5	6.2
10	10	2.7	**8.1**	4.6	6.9
10	∞	12.9	**14.7**	5.8	10.2
∞	-10	1.9	**7.6**	1.7	6.1
∞	0	0.9	**7.1**	3.5	6.1
∞	10	2.2	**6.3**	4.6	6.1

Prob. source+ prob. noise model: correct, but general

Results: Anechoic test conditions

Anechoic environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	3.0	**9.6**	-1.0	6.9
-10	0	7.7	**15.1**	-1.5	8.9
-10	10	12.9	**20.8**	-0.8	10.0
-10	∞	18.6	**26.3**	0.8	10.2
0	-10	1.7	**7.8**	1.4	6.1
0	0	2.6	**9.1**	2.2	6.9
0	10	7.0	**13.4**	2.6	8.8
0	∞	16.3	**20.8**	3.7	10.2
10	-10	1.7	**7.6**	1.7	6.1
10	0	1.5	**7.3**	3.5	6.2
10	10	2.7	**8.1**	4.6	6.9
10	∞	12.9	**14.7**	5.8	10.2
∞	-10	1.9	**7.6**	1.7	6.1
∞	0	0.9	**7.1**	3.5	6.1
∞	10	2.2	**6.3**	4.6	6.1

Prob. source+ prob. noise model:
correct, but general

Free-field source + prob. noise model:
correct, with Ff constraint

Results: Anechoic test conditions

Anechoic environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	3.0	**9.6**	-1.0	6.9
-10	0	7.7	**15.1**	-1.5	8.9
-10	10	12.9	**20.8**	-0.8	10.0
-10	∞	18.6	**26.3**	0.8	10.2
0	-10	1.7	**7.8**	1.4	6.1
0	0	2.6	**9.1**	2.2	6.9
0	10	7.0	**13.4**	2.6	8.8
0	∞	16.3	**20.8**	3.7	10.2
10	-10	1.7	**7.6**	1.7	6.1
10	0	1.5	**7.3**	3.5	6.2
10	10	2.7	**8.1**	4.6	6.9
10	∞	12.9	**14.7**	5.8	10.2
∞	-10	1.9	**7.6**	1.7	6.1
∞	0	0.9	**7.1**	3.5	6.1
∞	10	2.2	**6.3**	4.6	6.1

Prob. source+ prob. noise model:
correct, but general

Free-field source + prob. noise model:
correct, with Ff constraint

Prob. source + isotropic noise model:
correct at high SIR, but general

Results: Anechoic test conditions

Anechoic environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	3.0	**9.6**	-1.0	6.9
-10	0	7.7	**15.1**	-1.5	8.9
-10	10	12.9	**20.8**	-0.8	10.0
-10	∞	18.6	**26.3**	0.8	10.2
0	-10	1.7	**7.8**	1.4	6.1
0	0	2.6	**9.1**	2.2	6.9
0	10	7.0	**13.4**	2.6	8.8
0	∞	16.3	**20.8**	3.7	10.2
10	-10	1.7	**7.6**	1.7	6.1
10	0	1.5	**7.3**	3.5	6.2
10	10	2.7	**8.1**	4.6	6.9
10	∞	12.9	**14.7**	5.8	10.2
∞	-10	1.9	**7.6**	1.7	6.1
∞	0	0.9	**7.1**	3.5	6.1
∞	10	2.2	**6.3**	4.6	6.1

Prob. source+ prob. noise model:
correct, but general

Free-field source + prob. noise model:
correct, with Ff constraint

Prob. source + isotropic noise model:
correct at high SIR, but general

Free-field source + isotropic noise model:
correct at high SIR, with Ff constraint

Results: Reverberant test conditions

(nb: training was anechoic)

Office environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	**6.0**	4.5	2.5	3.7
-10	0	**8.2**	7.4	1.4	3.4
-10	10	**10.2**	9.8	1.6	3.3
-10	∞	**10.9**	10.6	2.1	3.2
0	-10	**5.6**	2.7	5.0	4.3
0	0	3.8	2.0	**4.0**	3.7
0	10	**5.0**	3.6	4.2	3.4
0	∞	**6.4**	5.4	4.3	3.3
10	-10	**6.1**	2.5	5.3	4.4
10	0	3.1	-0.0	**5.2**	4.3
10	10	1.0	-1.1	**5.6**	3.8
10	∞	1.4	0.1	**6.1**	3.2
∞	-10	**6.3**	2.6	5.1	4.5
∞	0	4.1	0.1	**5.2**	4.6
∞	10	0.7	-2.5	**6.4**	4.5

