http://www.twoears.eu/

Listening and Assessing with binaural models

Alexander RAAKE
Assessment of IP-based Applications
Telekom Innovation Labs (T-Labs)
Technical University Berlin, Germany
www.aipa.tu-berlin.de

Jens Bauert
Institute of Communication Acoustics
Ruhr-University Bochum
Germany
www.ika.ruhr-uni-bochum.de

EAA Joint Symposium on Auralization and Ambisonics 2014, Berlin, 3-5 April 2014

Part I: Jens Blauert
Reading the World with TWO!EARS

EU project of the 7th framework program, starting Dec. 1, 2013
FP7-ICT-2011-C (FET open call)

Reading the World with TWO!EARS

Partners

TECHNISCHE UNIVERSITÄT BERLIN

DANMARKS TEKNISKE UNIVERSITET

RUHR-UNIVERSITÄT BOCHUM

UNIV. PIERRE ET MARIE CURIE, PARIS 6

CNRS–LAAS, UNIV. PAUL SABATIER, TOULOUSE

UNIVERSITÄT ROSTOCK

THE UNIVERSITY OF SHEFFIELD

TECHNISCHE UNIVERSITEIT EINDHOVEN

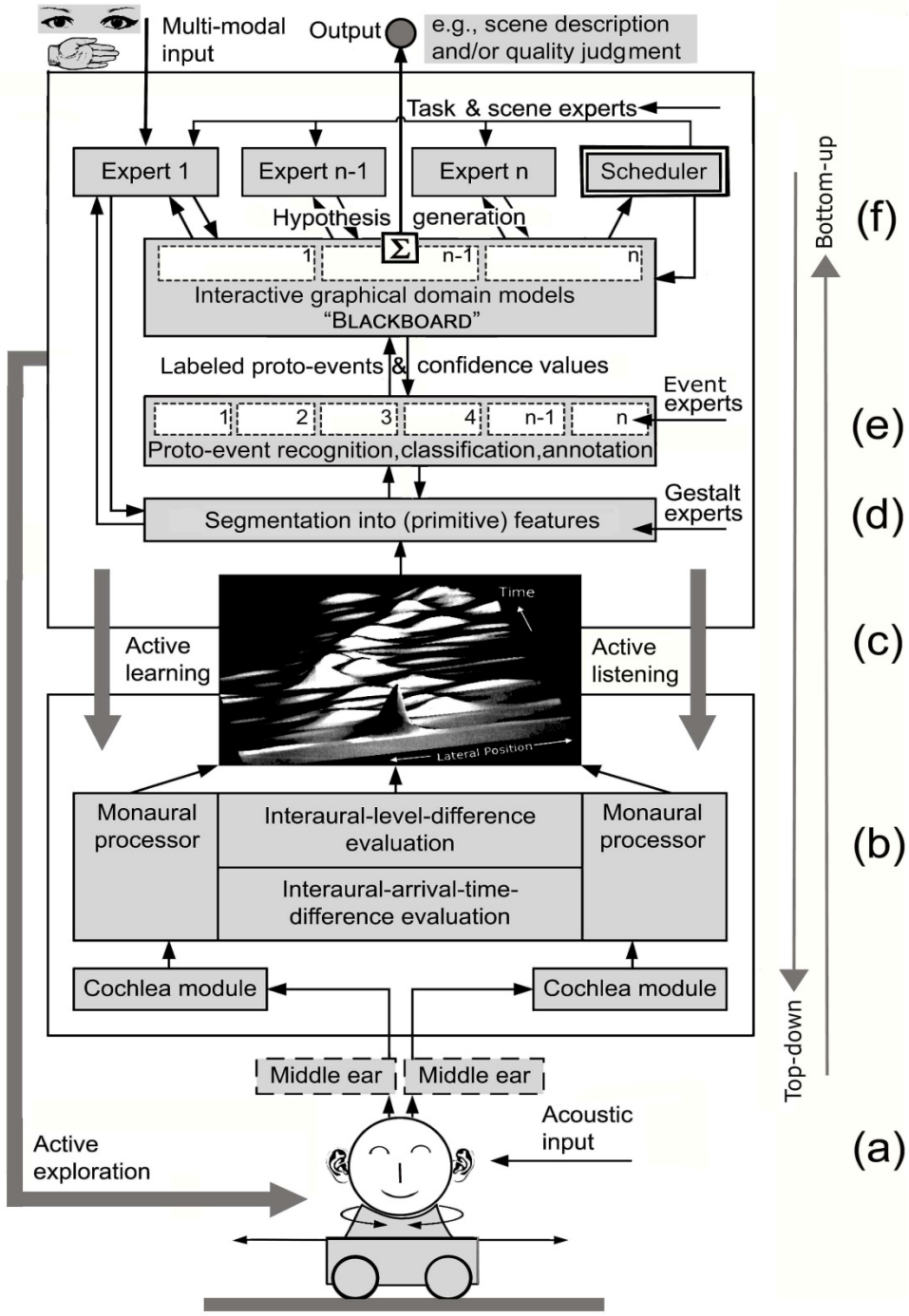
RENSSELAER POLYTECHNIC INSTITUTE, TROY NY

TWO!EARS

An intelligent, active computational model of human auditory listening and experience in a multimodal context

Unique features

- Modeling listeners as multimodal agents that develop their concept of the world by exploratory interaction
- Structural link from binaural perception to judgment and action, realized by interleaving bottom-up and top-down processes, including inference from domain-specific expert systems
- Meaning assignment by combining signal and symbol processing in a joint model structure, integrating visual and proprioceptive cues
- Robotic-platform front-end that actively parses its physical environment, orients itself and moves its sensor in a humanoid manner
- Open, modularized architecture that can easily be modified and extended – also by modules from other research groups



World model, meaning assignments, judgments,

Proto-event recognition, object building

Segmentation into features

Binaural-activity map

Auditory signal processing

Physical front end

TWO!EARS

platform for translatory movements, with head panning and tilting

TWO!EARS

A Couple of Hypotheses (1)

- In order to *understand* acoustic scenes in any sense of the word, it is necessary to match them with an internal world model

- This can in turn enable an adaptation of the world model to match more closely what is observed—it might even allow the construction of a world model purely from observations („learning structure from data“, e.g., Heckermann 1995, Schulte 2010)

after D. Kolossa, IKA Bochum

TWO!EARS

A Couple of Hypotheses (2)

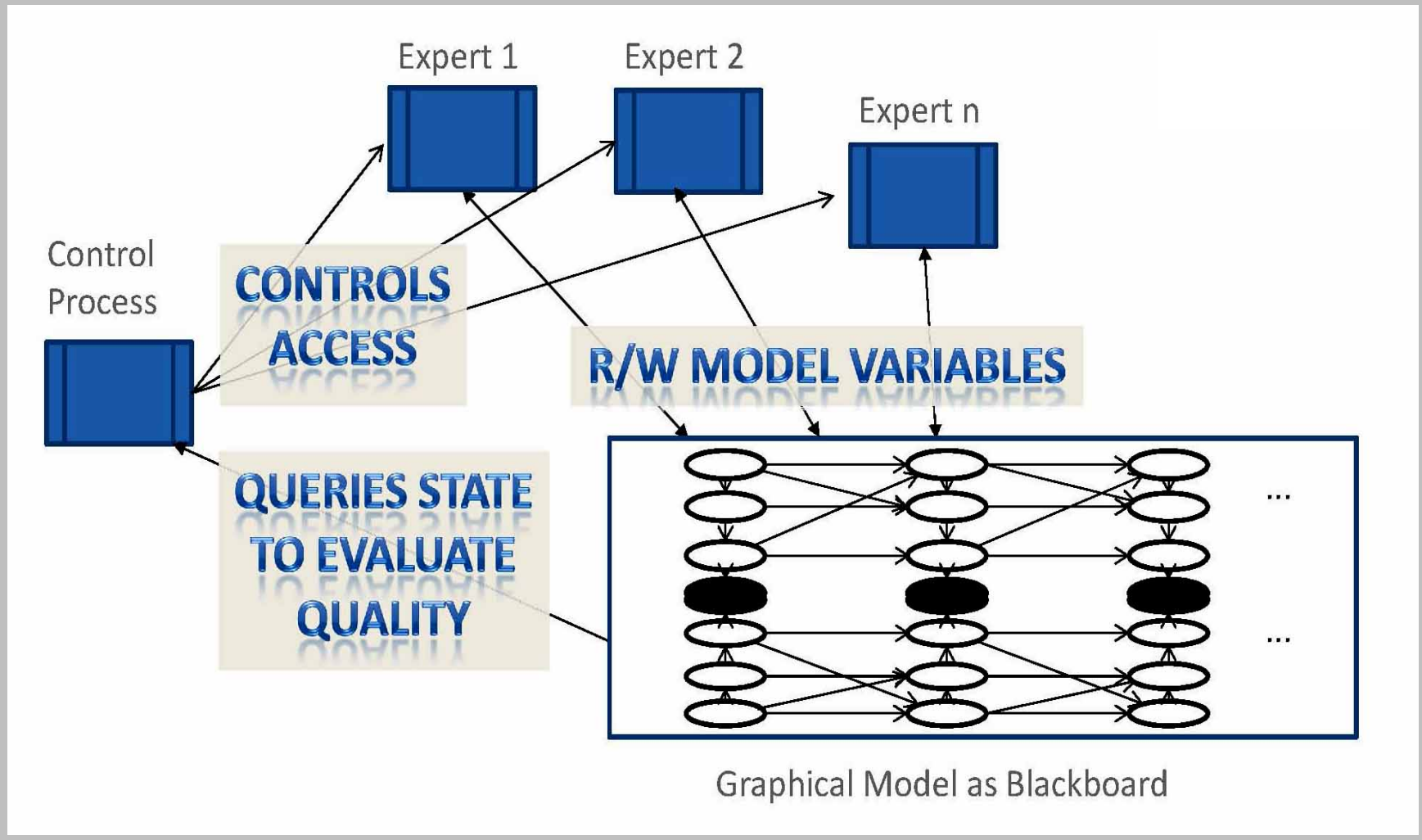
- A world model should be able to use *prior knowledge* and *data* to analyze a temporally evolving environment
- *Prior knowledge* can be statistical (conditional probabilities, discrete or continuous-valued) and rule-based (e.g. lexical knowledge, acoustic wave propagation)
- The *Data* of interest can be fully known or partially observable

Strategy

- Use joint statistical and rule-based representation of the (auditory) world in form of a *dynamical Bayesian network*
- Match network with observations by statistical inference, allowing external control of hypothesis- and/or rule-guided search

