RUHR-UNIVERSITÄT BOCHUM

RUB

Binaural Sound Source Localisation using a Bayesian-network-based Blackboard System and Hypothesis-driven Feedback

Christopher Schymura, Thomas Walther, Dorothea Kolossa
Ruhr-Universität Bochum

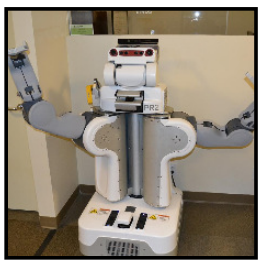
Ning Ma, Guy J. Brown
University of Sheffield

October 17, 2014

The Two!Ears project

KEMAR dummy head

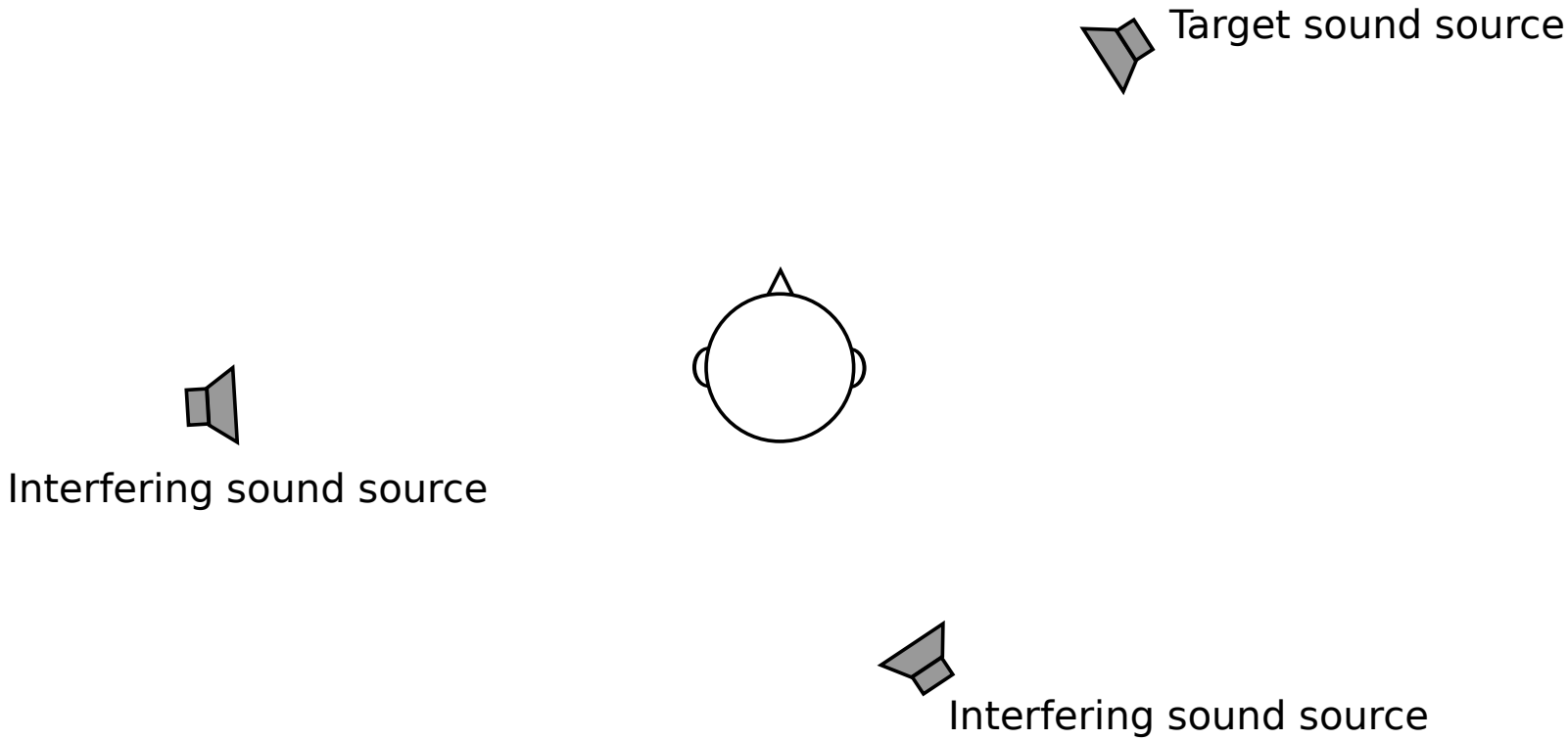
mounted on a



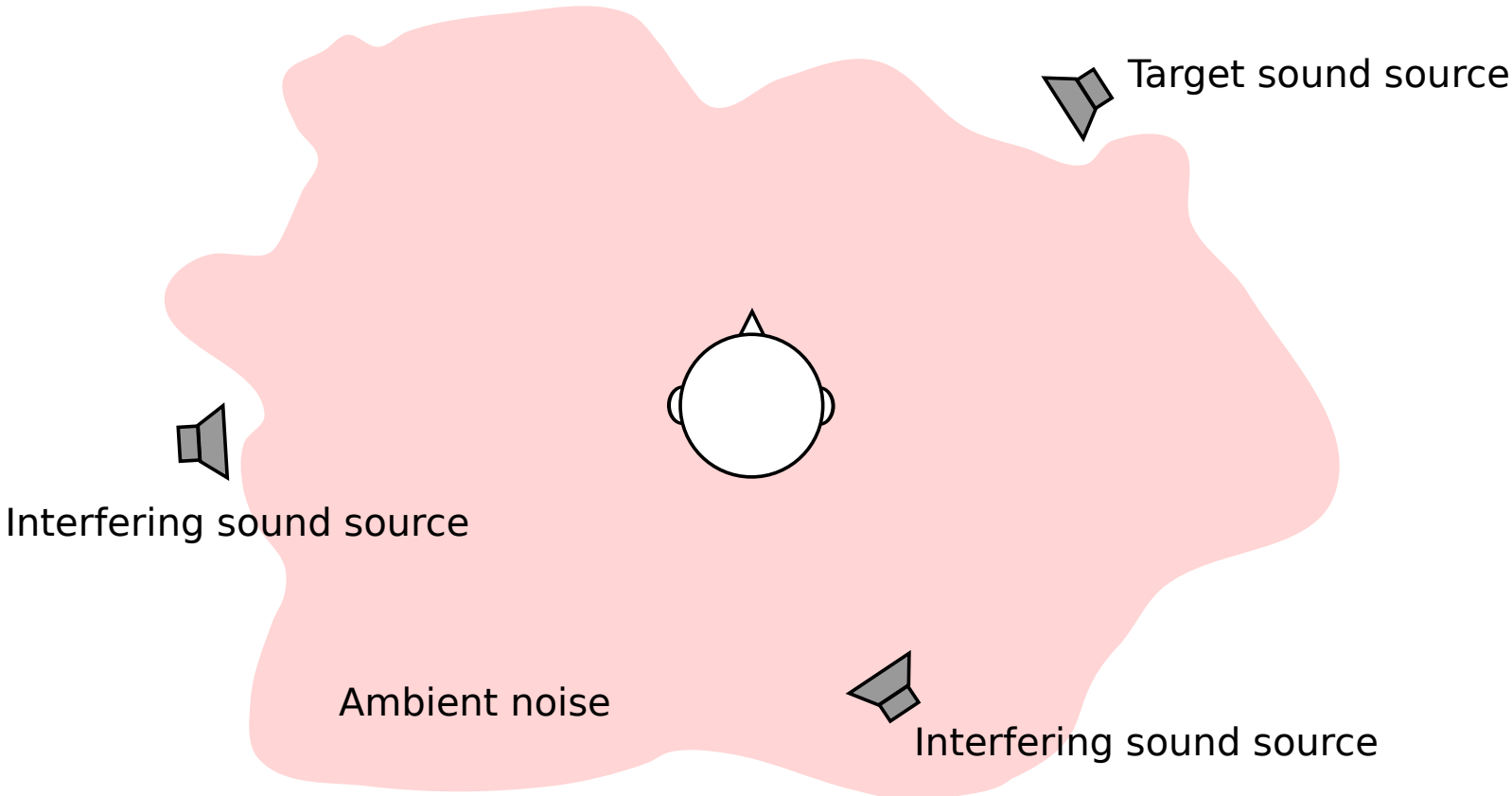
PR2 robotics platform

The Two!Ears project

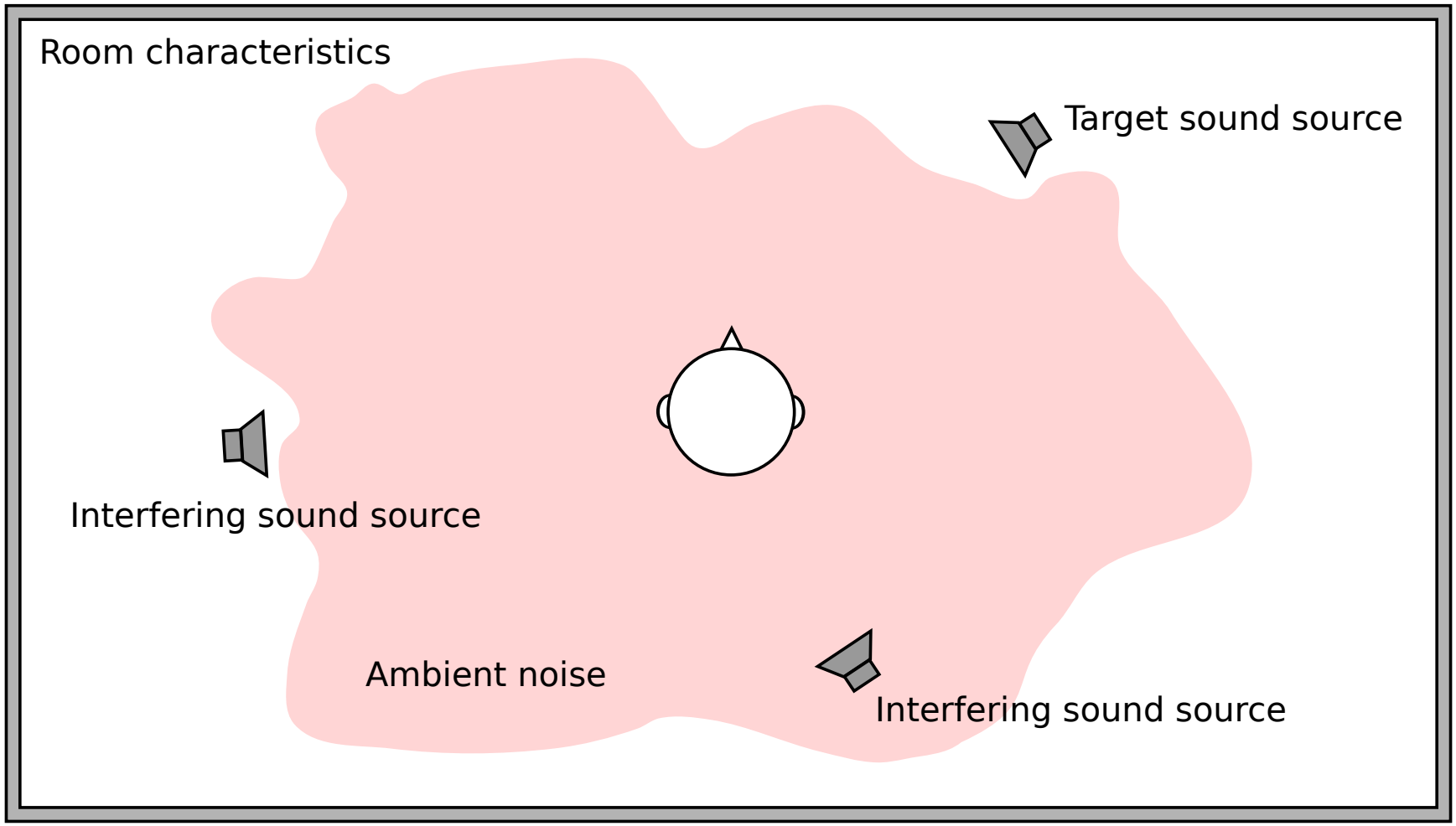
The Two!Ears project



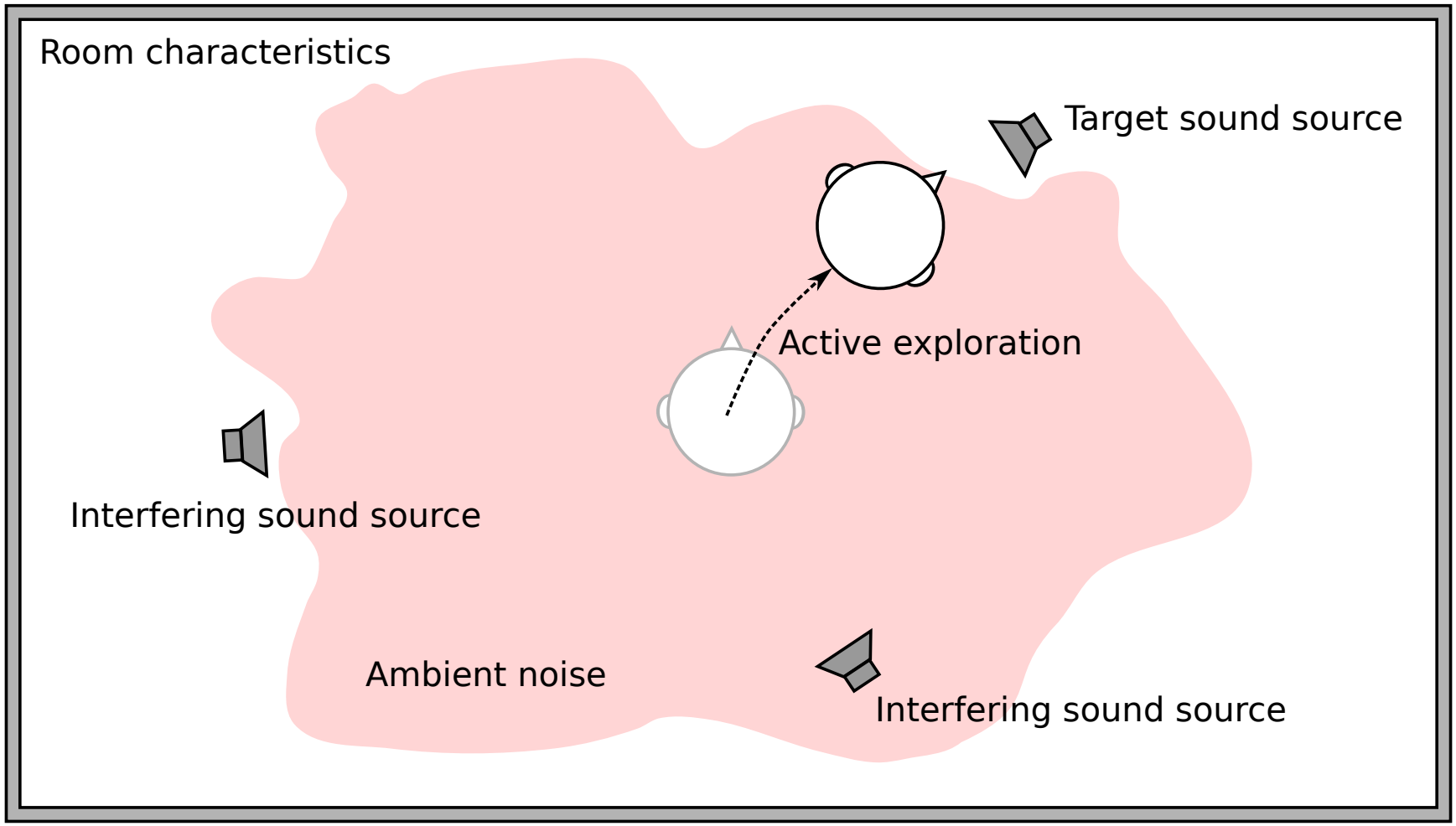