Results: Reverberant test conditions

(nb: training was anechoic)

Office environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	**6.0**	4.5	2.5	3.7
-10	0	**8.2**	7.4	1.4	3.4
-10	10	**10.2**	9.8	1.6	3.3
-10	∞	**10.9**	10.6	2.1	3.2
0	-10	**5.6**	2.7	5.0	4.3
0	0	3.8	2.0	**4.0**	3.7
0	10	**5.0**	3.6	4.2	3.4
0	∞	**6.4**	5.4	4.3	3.3
10	-10	**6.1**	2.5	5.3	4.4
10	0	3.1	-0.0	**5.2**	4.3
10	10	1.0	-1.1	**5.6**	3.8
10	∞	1.4	0.1	**6.1**	3.2
∞	-10	**6.3**	2.6	5.1	4.5
∞	0	4.1	0.1	**5.2**	4.6
∞	10	0.7	-2.5	**6.4**	4.5

Prob. source+ prob. noise model: correct

Results: Reverberant test conditions

(nb: training was anechoic)

Office environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	**6.0**	4.5	2.5	3.7
-10	0	**8.2**	7.4	1.4	3.4
-10	10	**10.2**	9.8	1.6	3.3
-10	∞	**10.9**	10.6	2.1	3.2
0	-10	**5.6**	2.7	5.0	4.3
0	0	3.8	2.0	**4.0**	3.7
0	10	**5.0**	3.6	4.2	3.4
0	∞	**6.4**	5.4	4.3	3.3
10	-10	**6.1**	2.5	5.3	4.4
10	0	3.1	-0.0	**5.2**	4.3
10	10	1.0	-1.1	**5.6**	3.8
10	∞	1.4	0.1	**6.1**	3.2
∞	-10	**6.3**	2.6	5.1	4.5
∞	0	4.1	0.1	**5.2**	4.6
∞	10	0.7	-2.5	**6.4**	4.5

Prob. source+ prob. noise model: correct

Free-field source + prob. noise model: incorrect, with Ff constraint

Results: Reverberant test conditions

(nb: training was anechoic)

Office environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	**6.0**	4.5	2.5	3.7
-10	0	**8.2**	7.4	1.4	3.4
-10	10	**10.2**	9.8	1.6	3.3
-10	∞	**10.9**	10.6	2.1	3.2
0	-10	**5.6**	2.7	5.0	4.3
0	0	3.8	2.0	**4.0**	3.7
0	10	**5.0**	3.6	4.2	3.4
0	∞	**6.4**	5.4	4.3	3.3
10	-10	**6.1**	2.5	5.3	4.4
10	0	3.1	-0.0	**5.2**	4.3
10	10	1.0	-1.1	**5.6**	3.8
10	∞	1.4	0.1	**6.1**	3.2
∞	-10	**6.3**	2.6	5.1	4.5
∞	0	4.1	0.1	**5.2**	4.6
∞	10	0.7	-2.5	**6.4**	4.5

Prob. source+ prob. noise model: correct

Free-field source + prob. noise model: incorrect, with Ff constraint

Prob. source + isotropic noise model: approx. correct at high SIR, but general

Results: Reverberant test conditions

(nb: training was anechoic)

Office environment					
Input		**SINR improvement (dB)**			
SIR (dB)	SNR (dB)	PrS +PrN	FfS +PrN	PrS +IsoN	FfS +IsoN
-10	-10	**6.0**	4.5	2.5	3.7
-10	0	**8.2**	7.4	1.4	3.4
-10	10	**10.2**	9.8	1.6	3.3
-10	∞	**10.9**	10.6	2.1	3.2
0	-10	**5.6**	2.7	5.0	4.3
0	0	3.8	2.0	**4.0**	3.7
0	10	**5.0**	3.6	4.2	3.4
0	∞	**6.4**	5.4	4.3	3.3
10	-10	**6.1**	2.5	5.3	4.4
10	0	3.1	-0.0	**5.2**	4.3
10	10	1.0	-1.1	**5.6**	3.8
10	∞	1.4	0.1	**6.1**	3.2
∞	-10	**6.3**	2.6	5.1	4.5
∞	0	4.1	0.1	**5.2**	4.6
∞	10	0.7	-2.5	**6.4**	4.5