after D. Kolossa, IKA Bochum

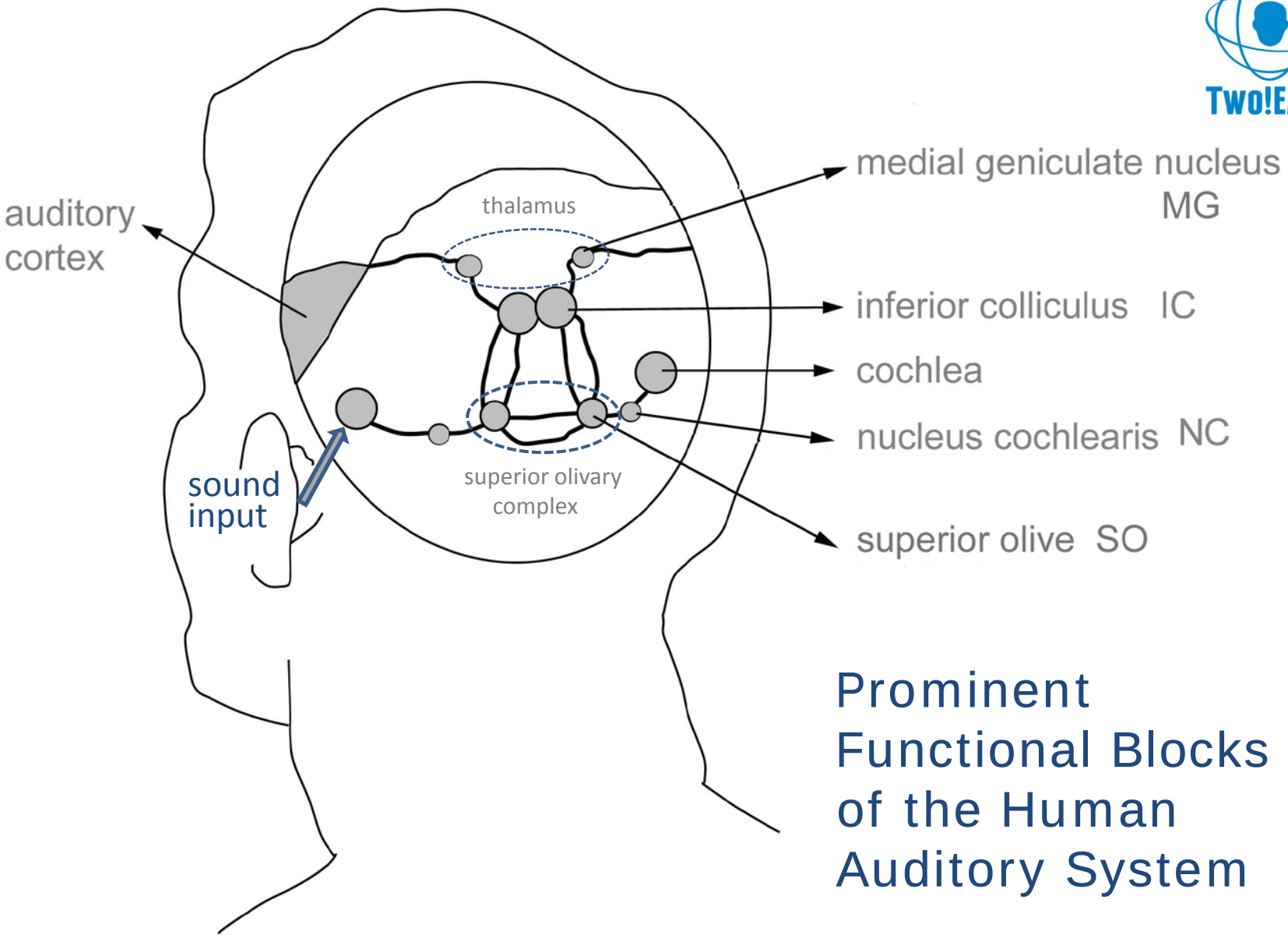
TWO!EARS



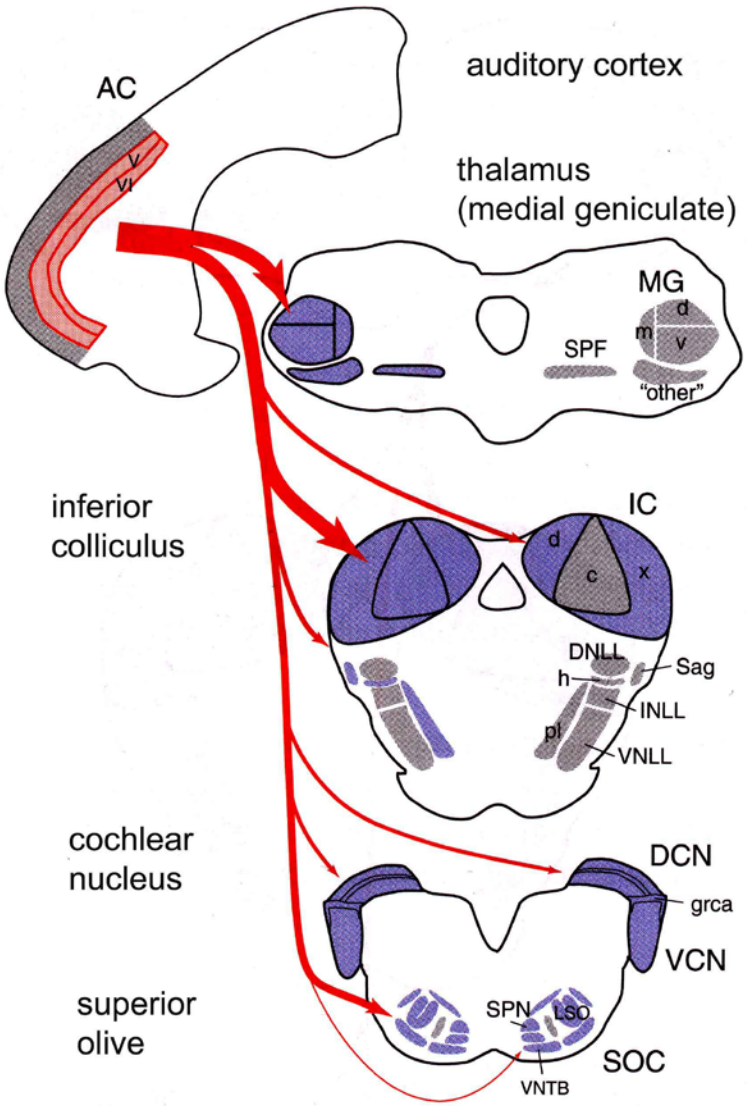
Architecture for the World Model

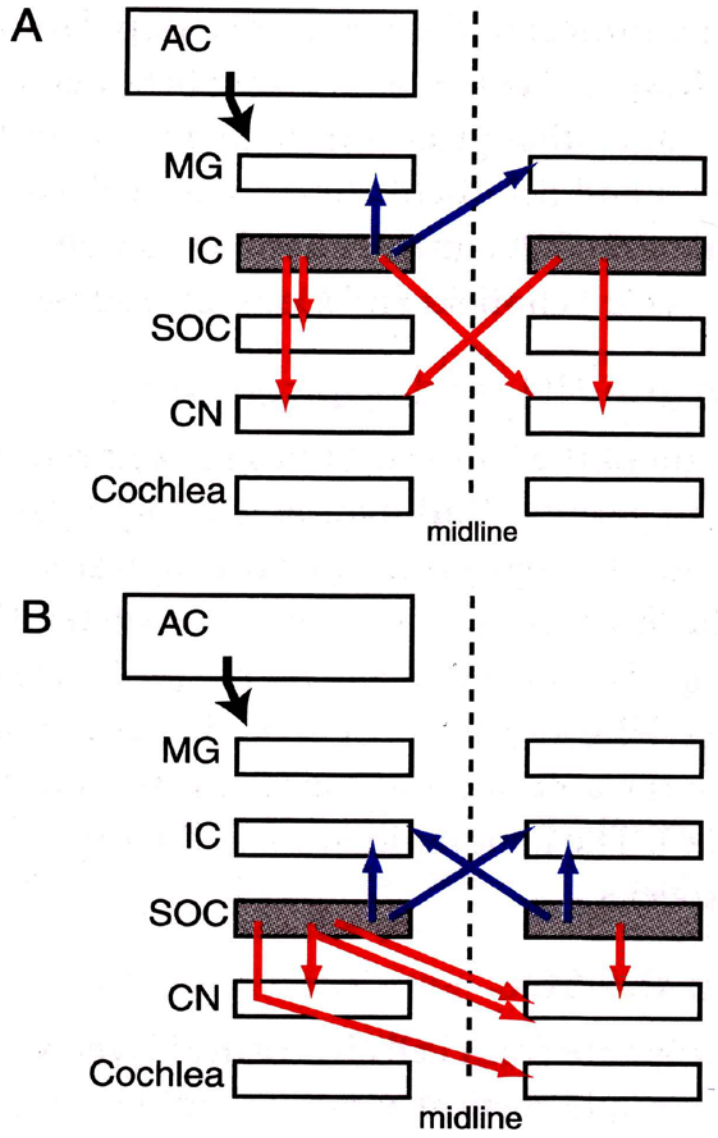
after D. Kolossa & Chr. Schymura, IKA Bochum

TWO!EARS



Prominent Functional Blocks of the Human Auditory System





From: Schofield 2010, Oxford Handbook of Auditory Science, Vol. 2

- Where are uncertainties and/or doubts in the results?
  - at the signals level: ⟶ Variances too high
  - at the symbolic level: ⟶ Logical inconsistencies
  - at the level of meaning: ⟶ Implausibilities
- Which parameters can be manipulated/tuned/tweaked?
- Which additional cues can be provided?

Active Listening, Feedback-Loops and Integration of Crossmodal Information

TWO!EARS

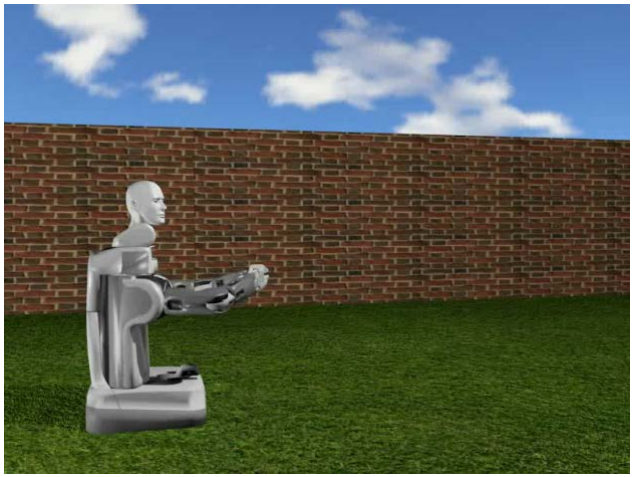
- Turning the acoustic sensors into optimal position (turn-to reflex)
- Advanced exploration of the environment by active head-&-torso movements
- Increasing signal-to-noise ratio by specific enhancement of spectral and temporal selectivity
- Paying attention to specific signal features to deliver additional information as required by the cognitive stage
- Activation of specific signal-processing procedures, such as echo cancelling, de-reverberation, precedence-effect preprocessing, reconsideration to solve ambiguities
- Improvement of object recognition, auditory grouping, aural-stream segregation, aural-scene analysis
- Improvement of scene understanding, assignment of meaning, quality judgments, attention focusing

Expected Functional Improvements Due to Auditory Feedback

TWO!EARS

1. Search & Rescue Mission

→ sketch by Th. Walther



2. Sound-Quality Assessment

→ as to which Alexander Raake will now continue this talk

Two Proof-of-concept Applications for System Demonstration

TWO!EARS

Part II: Alexander Raake
Assessing with TWO!EARS

TWO!EARS applications & proof of concept (1)

- **Reminder: Goal of project is "Reading World with TWO!EARS"**
- **Components of reading**
  - Basic scene analysis
  - Exploration
  - Extracting meaning of scene & objects
  - Highest level of meaning extraction:
    Evaluation beyond signal-level / form → Quality of Experience
- **Two proof-of-concept areas to reflect goals**
  - Dynamic auditory scene analysis – search & rescue
  - Quality of Experience assessment
- **Evaluate performance improvement for these two tasks with different choices of model components & functionalities**

TWO!EARS

TU berlin

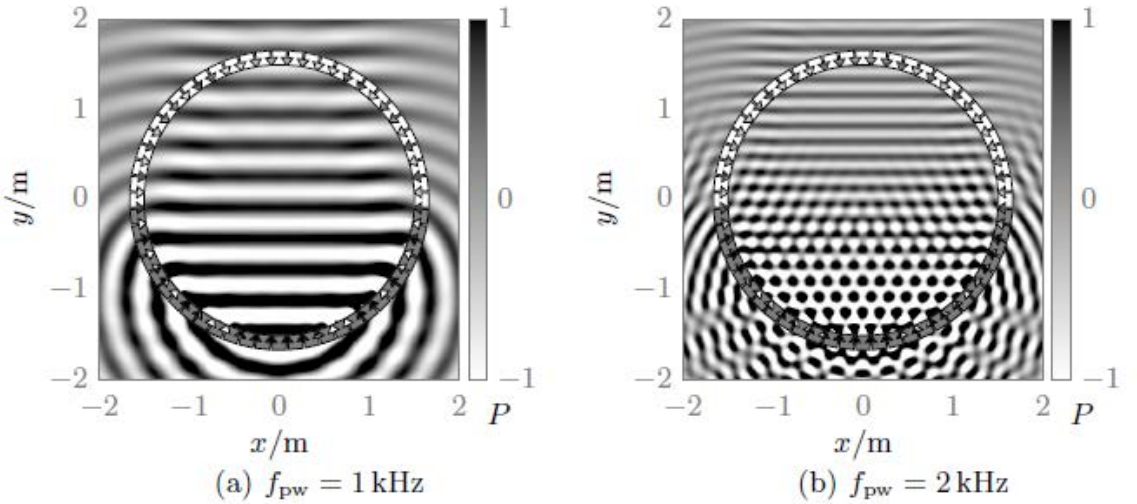
TWO!EARS applications & proof of concept (2)