The Two!Ears project



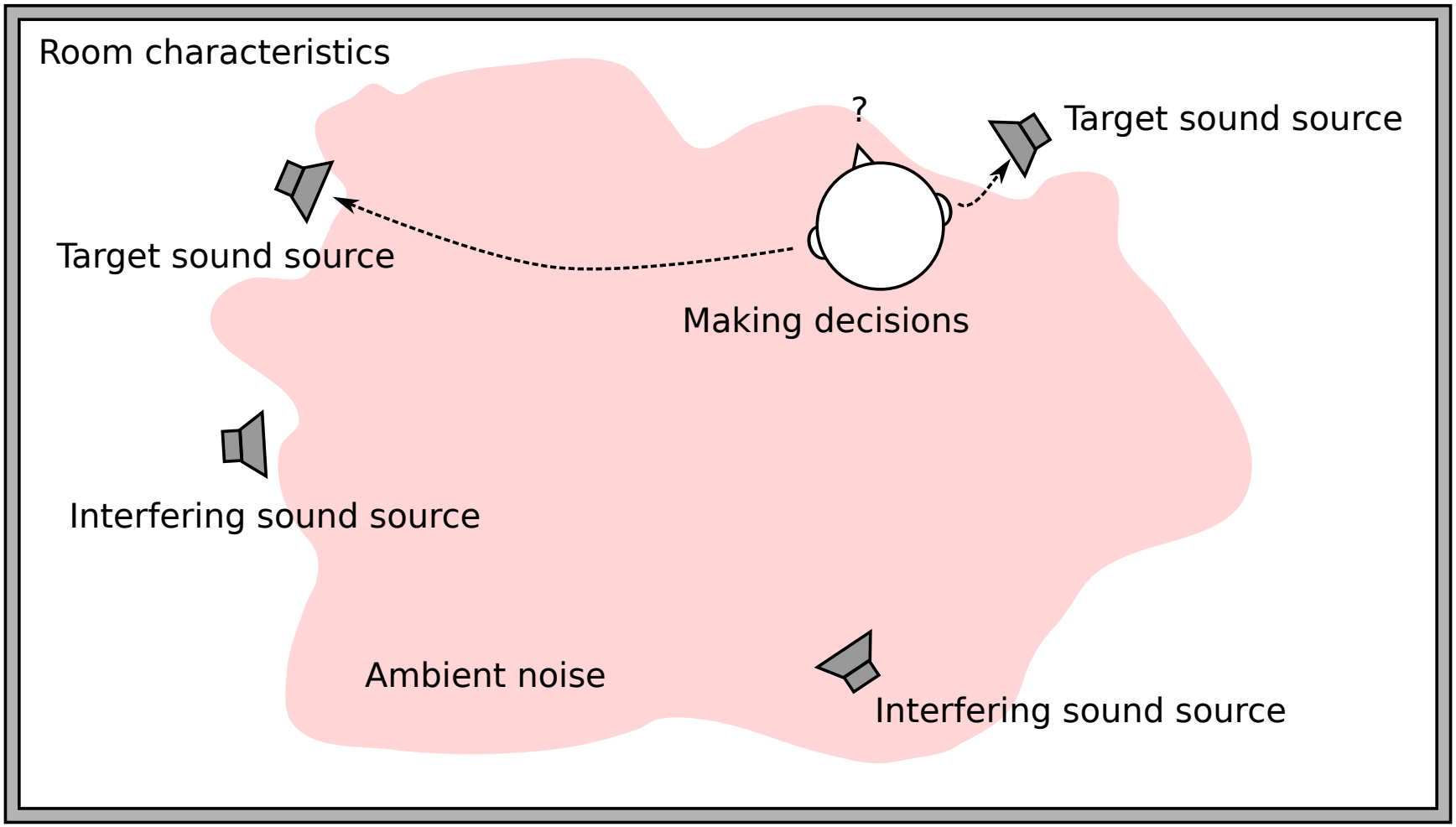
The Two!Ears project



The Two!Ears project



The Two!Ears project



The Two!Ears project

How can this be achieved?

The Two!Ears project

How can this be achieved?

- Build an internal represenation of the world that is currently perceived, referred to as the „internal world model“

The Two!Ears project

How can this be achieved?

- Build an internal represenation of the world that is currently perceived, referred to as the „internal world model“
- Use multimodal information (auditory, audiovisual and sensorimotor cues) to continously update the internal world model (bottom-up processing)