Prob. source+ prob. noise model:
correct

Free-field source + prob. noise model:
incorrect, with Ff constraint

Prob. source + isotropic noise model:
approx. correct at high SIR, but general

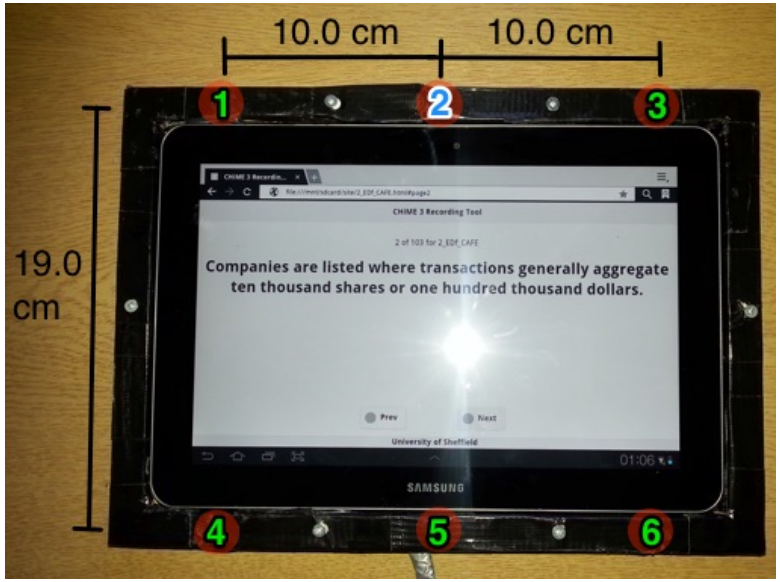
Free-field source + isotropic noise model:
incorrect, highly constrained

Localization-based signal enhancement: CHiME-3 speech recognition task

Six-channel tablet recordings

3-D localization (x, y + depth)

8 speakers, noisy environments



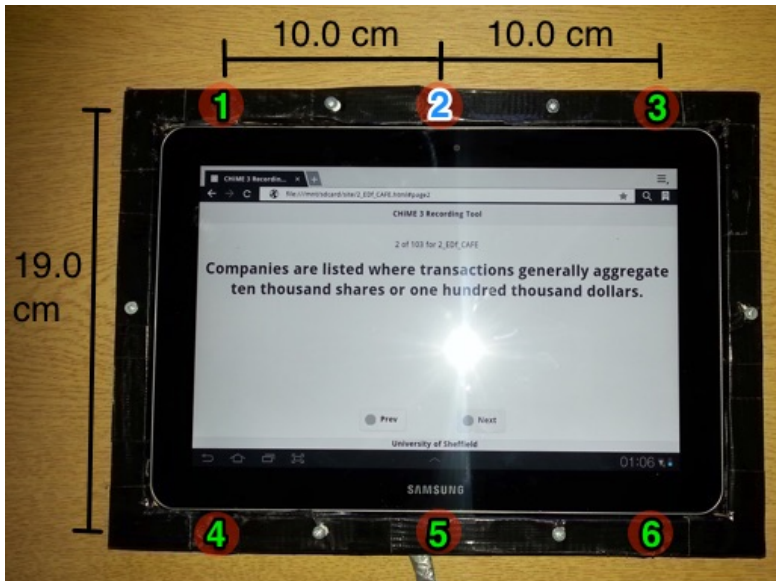
http://spandh.dcs.shef.ac.uk/chime_challenge/chime2015/overview.html

Localization-based signal enhancement: CHiME-3 speech recognition task

Six-channel tablet recordings

3-D localization (x, y + depth)