Dynamic auditory scene analysis – search & rescue

- Audio SLAM (simultaneous localization and mapping)
- Speaker identification
- Keyword-type speech recognition
- Relevance identification
- Coarse audio-type identification

Quality of Experience

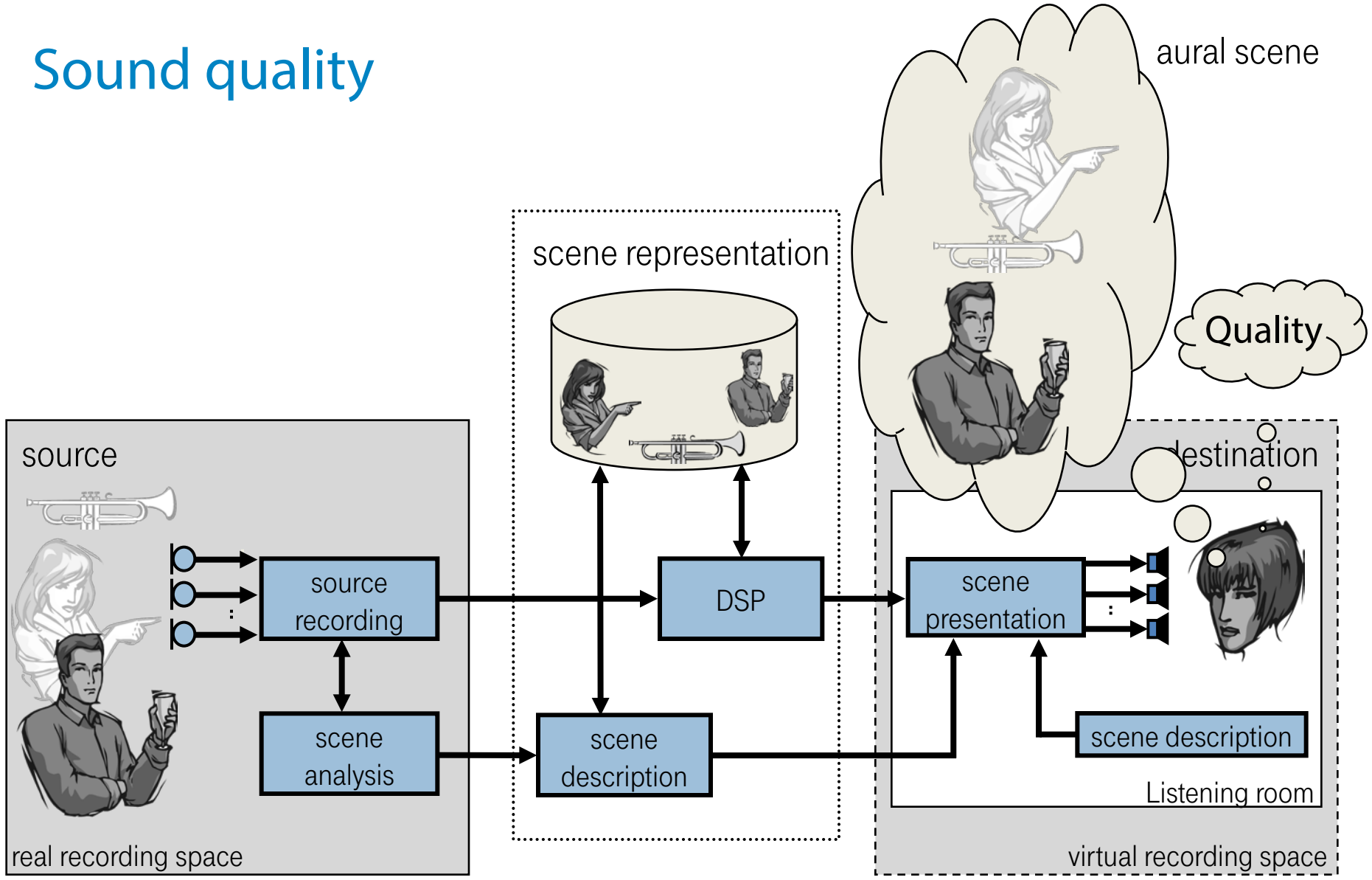
- Applied to multi-loudspeaker audio reproduction
- Active exploration of listening area
- Internal reference
- Meaning assignment



(a) $f_{\mathrm{pw}} = 1\,\mathrm{kHz}$ (b) $f_{\mathrm{pw}} = 2\,\mathrm{kHz}$

(Wierstorf et al., 2014)

Sound quality



(Raake & Spors 2006, Spors et al., 2013)

Overview

- Introduction
- SoA: Sound reproduction quality
- Model & application to sound quality assessment
- Outlook

Plausibility vs. authenticity

Authentic

- Indiscernible from an explicit or implicit reference
- True to the original
- Linked with concept of **fidelity**

Plausible

- Perceived features of the reproduced scenes show believable & creditable correspondence with the listeners' expectations in given context
- Not necessarily authentic
- Linked with concept of **QoE**
  - *But two plausible scenes may lead to quite different QoE*
- Jekosch: Not plausible if sth. sounds "strange"

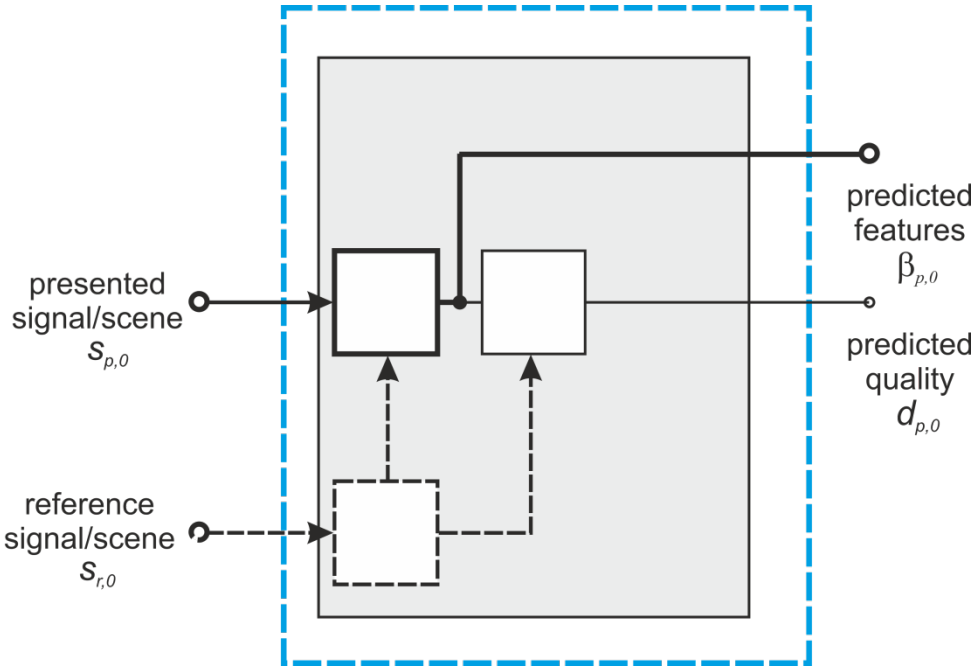
Evaluation criteria

QoE

- Immersion
- Envelopment
- **Quality**
- Attributes of quality
- Preference
- ...

Sound quality models
State of the Art

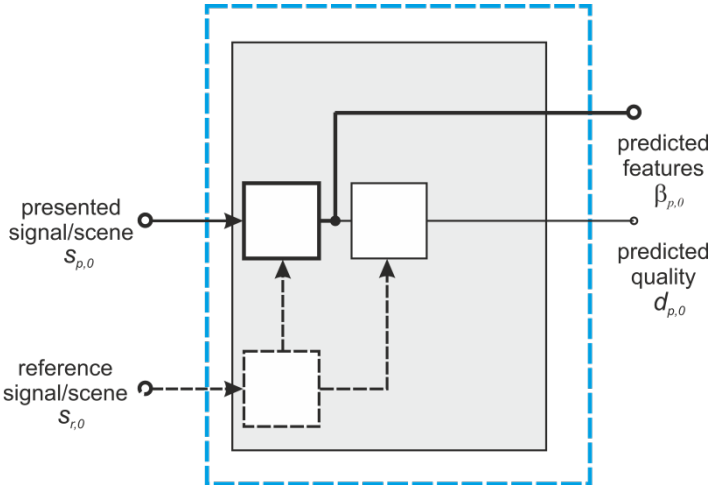
- Criterion: Quality!
- Considered model types
  - Full-reference (FR)
  - No-reference (NR)
- Speech & audio
- Signals transformed into perceptual domain using models of sub-cortical auditory system
  → bottom-up processing
- Key problems
  - Models require explicit reference, world knowledge of listeners not reflected
  - No top-down (feedback) information
  - No exploration



TWO!EARS

Speech and audio transmission quality

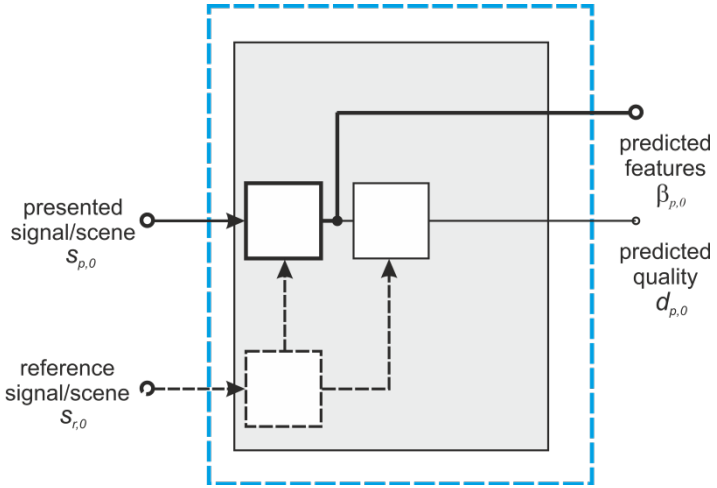
- Speech: Perception-based
  - PESQ (ITU-T Rec. P.862, 2001)
    → Explicit comparison with reference
  - POLQA (ITU-T Rec. P.863, 2011)
    → Idealization of reference(!)
- Speech: Perception-based, perceptual dimensions
  - Dimension-based (Wältermann et al., 2012)
    → Muldtimensional analysis of quality, distance from ideal point (≡ reference)
  - DIAL (Côté et al., 2012)
    → Perceptual model with explicit reference + quality dimensions
- Audio: Perception-based feature model
  - PEAQ (Thiede et al. 2000)
    → Individual features based on psychoacoustic model oncl. spectro-temporal analysis for coded audio & reference



TWO!EARS

Multichannel Audio

- Freqeuently: Technical or physical characteristics of acoustic scene mapped to low-level attributes / perceptive constructs
- Quality =f(spatial & timbral fidelities, artefacts) (Rumsey et al., 2004, 2005)
- Stereophonic: 70% spectral, 30% spatial (Rumsey 2005)
- Spatial fidelity: Wierstorf et al. 2012, 2013 based on Lindemann 1986, Dietz 2008, QUESTRAL (Rumsey et al., 2008)...