The Two!Ears project

How can this be achieved?

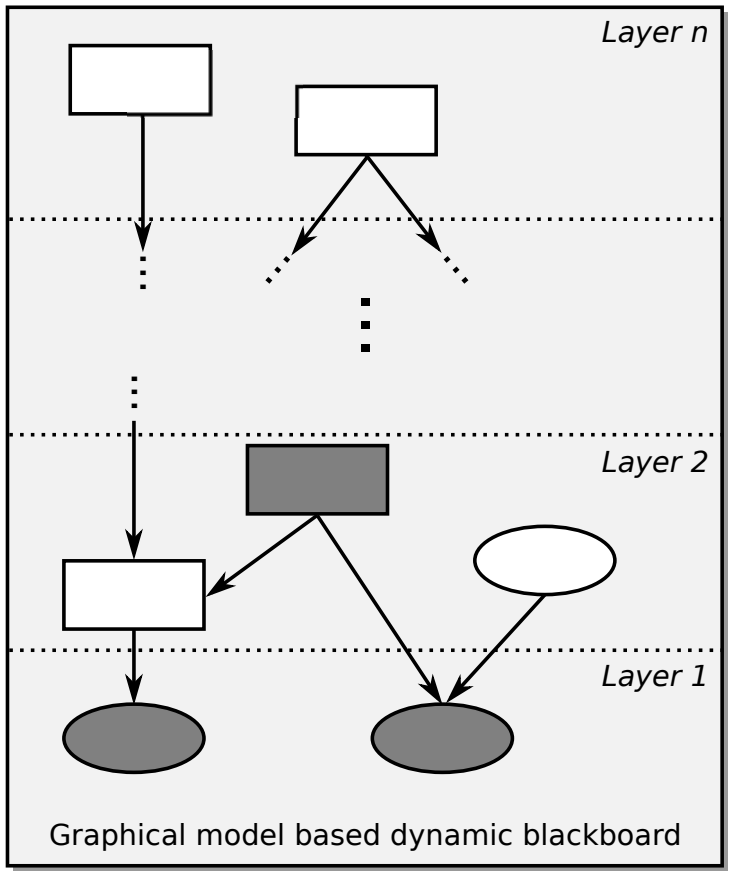
- Build an internal represenation of the world that is currently perceived, referred to as the „internal world model“
- Use multimodal information (auditory, audiovisual and sensorimotor cues) to continously update the internal world model (bottom-up processing)
- Generate hypotheses from the model and confirm or reject them by gathering new information by means of top-down feedback

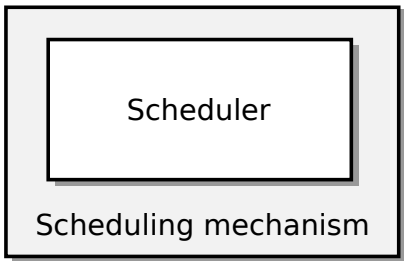
The Two!Ears project

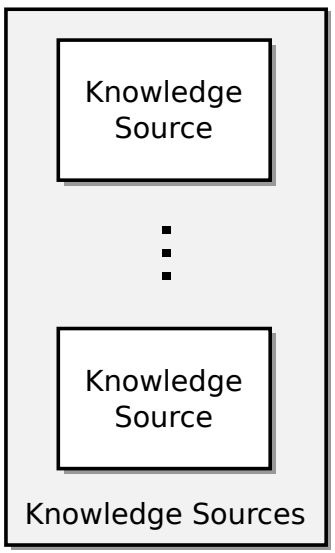
How can this be achieved?