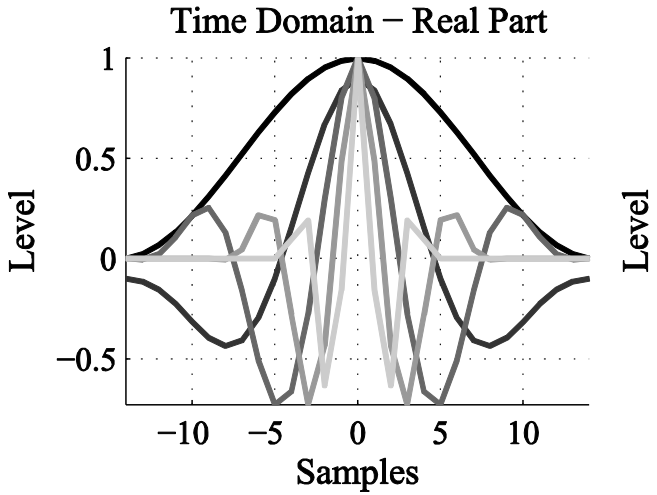
8 speakers, noisy environments



http://spandh.dcs.shef.ac.uk/chime_challenge/chime2015/overview.html

ASR system

Temporal modulation patterns as acoustic input features for ASR



7-layer hybrid DNN, 2047 sigmoid activation units

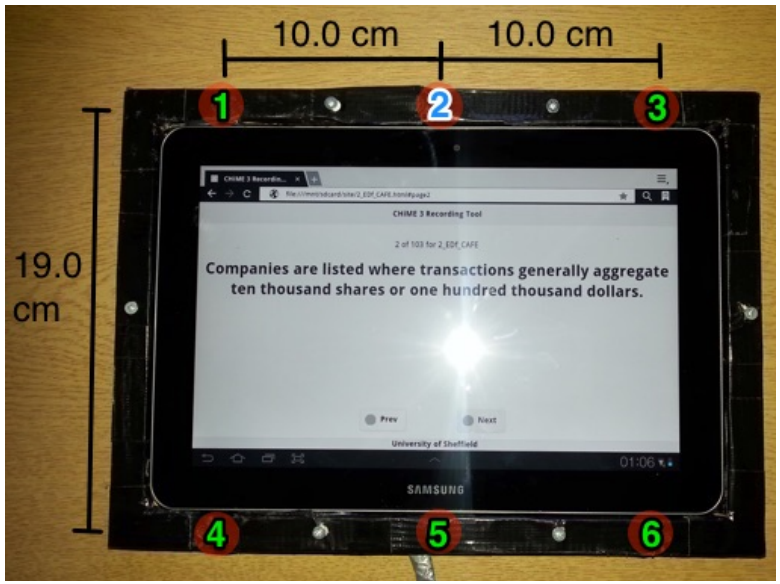
WSJ0 tri-gram,entropy pruning, RNN-based LM rescoring