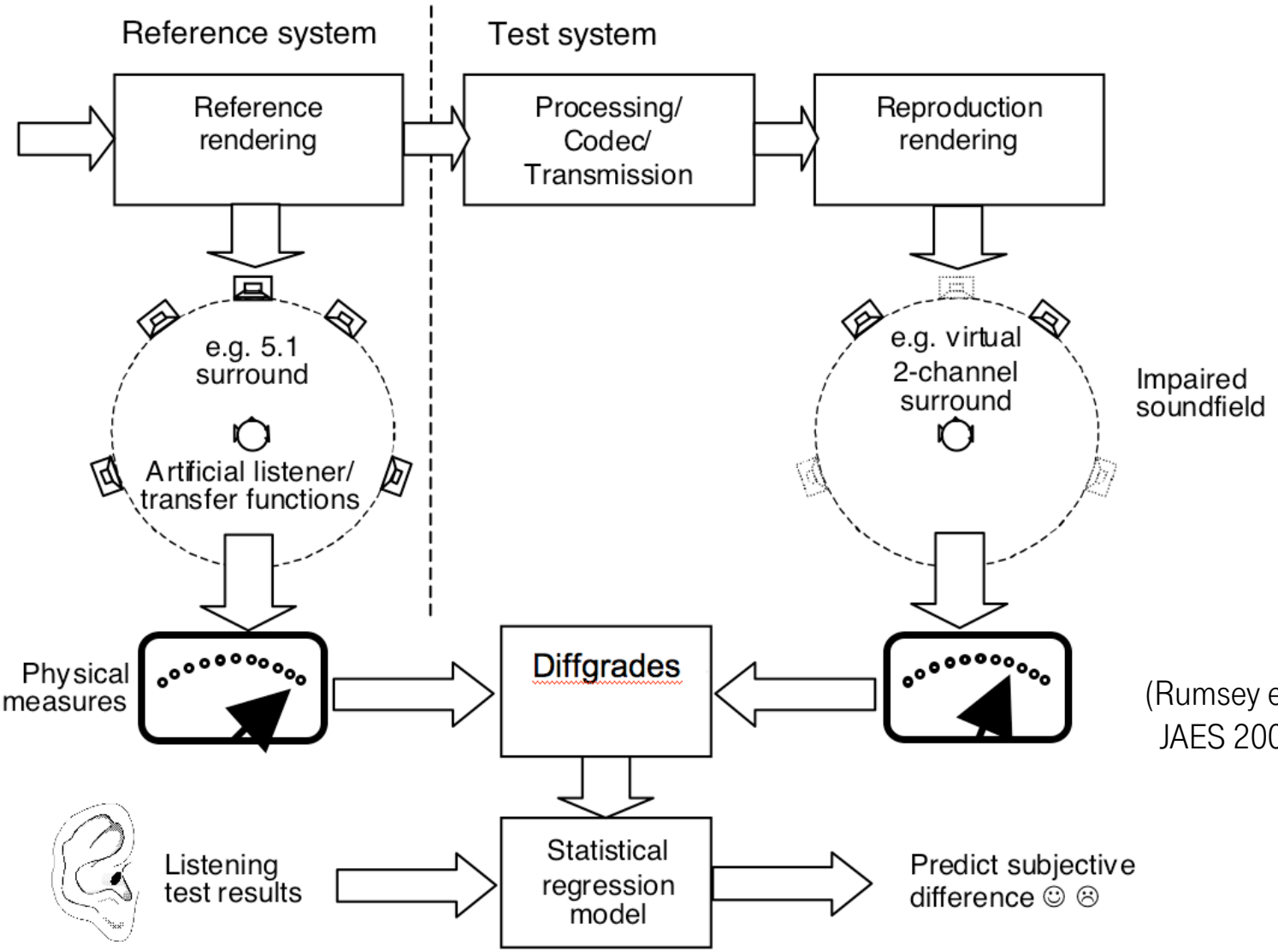


- Timbral fidelity: Pulkki 2001, Brüggen 2001, Moore & Tan 2004, Raake 2006, ...
- Artefacts: ???
- Multichannel audio reproduction: Models still under development by ITU-R SG 6 (cf. e.g. Liebetrau et al., 2010)
- Problem: All target fidelity

TWO!EARS

TU berlin

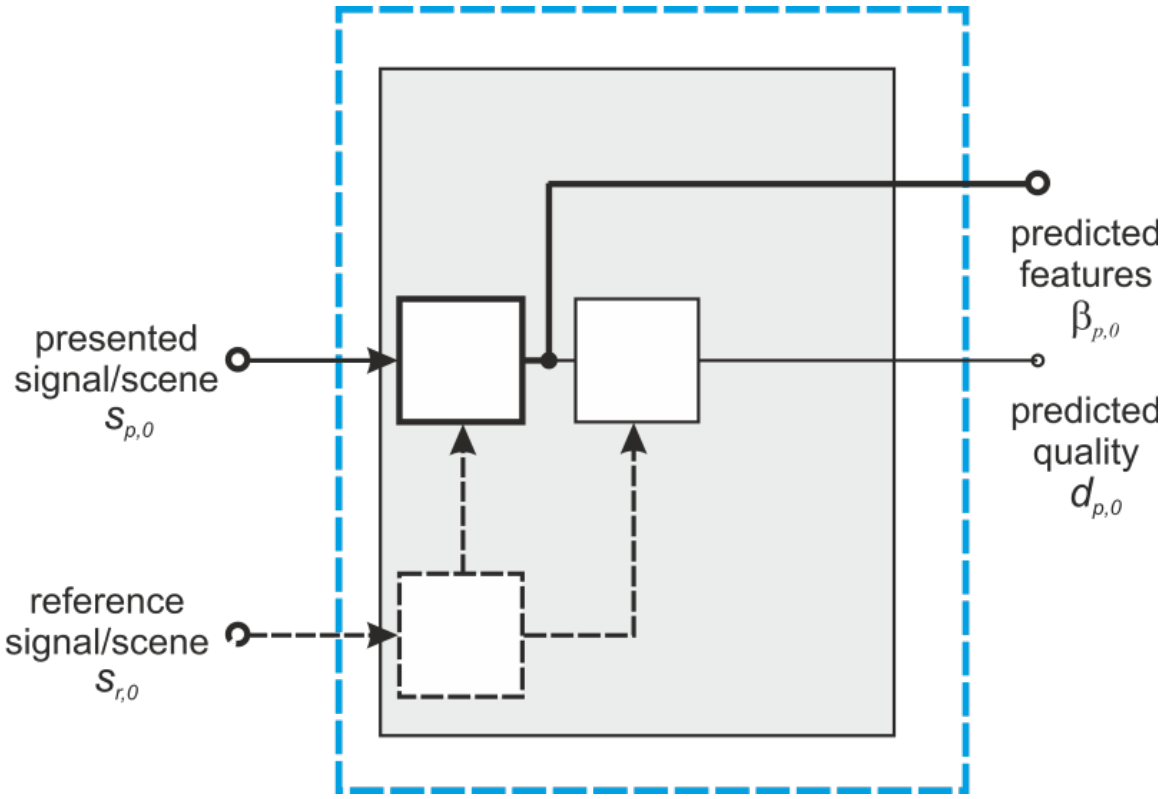
Example: QESTRAL
Spatial fidelity



Fidelity!

(Rumsey et al., JAES 2008)

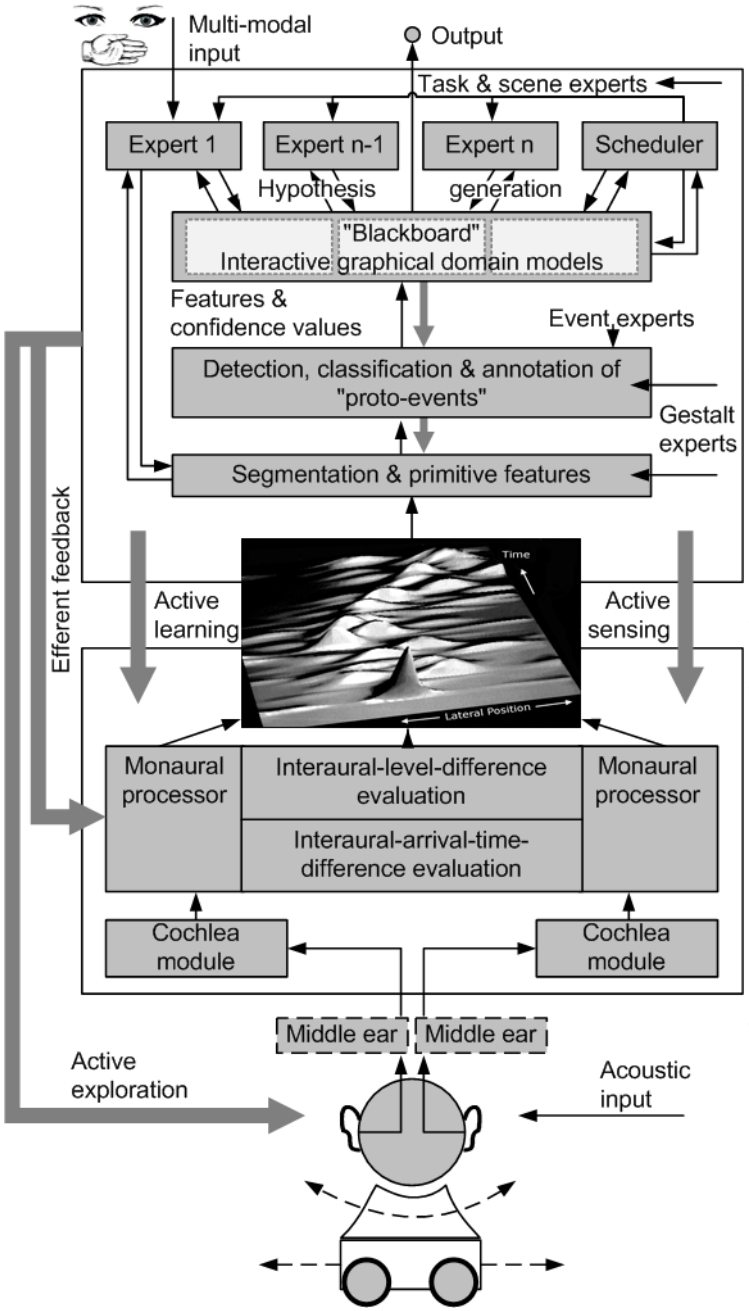
Remaining issues



Overview

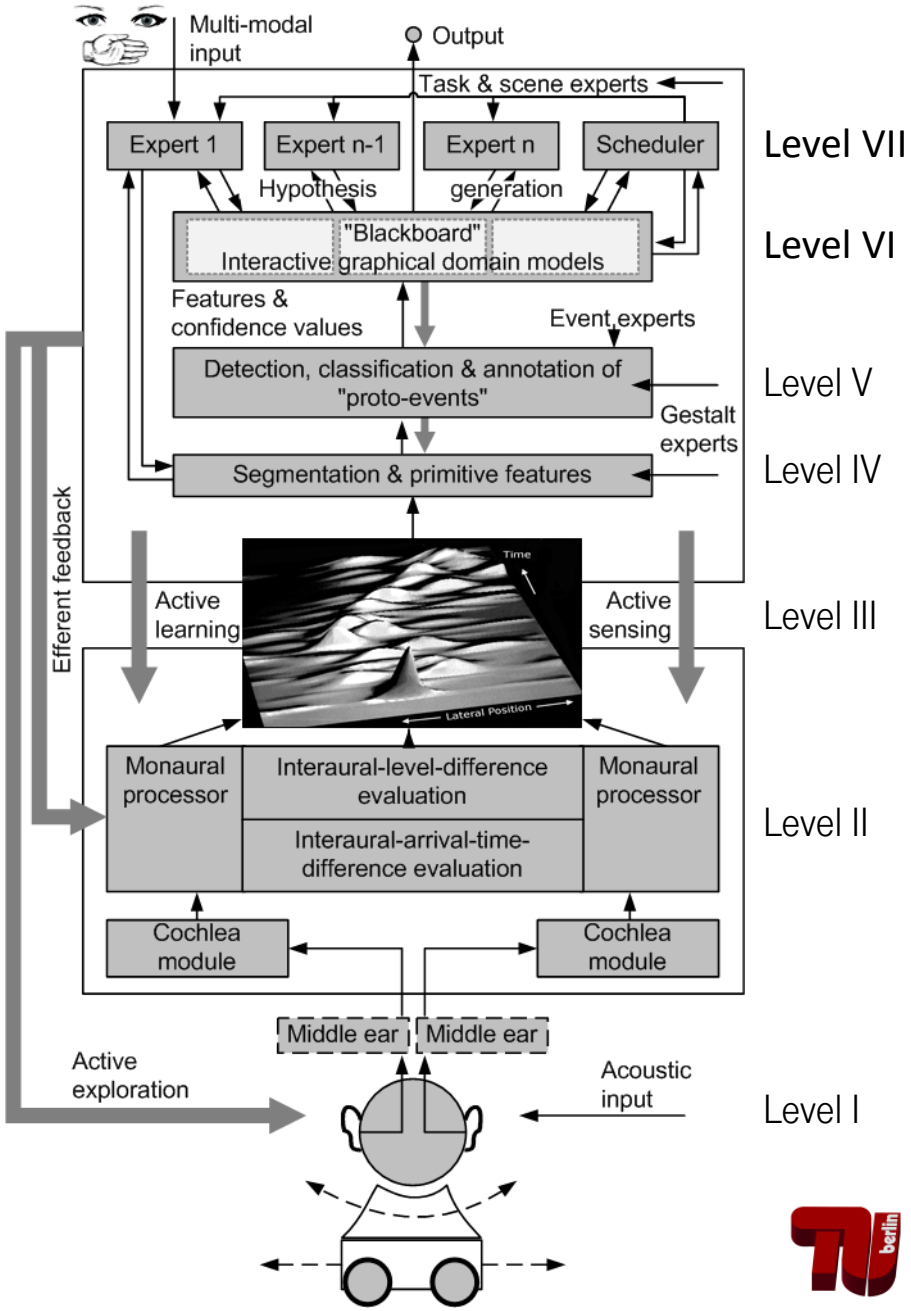
- Introduction
- SoA: Sound reproduction quality
- Model & application to sound quality assessment
- Outlook