- Build an internal represenation of the world that is currently perceived, referred to as the „internal world model“
- Use multimodal information (auditory, audiovisual and sensorimotor cues) to continously update the internal world model (bottom-up processing)
- Generate hypotheses from the model and confirm or reject them by gathering new information by means of top-down feedback
- Use the results to make appropriate decisions that incrementally lead to the accomplishment of a given task

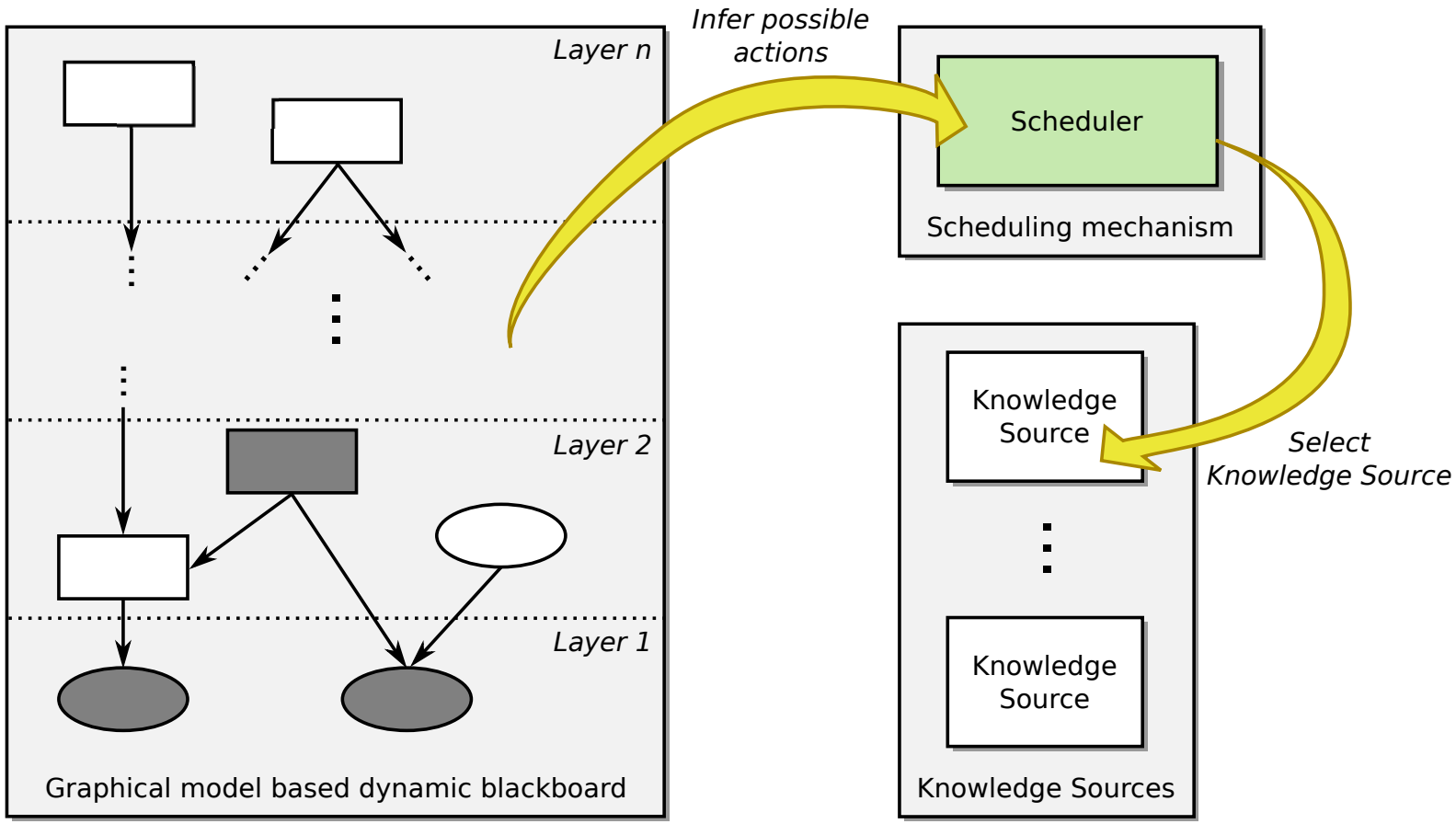
Graphical-model based Blackboard Systems