Localization-based signal enhancement: CHiME-3 speech recognition task

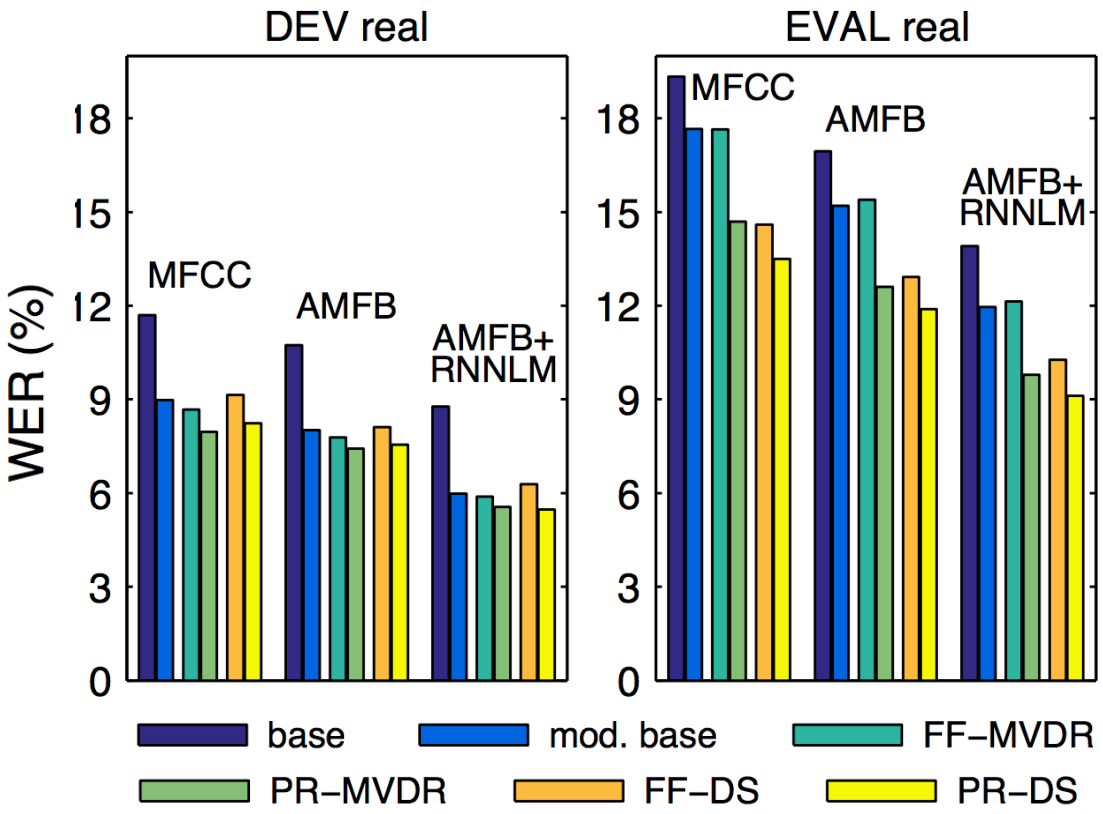
Six-channel tablet recordings

3-D localization (x, y + depth)

8 speakers, noisy environments



http://spandh.dcs.shef.ac.uk/chime_challenge/chime2015/overview.html



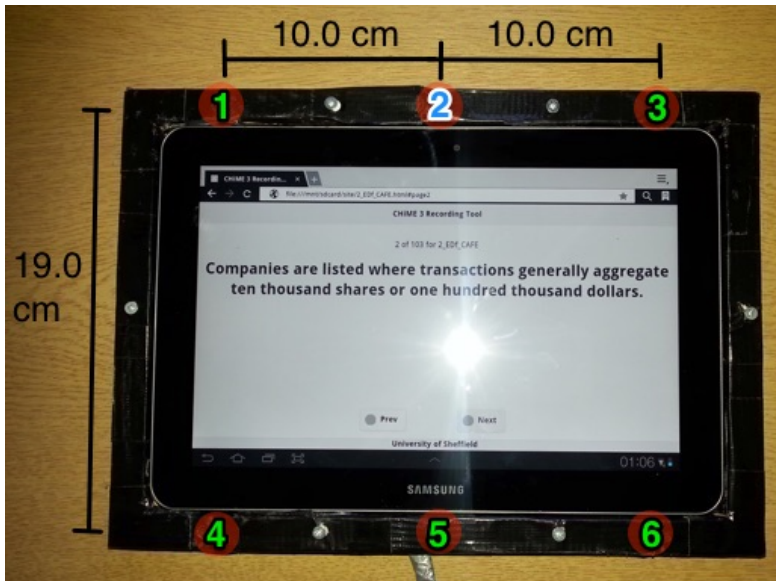
Kayser, Moritz, Anemüller (2016)