TWO!EARS-Model



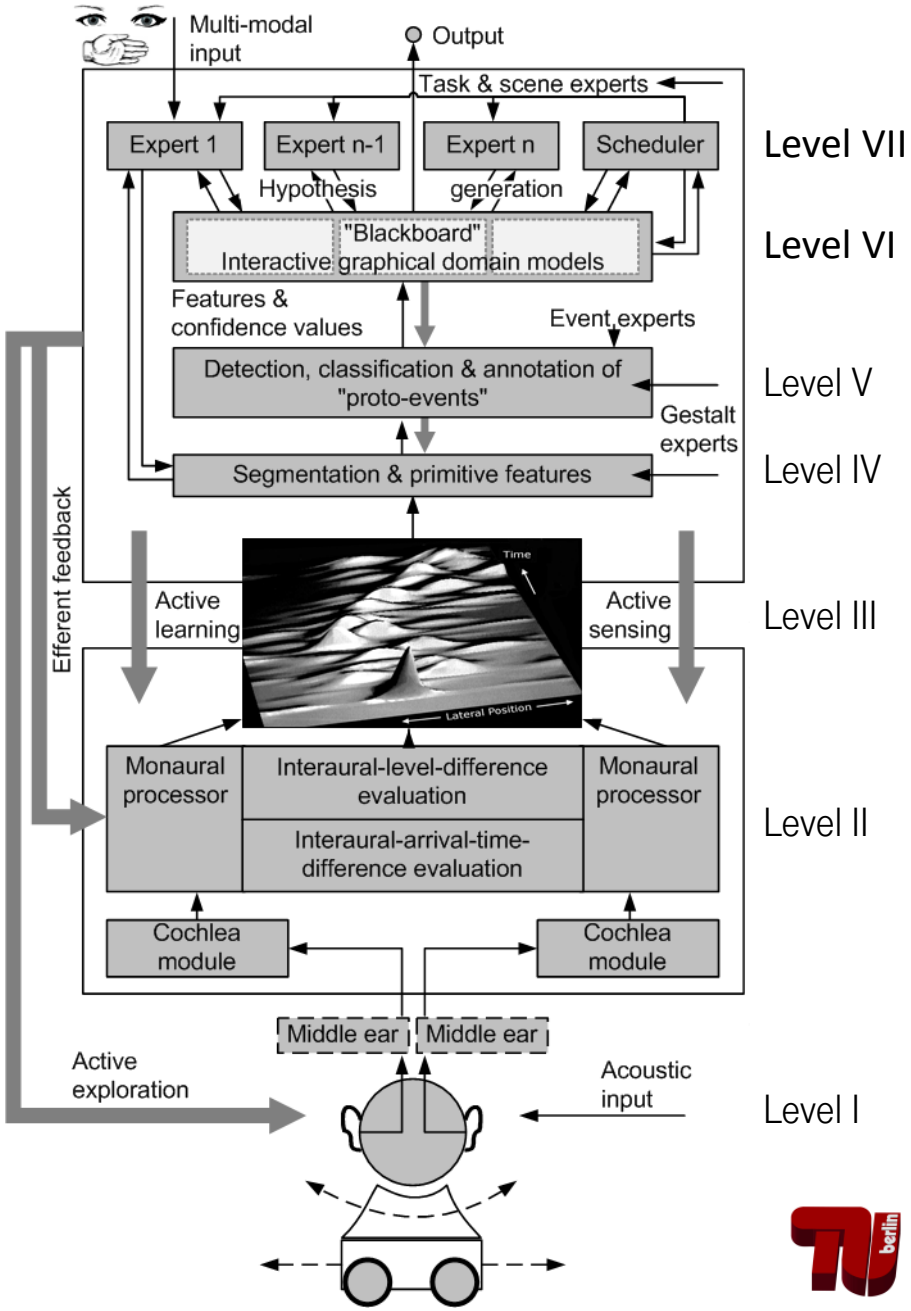
Quality model

- Layers II to III: Comprehensive set of perceptual features
- Layers II-V: Analysis of auditory scenes and objects
  - Object- and content-related information
  - Both Full-reference (FR) and No-reference (NR) models enabled
  - NR: Learned conceptual references (system-inherent) → machine-learning techniques
    - *“Quality-” & “Feature-” experts*.
  - FR: Adaptations of references in top-down manner using expert-system input
- System capable of exploratory movements

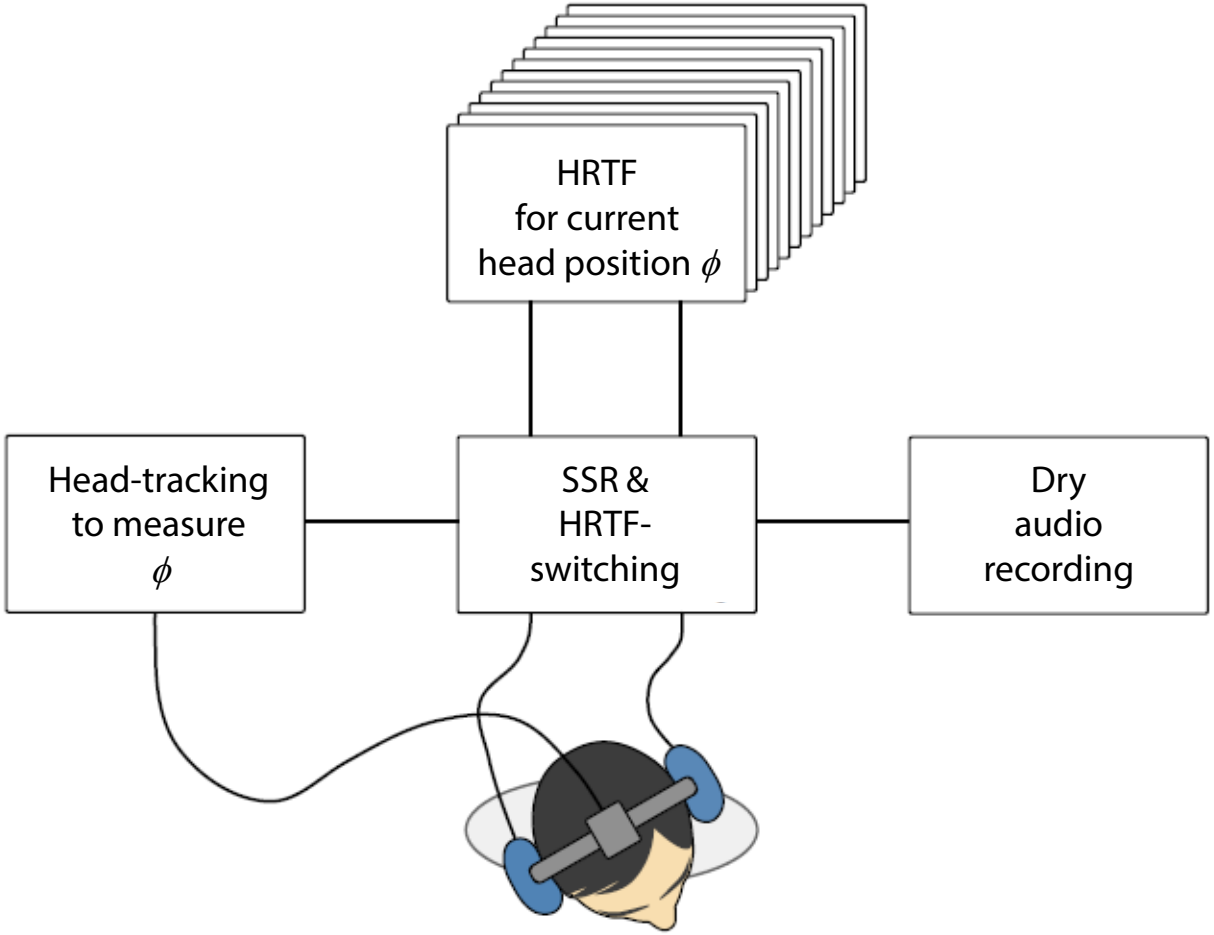


First work

- Braasch et al. (in Blauert, Springer 2013): Binaural model endowed with head movements for localization disambiguation
- Wierstorf, Spors, Raake (ongoing):
  - Wave Field Synthesis (WFS) & Higher Order Ambisonics (HOA)
  - Localization/coloration tests
  - Binaural models (e.g. AABBA's "auditory toolbox", Søndergaard et al., 2011-13)
  - Ongoing work: Binaural model equipped with sensorimotor information → head-orientation
  - Target: Improve predictions for sound field synthesis evaluation
  - Recordings of head-tracking data in listening tests



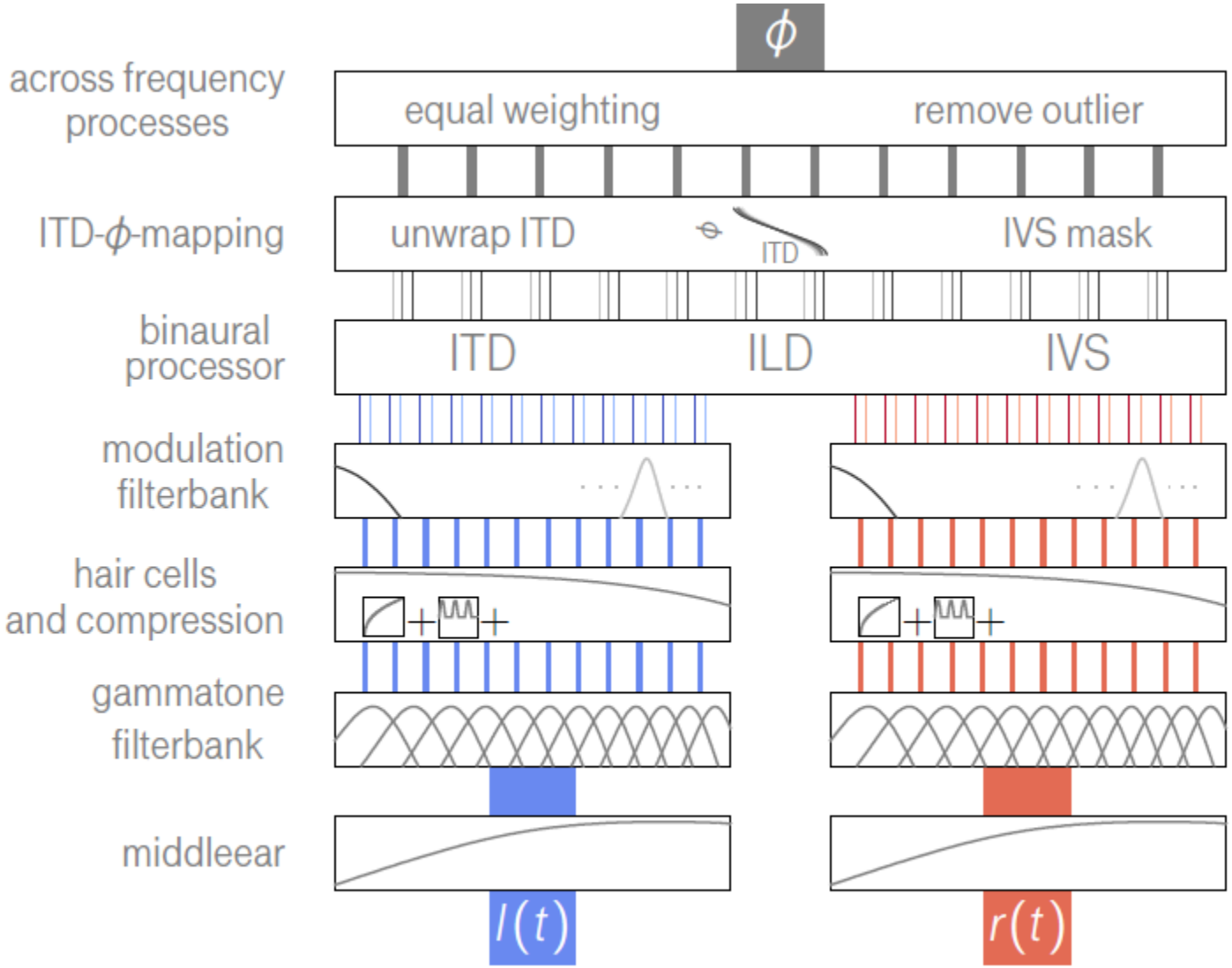
Localization test: Dynamic binaural re-synthesis