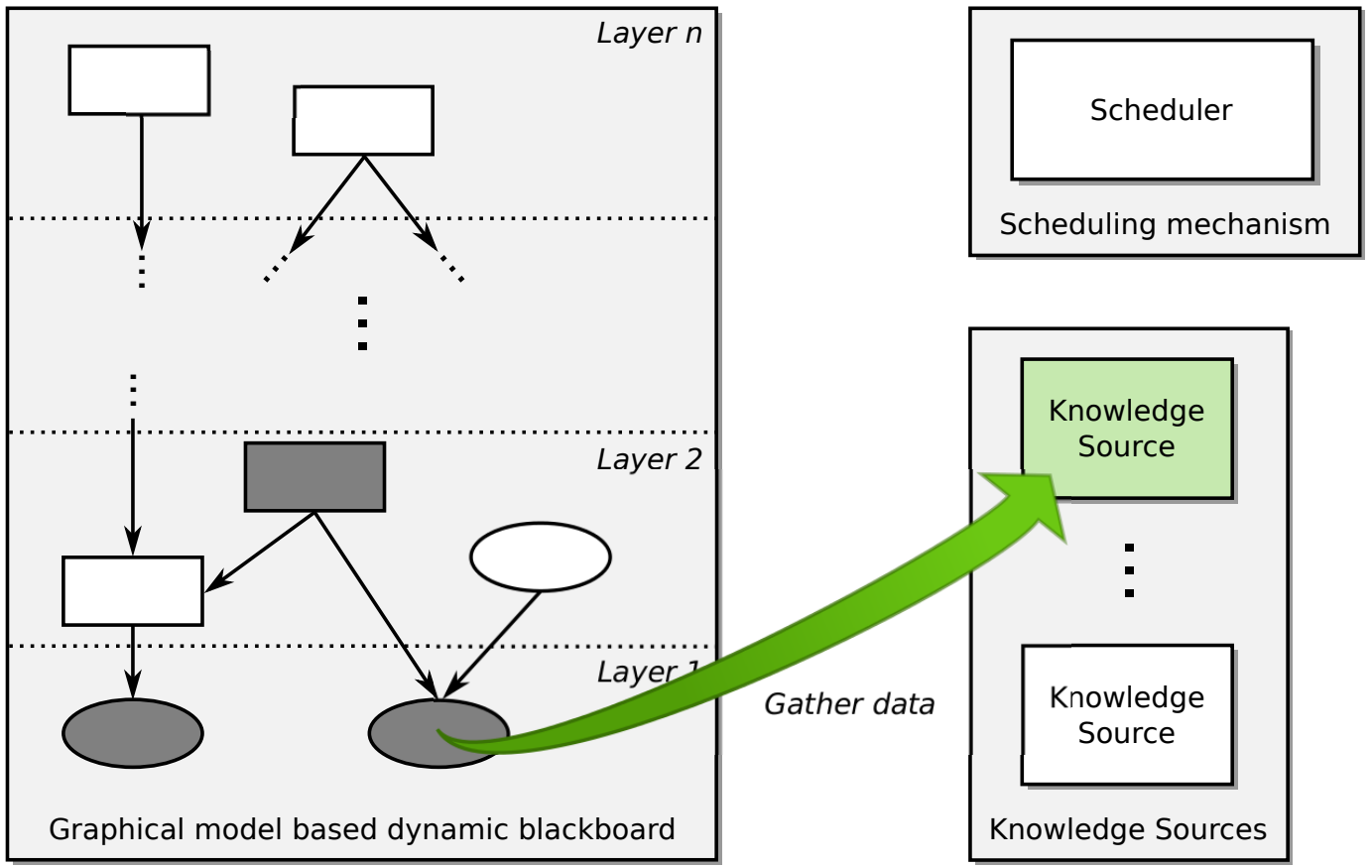




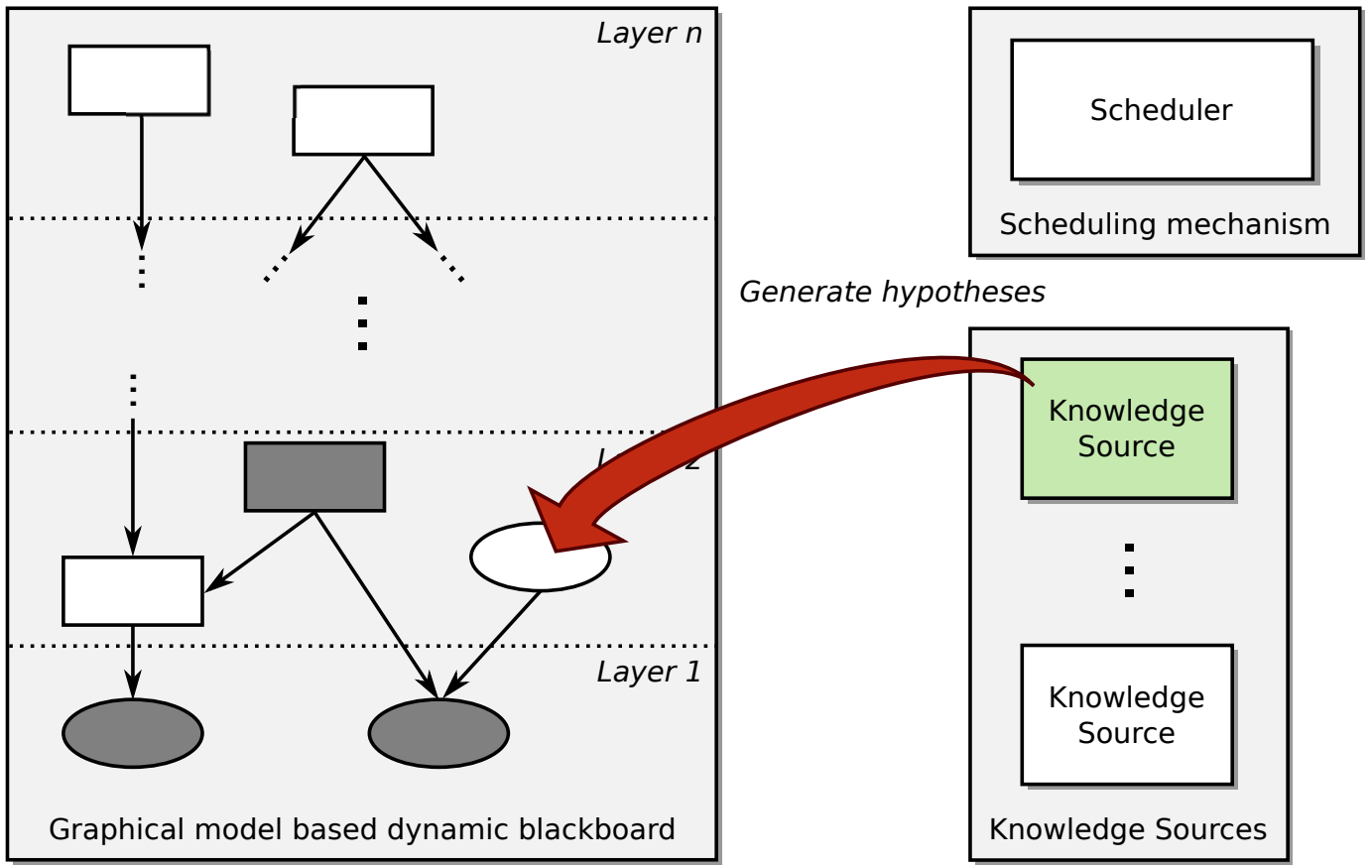
Graphical-model based Blackboard Systems



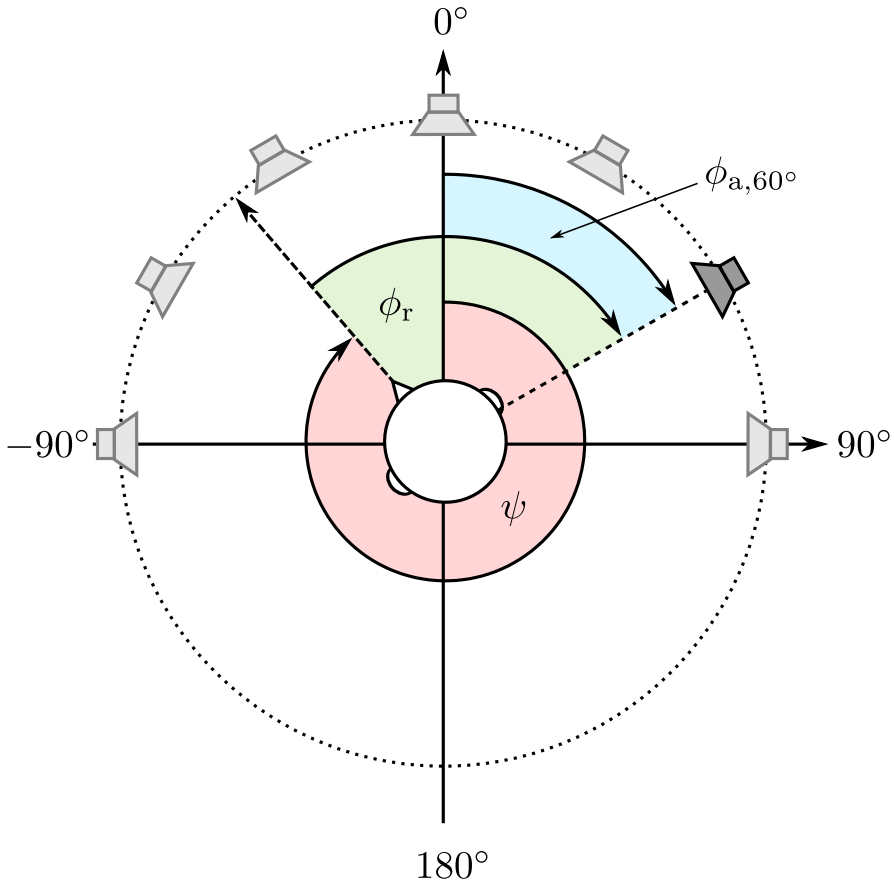
Graphical-model based Blackboard Systems



Graphical-model based Blackboard Systems



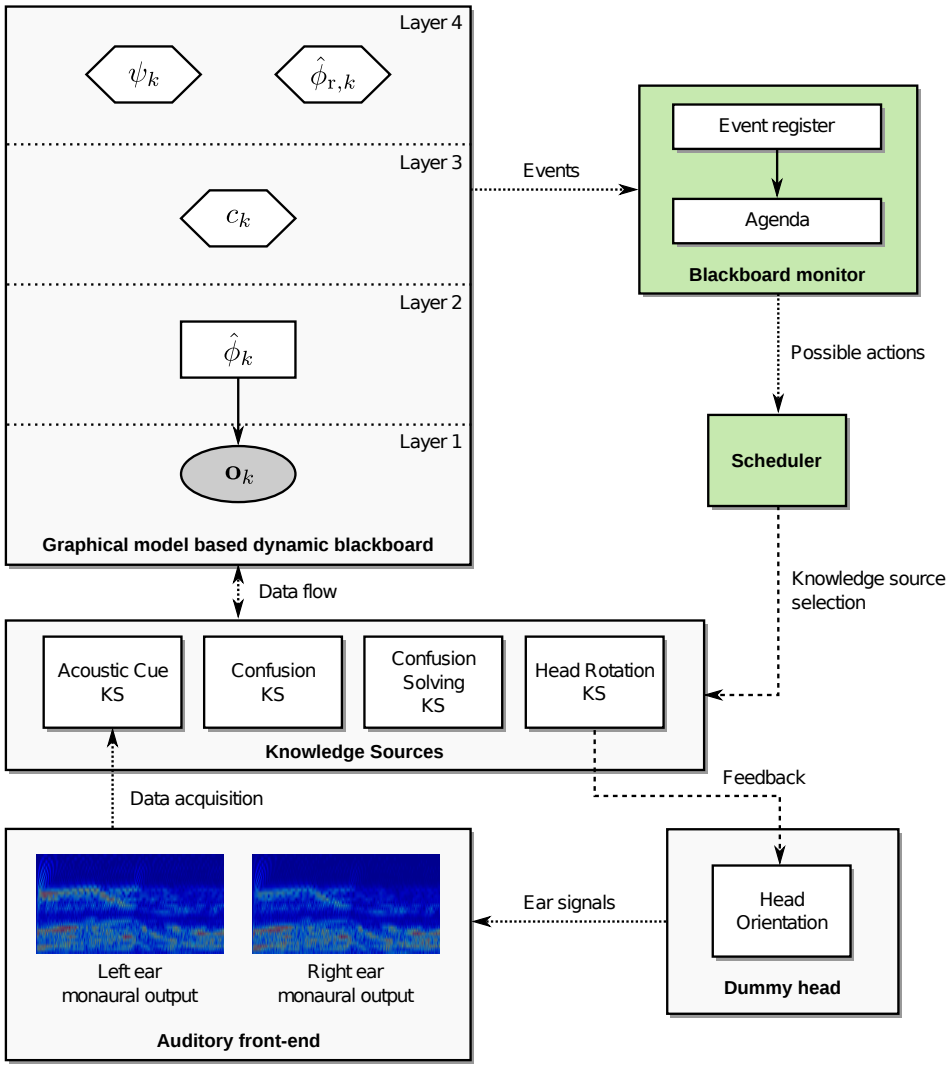
Proof of concept: Task



- 170 utterances of speech from the GRID corpus [1] for testing
- Free-field conditions (with optional ambient noise)
- Head rotations possible