Localization-based signal enhancement: CHiME-3 speech recognition task

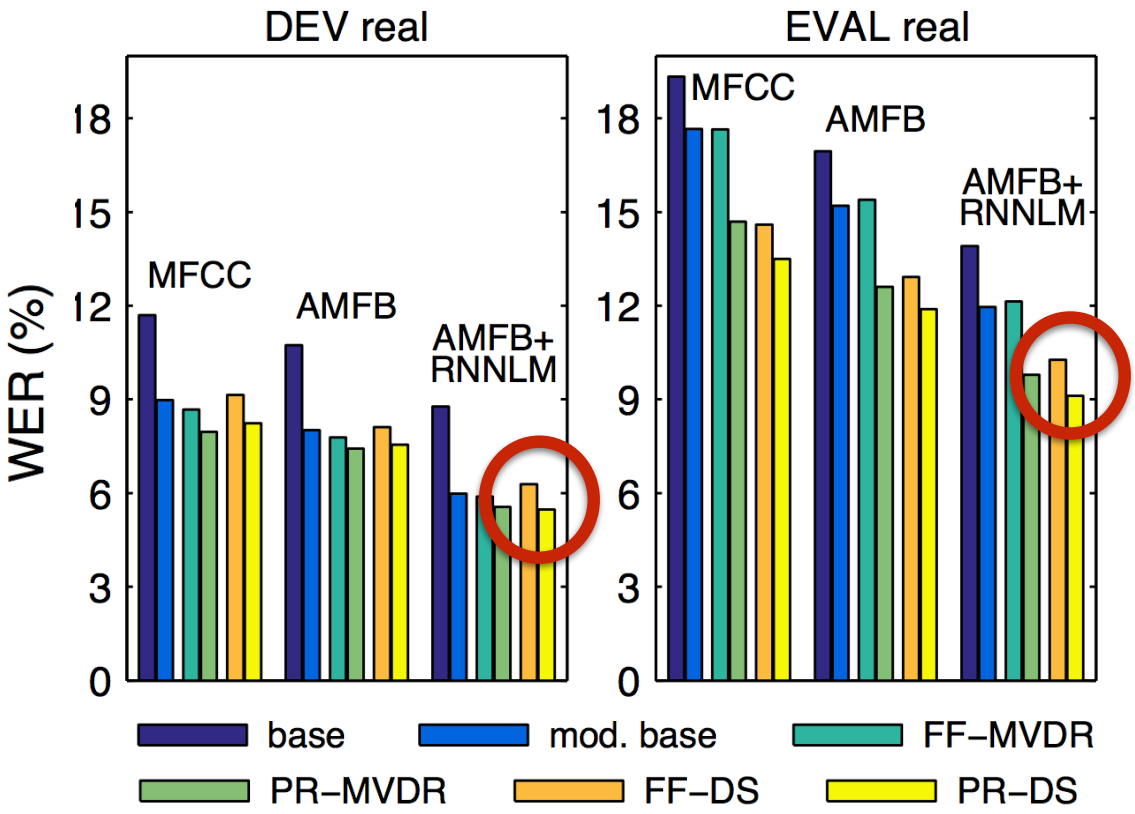
Six-channel tablet recordings

3-D localization (x, y + depth)

8 speakers, noisy environments



http://spandh.dcs.shef.ac.uk/chime_challenge/chime2015/overview.html

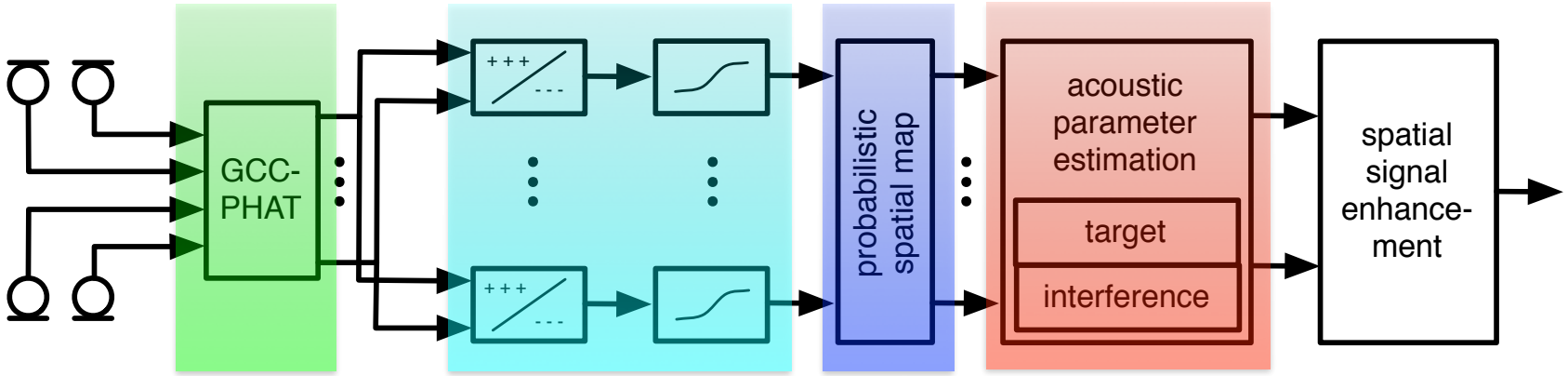


WER rel. impr.: 5.90% — 17.45%

Kayser, Moritz, Anemüller (2016)