(Wierstorf et al., 2014)

Simulated loudspeakers

Localization test: Binaural model

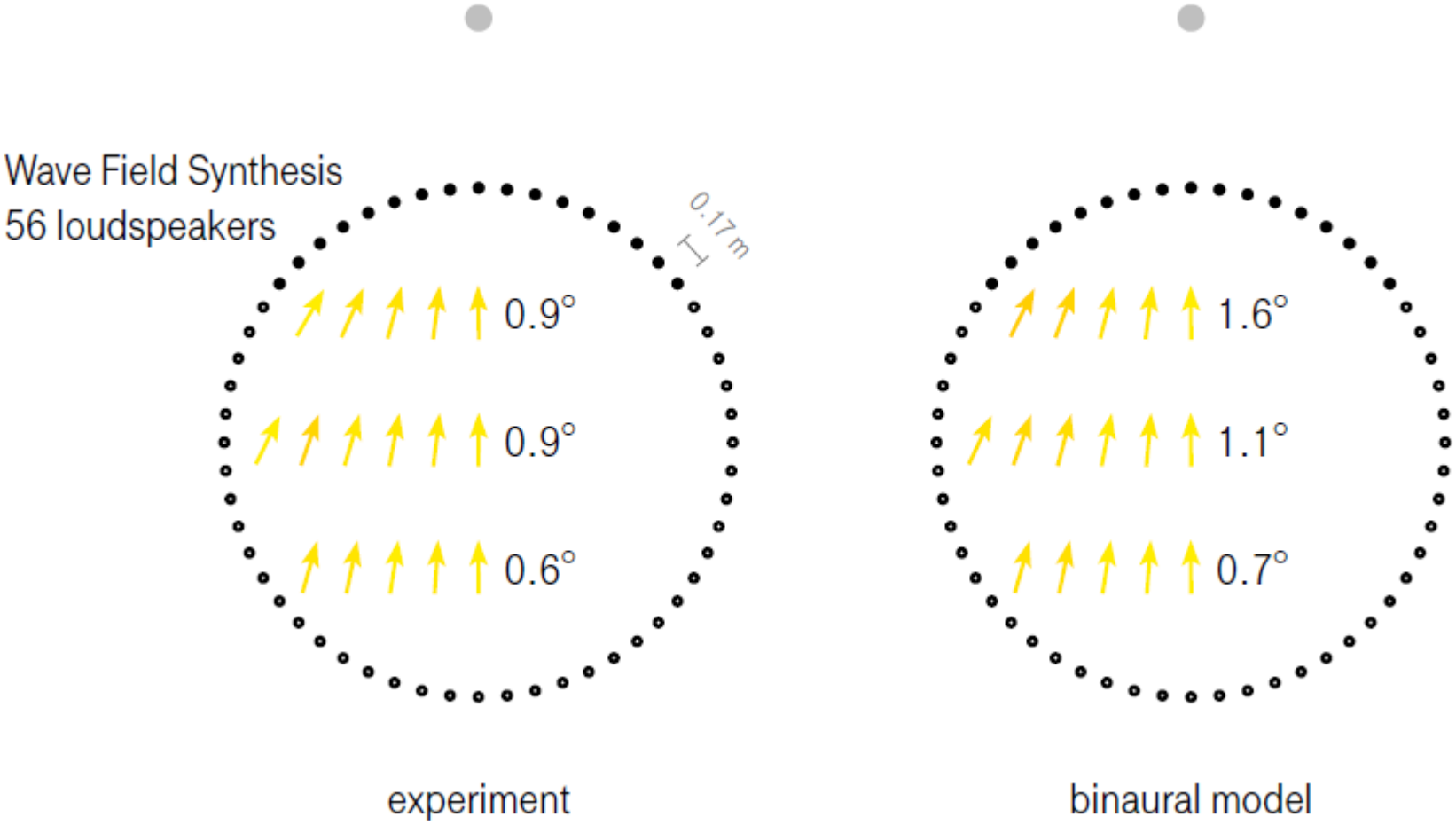


Extensions to Dietz et al. (2011)

(Wierstorf et al., 2014)

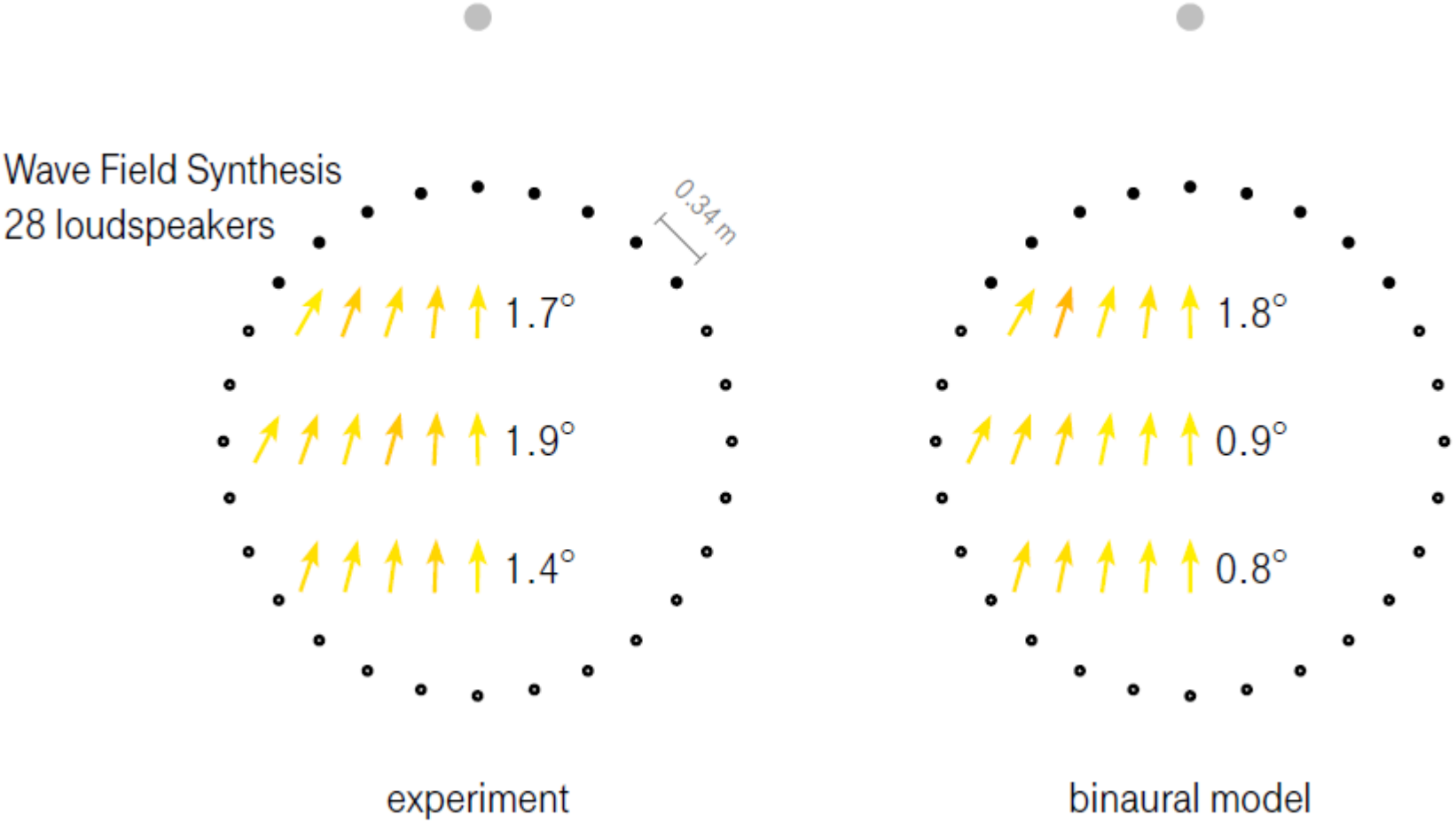
http://amtoolbox.sourceforge.net/

Localization test & model
WFS



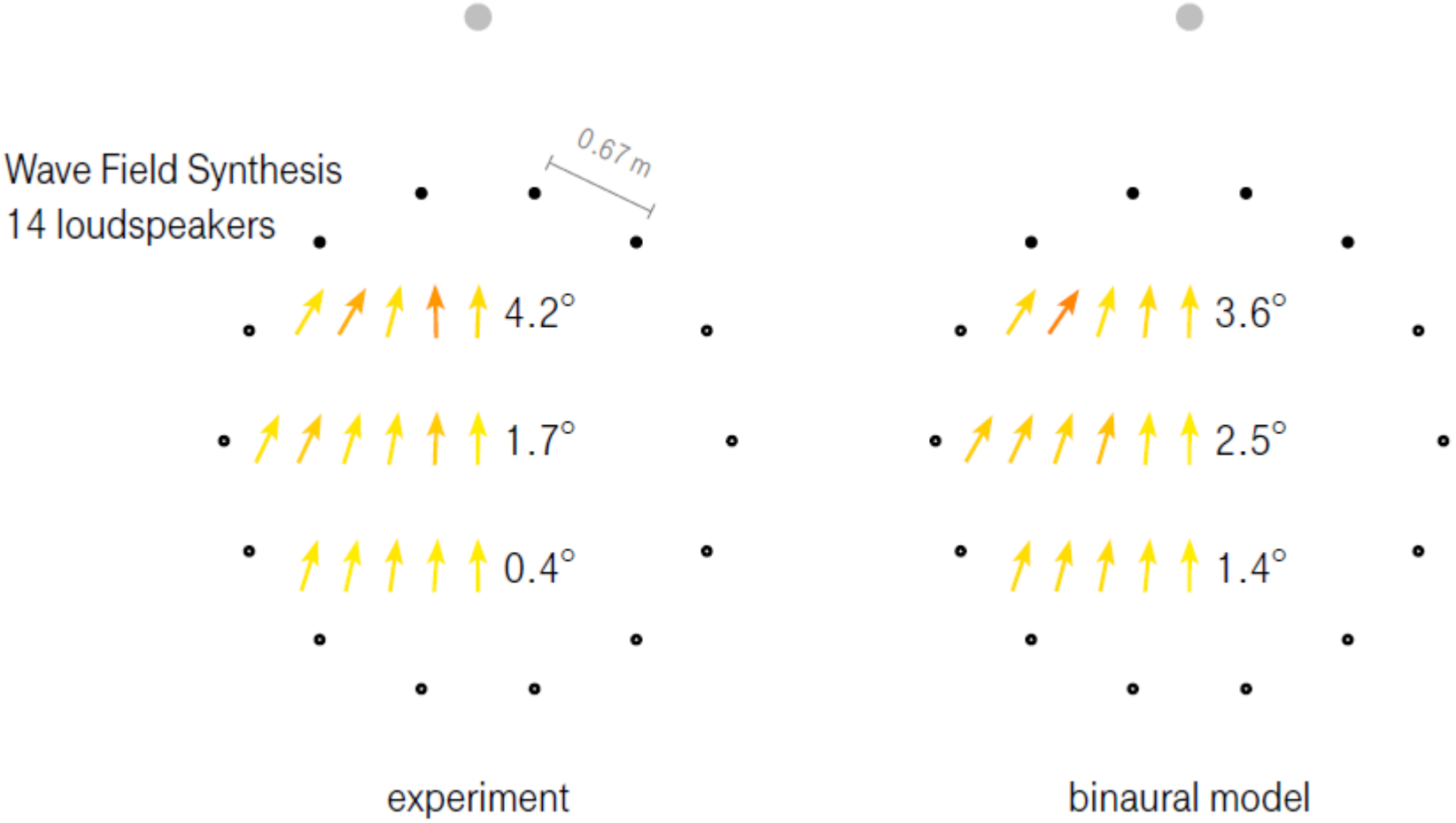
(Wierstorf et al., 2014)