[1] M. Cooke, J. Barker, S. Cunningham, X. Shao: An audio-visual corpus for speech perception and automatic speech recognition. Journal of the Acoustical Society of America, 2006

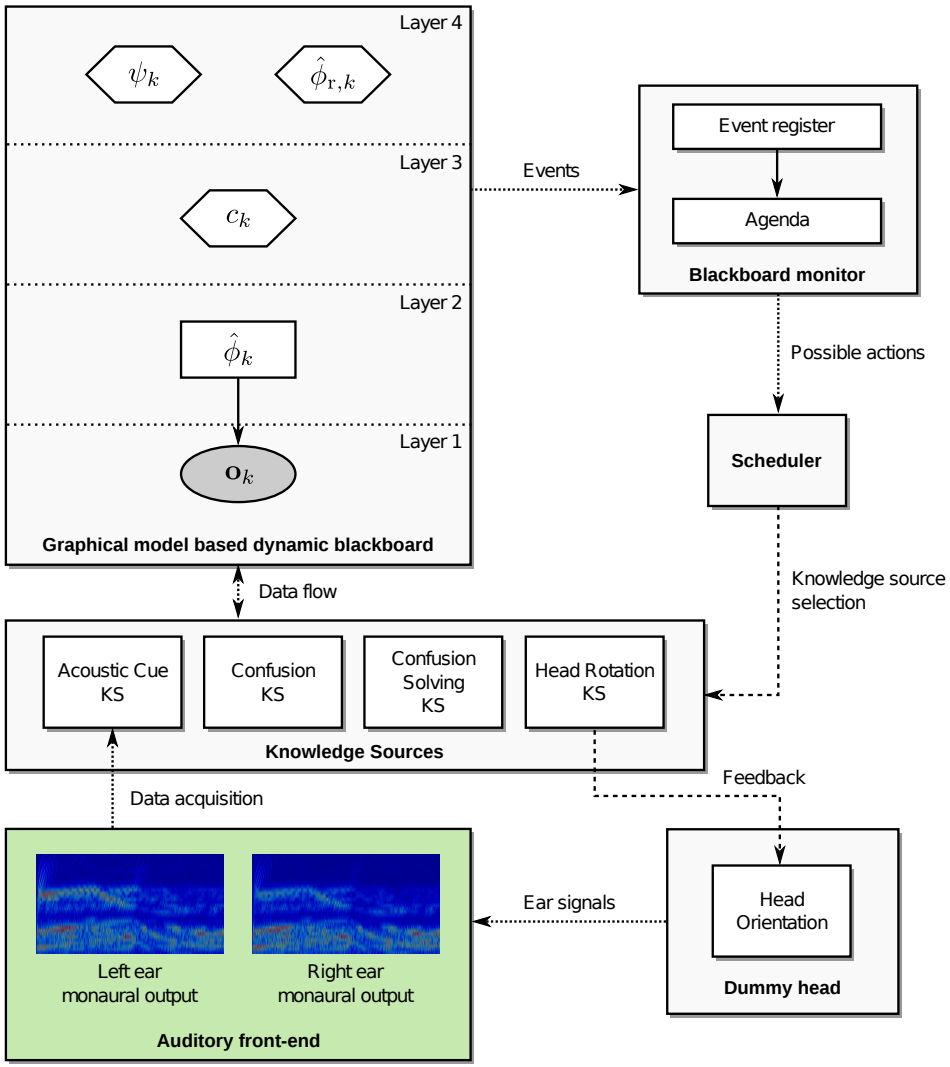
Proof of concept: System architecture



Computational framework:

- Event-based processing
- Blackboard monitor keeps track of events that have been generated by the blackboard
- Scheduler selects possible actions according to the triggered events

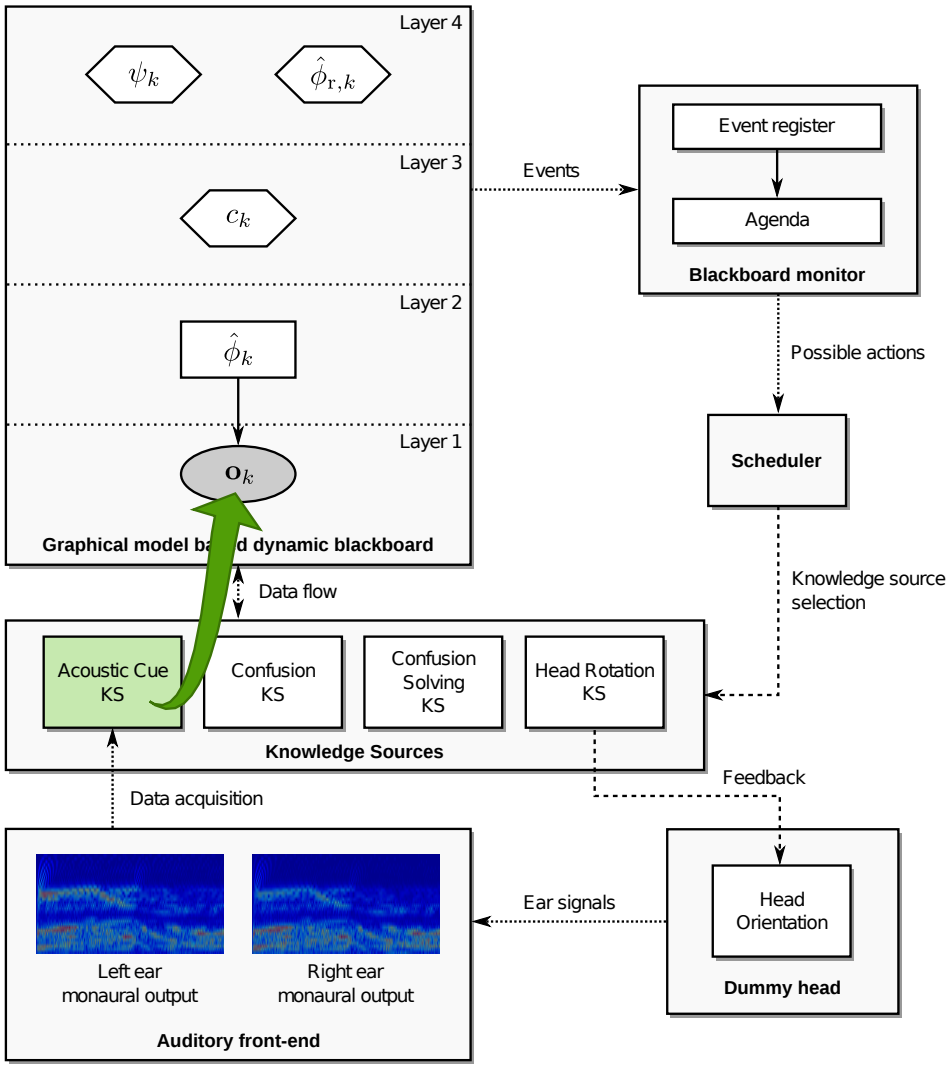
Proof of concept: System architecture



Auditory front-end:

- Gammatone filterbank
- Simple IHC model (half-wave rectification and square-root compression)