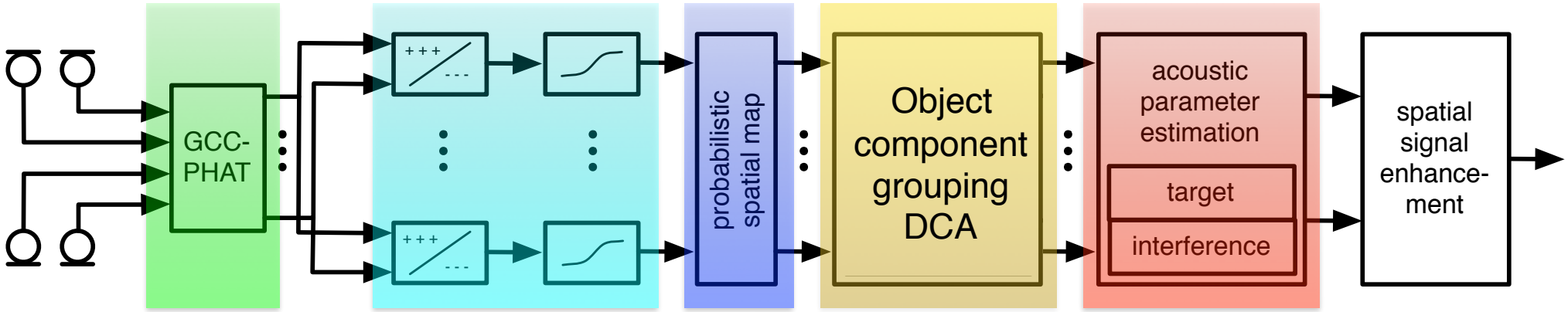
Current work:
Towards object based signal enhancement

Goal: Group spatial components of single object together using disjoint component analysis (DCA).



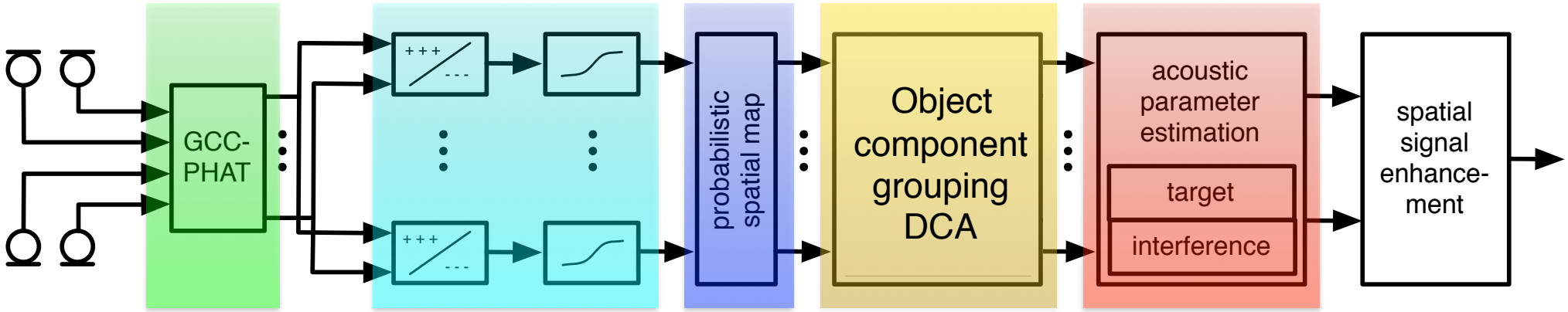
Current work:
Towards object based signal enhancement

Goal: Group spatial components of single object together using disjoint component analysis (DCA).



Current work:
Towards object based signal enhancement

Goal: Group spatial components of single object together using disjoint component analysis (DCA).



Anemüller, Kayser (2017b)

Conclusion

Supervised learning for probabilistic source localization:

- efficient: linear projection plus 1-dim. non-linearity
- derived from training data, no subsequent adaptation

(Re-) Estimation of acoustic parameters

- based on learned anechoic space representation
- adaptation per utterance to new acoustic environment

Results

- Anechoic environment: partly-fixed geometry model best
- Reverberant environment: full prob. re-estimation best

Conclusion

Supervised learning for probabilistic source localization:

- efficient: linear projection plus 1-dim. non-linearity
- derived from training data, no subsequent adaptation

(Re-) Estimation of acoustic parameters

- based on learned anechoic space representation
- adaptation per utterance to new acoustic environment

Results

- Anechoic environment: partly-fixed geometry model best
- Reverberant environment: full prob. re-estimation best