Localization test & model
WFS



(Wierstorf et al., 2014)

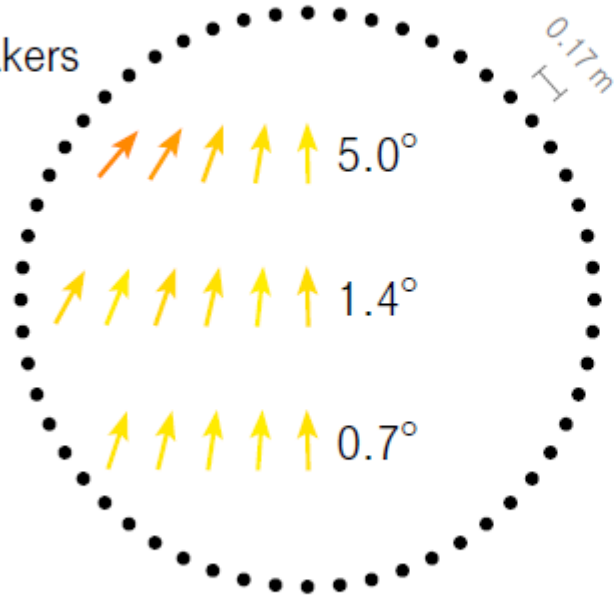
Localization test & model
WFS



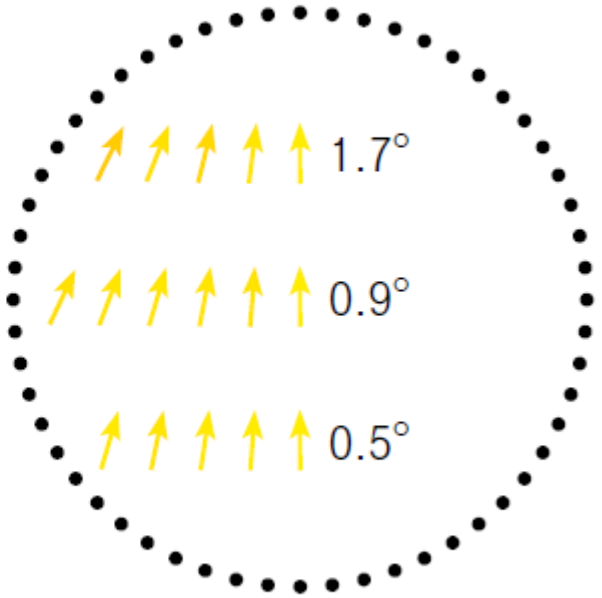
(Wierstorf et al., 2014)

Localization test & model
Ambisonics

Ambisonics
56 loudspeakers
$M = 28$



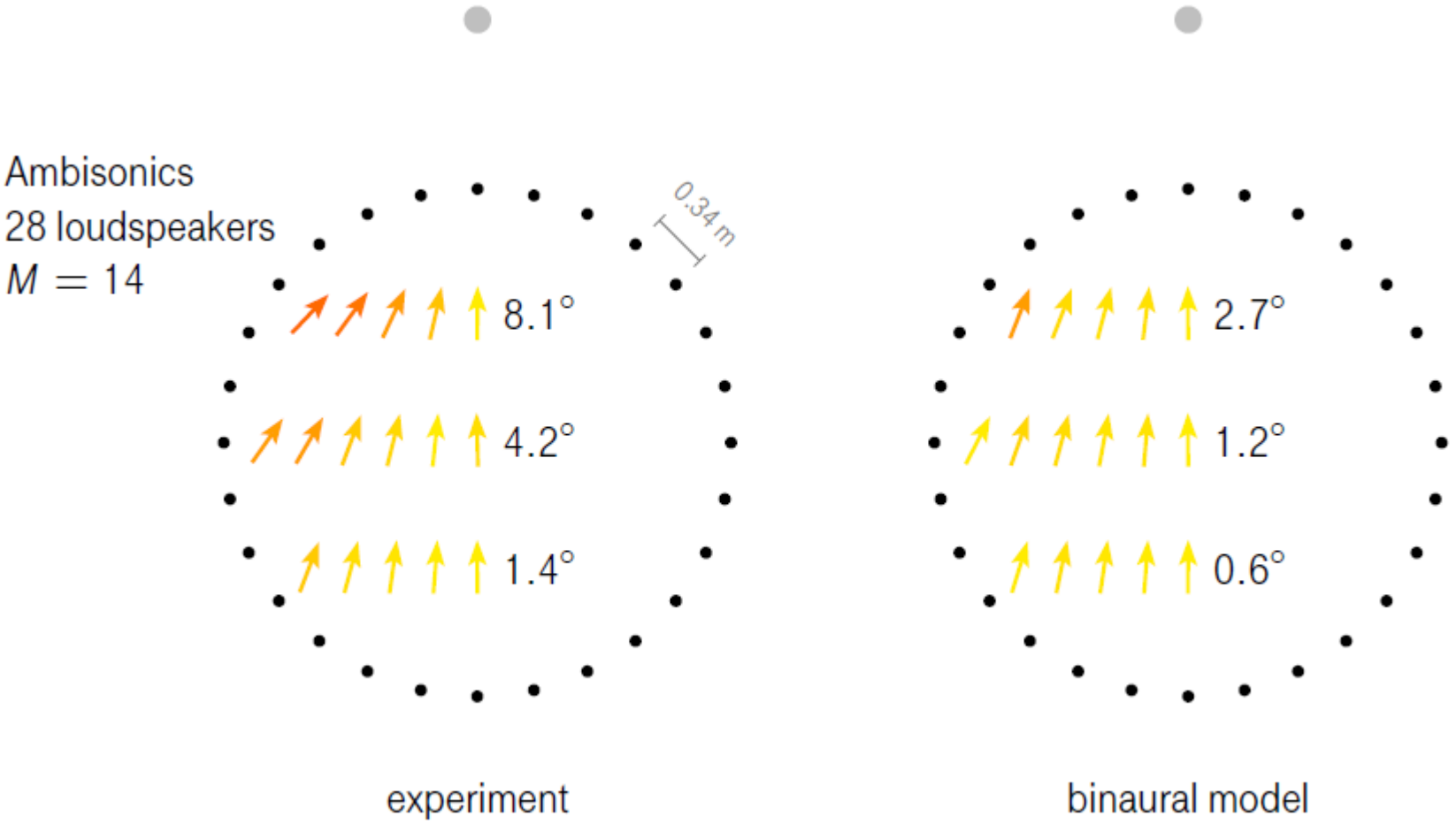
experiment



binaural model

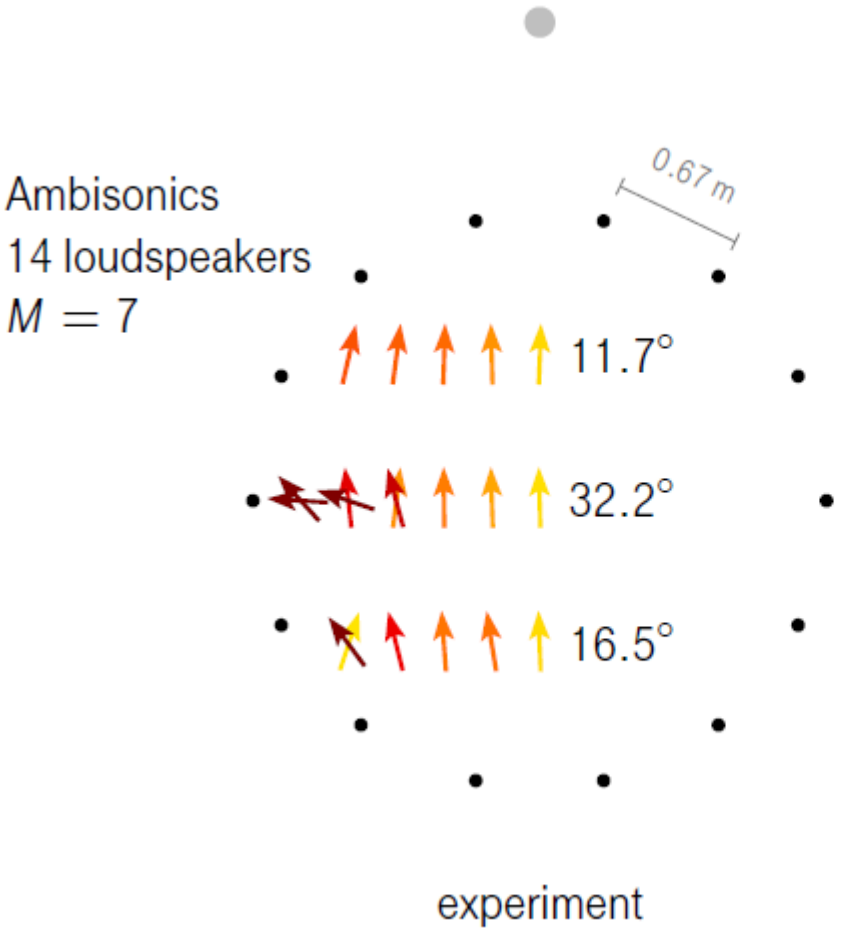
(Wierstorf et al., 2014)

Localization test & model
Ambisonics



(Wierstorf et al., 2014)

Localization test & model
Ambisonics



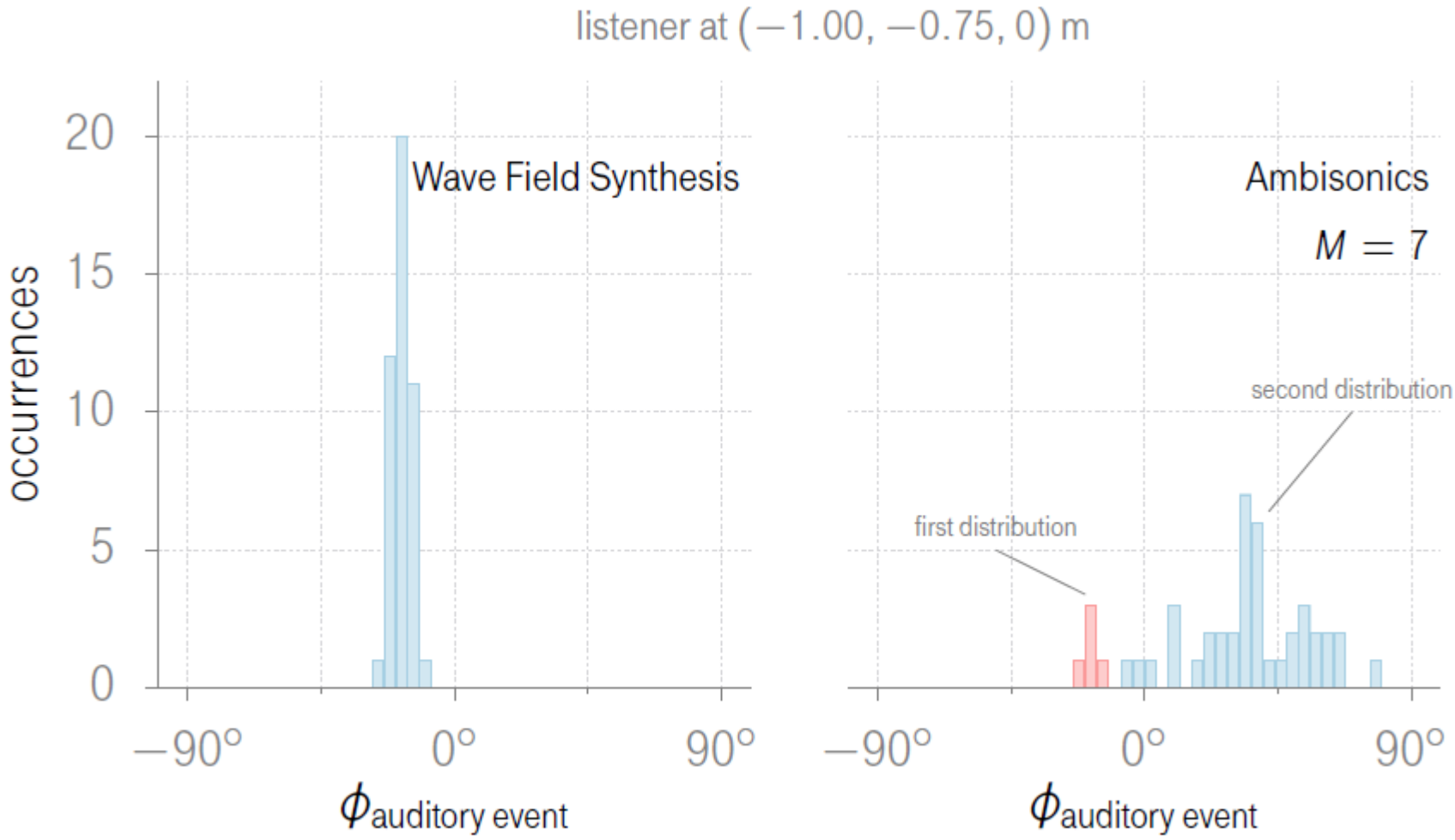
?