Proof of concept: System architecture



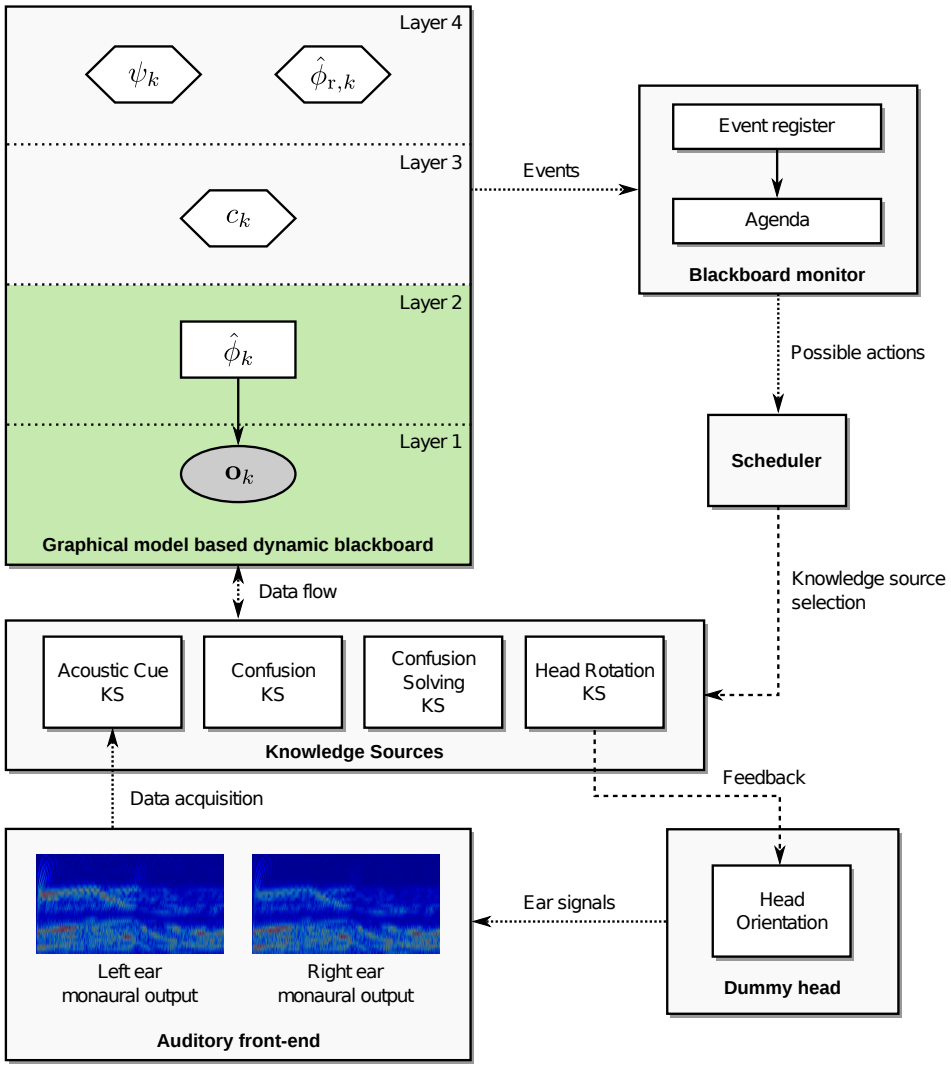
Auditory cues:

- ITDs $\tau_{k,m}$ and
- ILDs $\delta_{k,m}$, with

$$\boldsymbol{o}_k = \begin{pmatrix} \tau_{k,1} \\ \vdots \\ \tau_{k,M} \\ \delta_{k,1} \\ \vdots \\ \delta_{k,M} \end{pmatrix}$$

- k: Frame index
- m: Channel index
- M: Number of filterbank channels

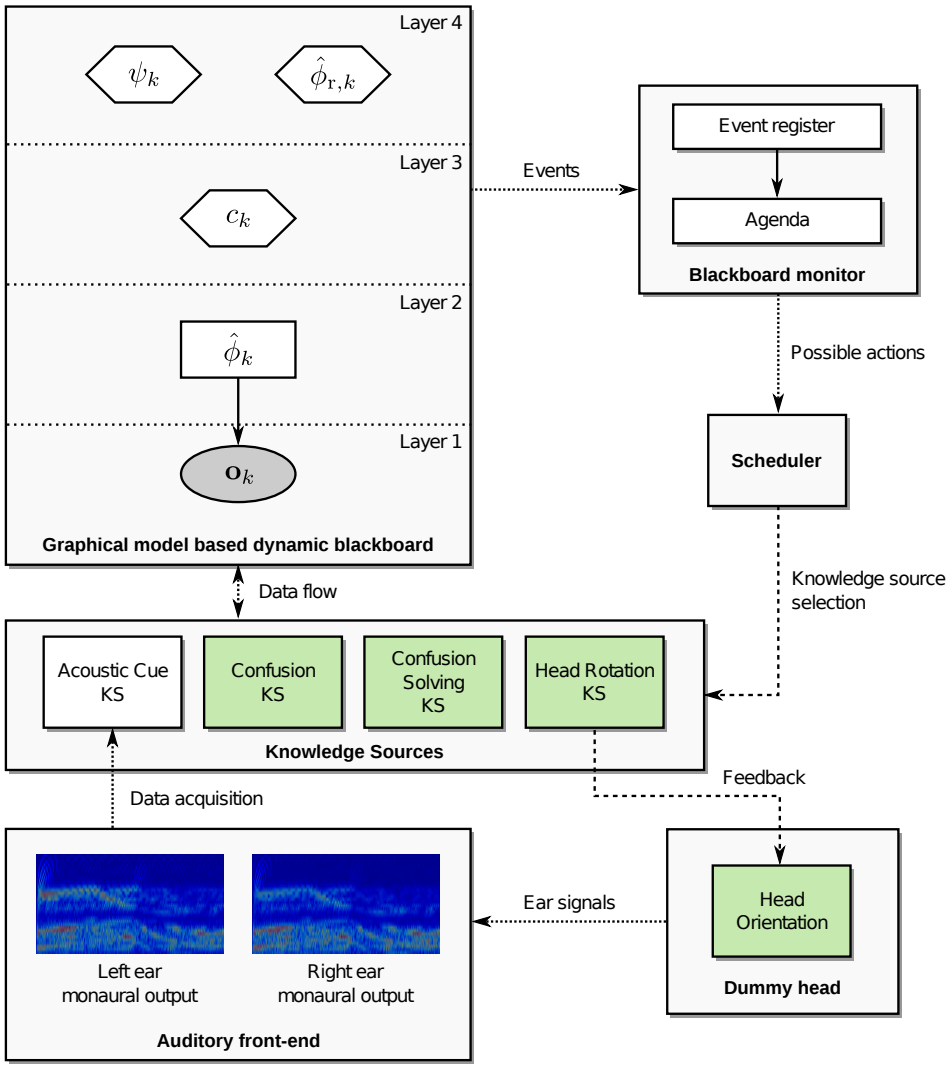
Proof of concept: System architecture



Graphical model for localisation:

- Gaussian-mixture models (GMMs)
- Trained on 340 utterances with 72 different angles (5° increment)
- Training with clean conditions only

Proof of concept: System architecture

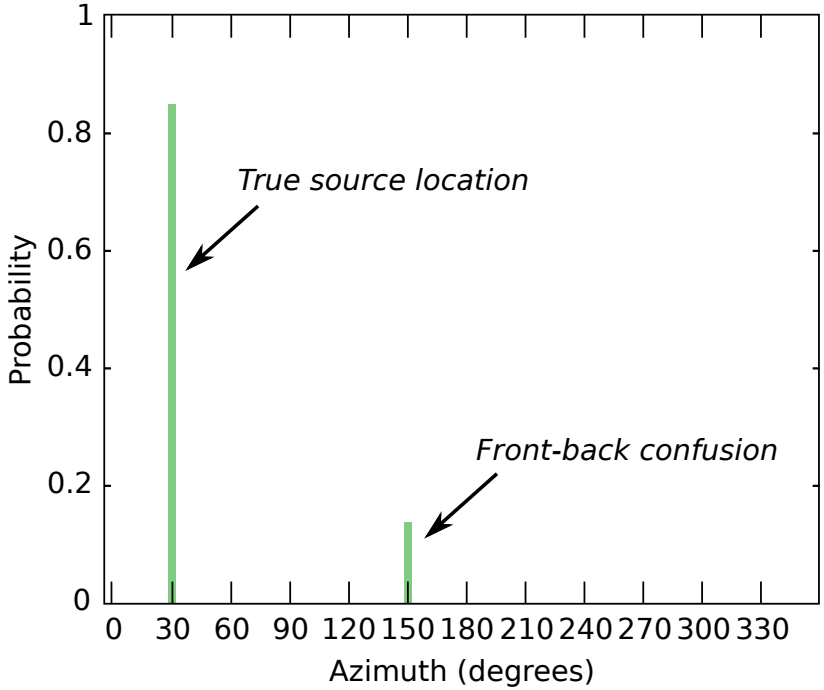


Feedback path:

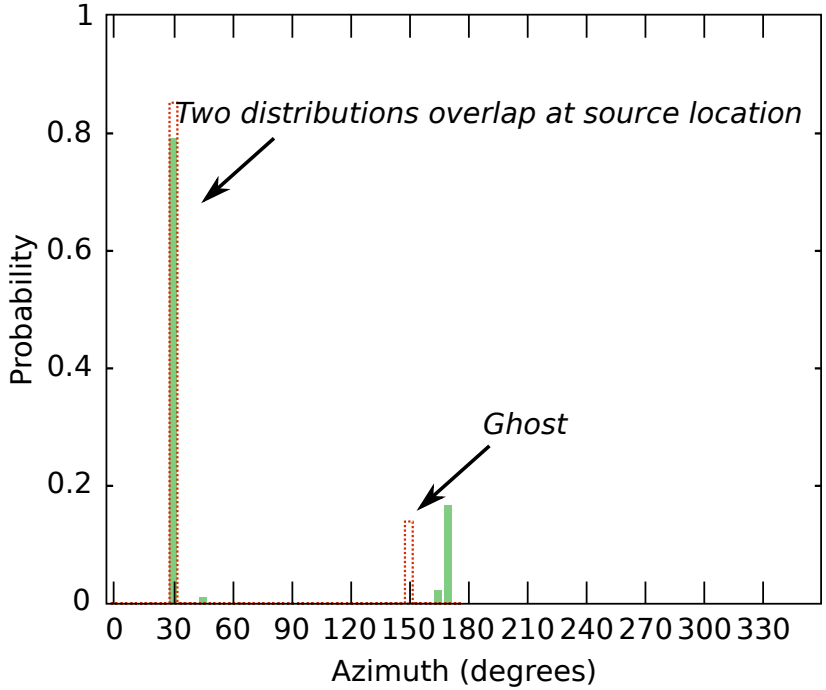
- Head rotations to reduce front/back ambiguities
- Feedback is triggered after evaluating the probabilistic output of the GMMs

Proof of concept: Feedback

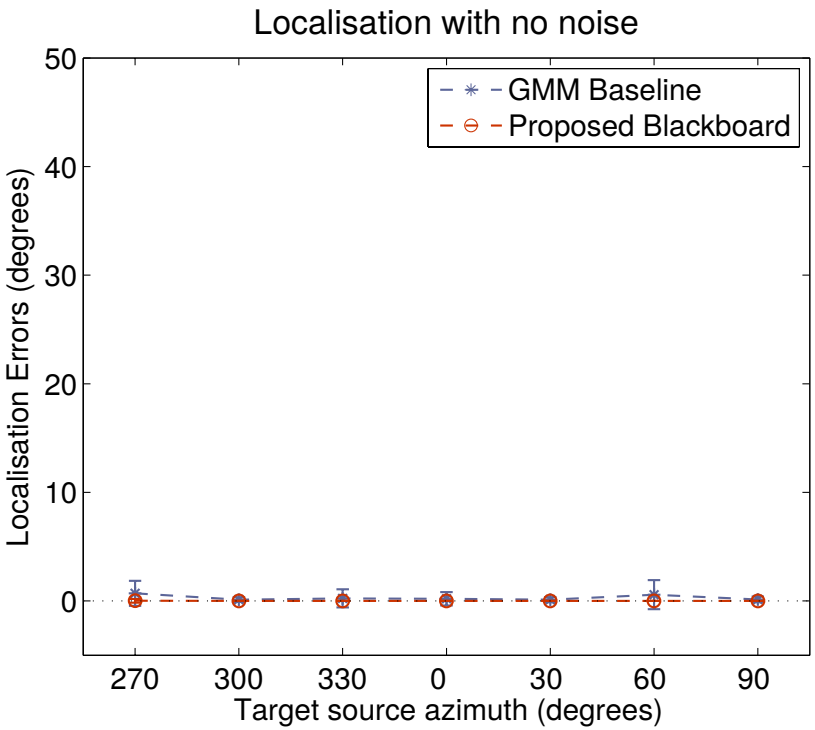
Probability distribution for a single source located at 30 degrees

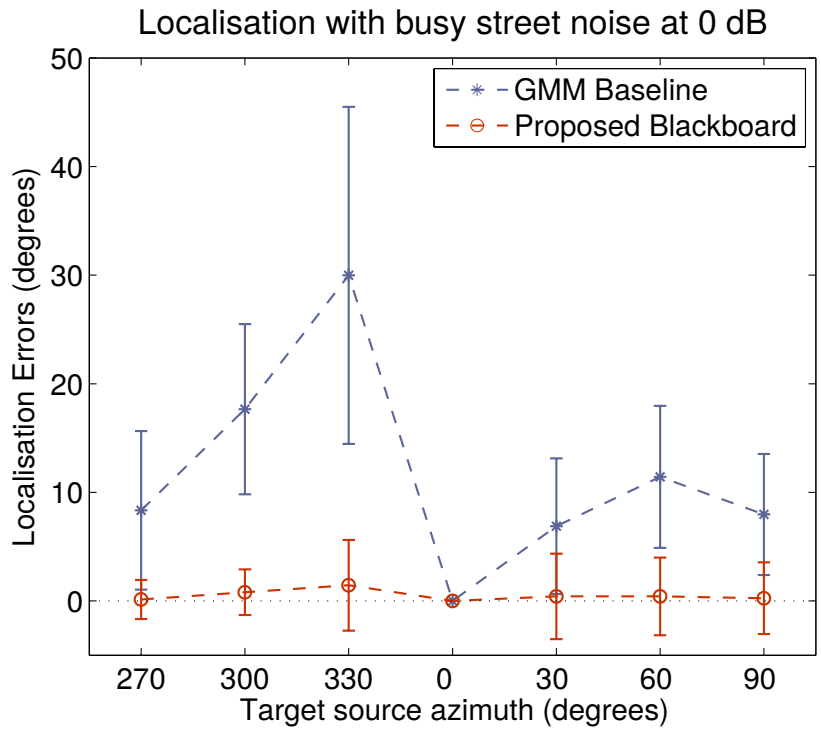


After head rotation of 30 degrees to the right



Results


Localisation with no noise
GMM Baseline
Proposed Blackboard
Localisation Errors (degrees)
50
40
30
20
10
0
270
300
330
0
30
60
90
Target source azimuth (degrees)


Localisation with busy street noise at 0 dB
GMM Baseline
Proposed Blackboard
Localisation Errors (degrees)
50
40
30
20
10
0
270
300
330
0
30
60
90
Target source azimuth (degrees)

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions
- This allows the system to postulate and test hypotheses about the current situation and make decisions accordingly

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions
- This allows the system to postulate and test hypotheses about the current situation and make decisions accordingly

What's next?

- Further improvement of the localisation performance via multiconditional training (ongoing work)

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions
- This allows the system to postulate and test hypotheses about the current situation and make decisions accordingly

What's next?

- Further improvement of the localisation performance via multiconditional training (ongoing work)
- Integration of dynamic state-space models for robust tracking of multiple sources (ongoing work)

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions
- This allows the system to postulate and test hypotheses about the current situation and make decisions accordingly

What's next?

- Further improvement of the localisation performance via multiconditional training (ongoing work)
- Integration of dynamic state-space models for robust tracking of multiple sources (ongoing work)
- Extension of the system to simulataneously perform sound source classification (already working)

Conclusions and future work

- Blackboard Systems provide a flexible computational framework for tasks related to Auditory Scene Analysis
- The integration of Graphical Models allows the generation of „world models“ as probabilistic expressions
- This allows the system to postulate and test hypotheses about the current situation and make decisions accordingly

What's next?

- Further improvement of the localisation performance via multiconditional training (ongoing work)
- Integration of dynamic state-space models for robust tracking of multiple sources (ongoing work)
- Extension of the system to simulataneously perform sound source classification (already working)
- Investigation of additional possibilities for including feedback

Questions?