experiment

binaural model

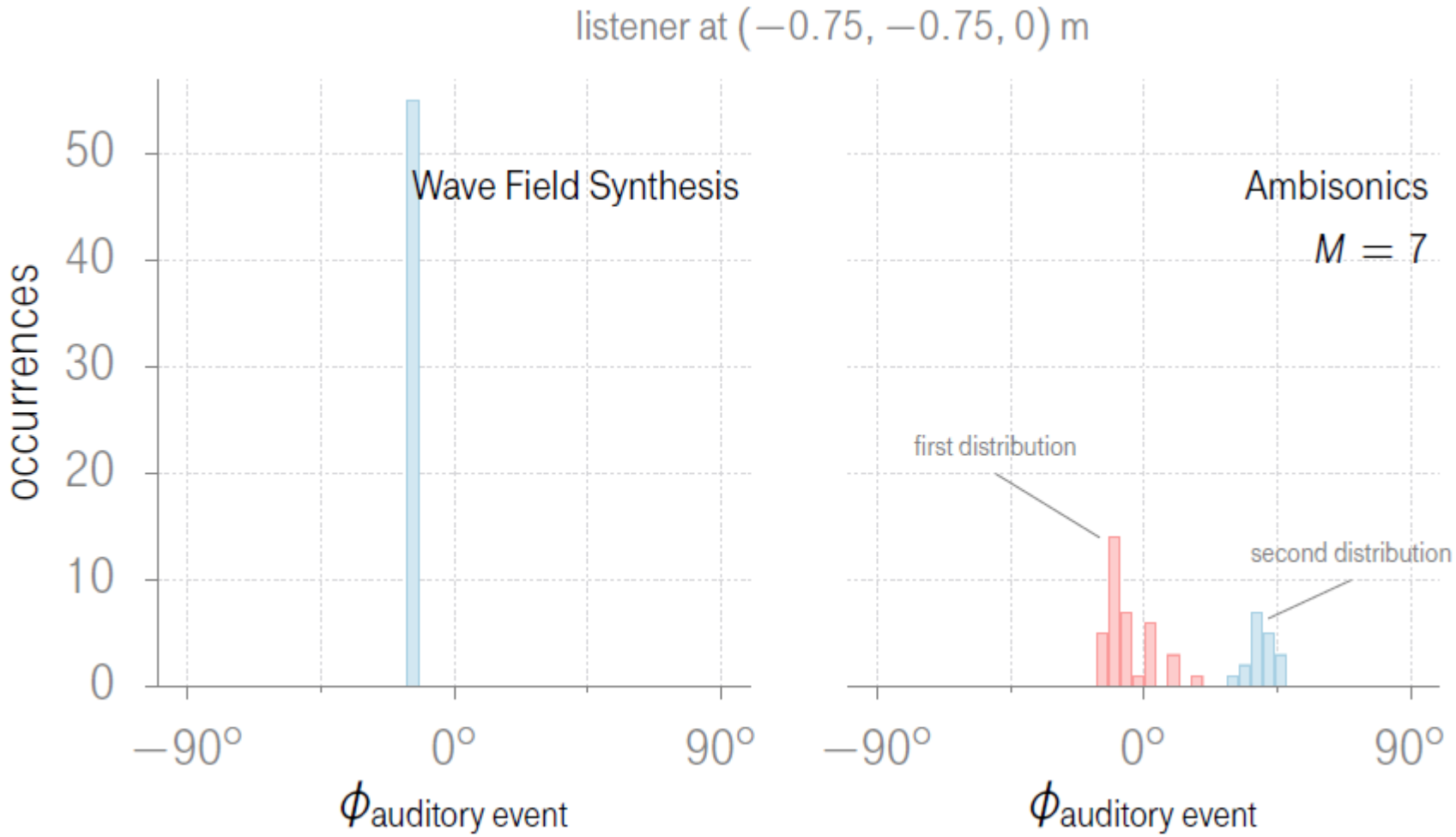
(Wierstorf et al., 2014)

Test vs. Model
Test: 11 subjects, 5 repetitions



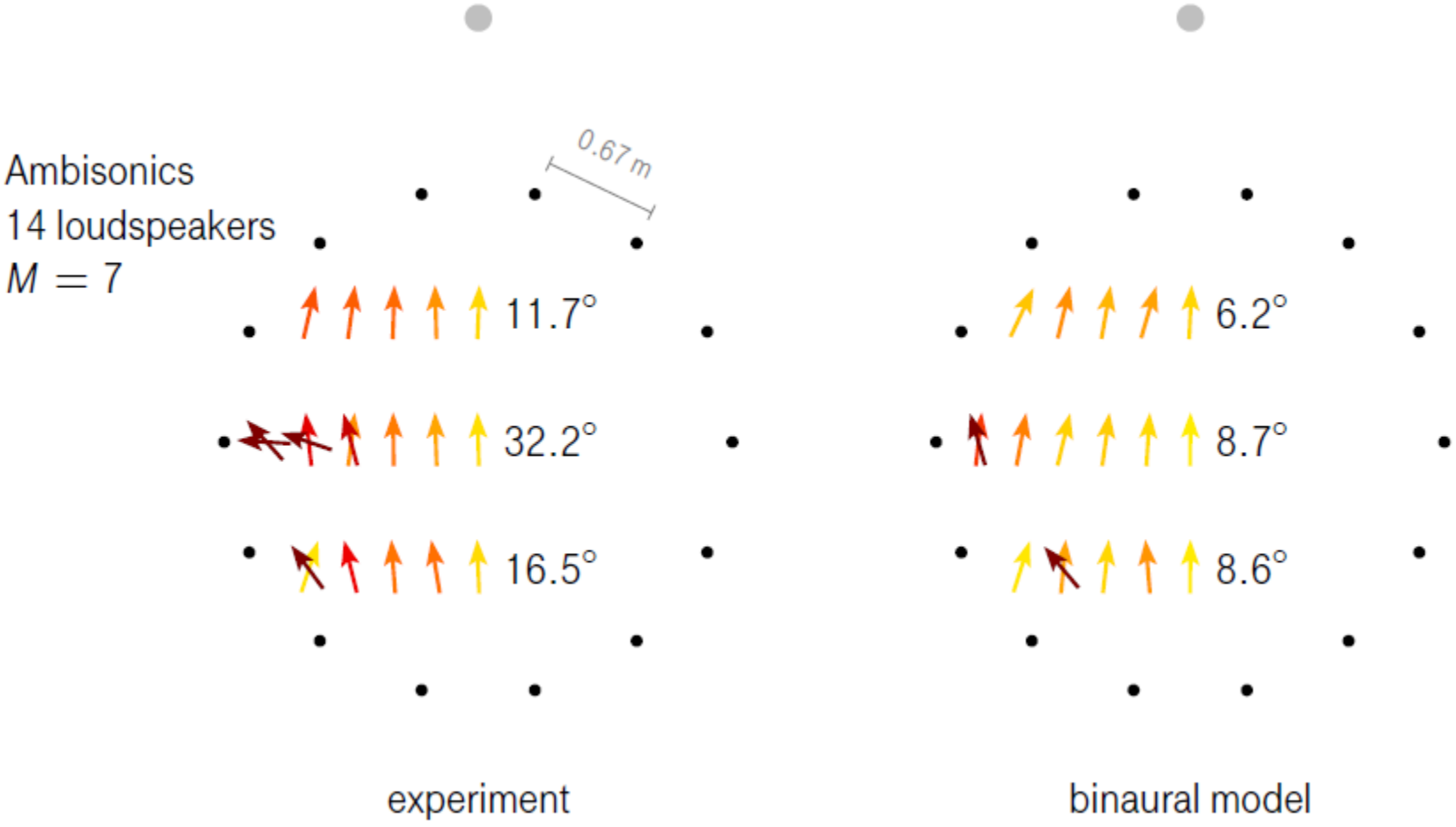
(Wierstorf et al., 2014)

Test vs. Model
Model: 11 instances, 5 repetitions



(Wierstorf et al., 2014)

Localization test & model
Ambisonics



(Wierstorf et al., 2014)

Conclusion
Model characteristics

- Learned internal references (not explicit reference signals)
  - NR model
  - FR model: Reference-adaptation
  - Plausibility instead of fidelity/authenticity
- Identification of scene and source types
  - Adjustment of low-level processing & internal reference in the light of given evaluation task (e.g. content-dependent)
- Attentional processes based on scene- & object-oriented paradigm
- Active exploration
  - Specific analysis of certain low-level features
  - Exploration of scene (e.g. identify sweet-spot of sound reproduction system)
- First test & modelling results

References

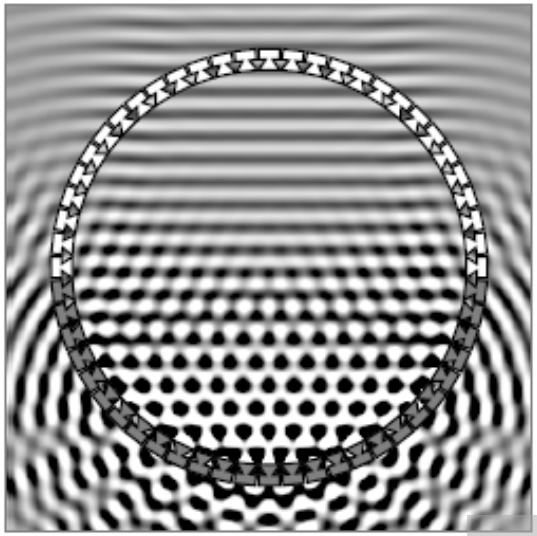
Blauert, J. (2013) Conceptual aspects regarding the qualification of spaces for aural performances, Acta Acustica united with Acustica 99, 1–13

Blauert, J. & Jekosch, U. (2012) A layer model of sound quality, J. Audio-Engr. Soc. 60, 4-12

Raake, A., Egger, S. (2014) Quality and Quality of Experience, In Quality of Experience: Advanced Concepts, Applications and Methods (S. Möller and A. Raake, eds.), Springer, DE-Berlin

Raake, A. & Blauert, J. (2013) Comprehensive modeling of the formation process of sound-quality. Proc. IEEE QoMEX 2013. Klagenfurt, Austria

Wierstorf, H., Raake, A., Spors, S. (2013) Binaural assessment of multi-channel reproduction, In: The technology of binaural listening (J. Blauert, ed.), Springer, USA-New York.

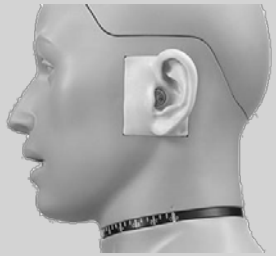


www.aipa.tu-berlin.de

www.ruhr-uni-bochum.de/ika/

www.twoears.eu

Thank you



TWO!EARS