On the Validity of Virtual Reality-based Auditory Experiments: A Case Study about Ratings of the Overall Listening Experience

Michael Schoeffler

Leibniz-Rechenzentrum Garching, Zentrum für Virtuelle Realität und Visualisierung, 09.07.2015

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer
IIS

Revised version (more accessible to people who did not attend the talk)
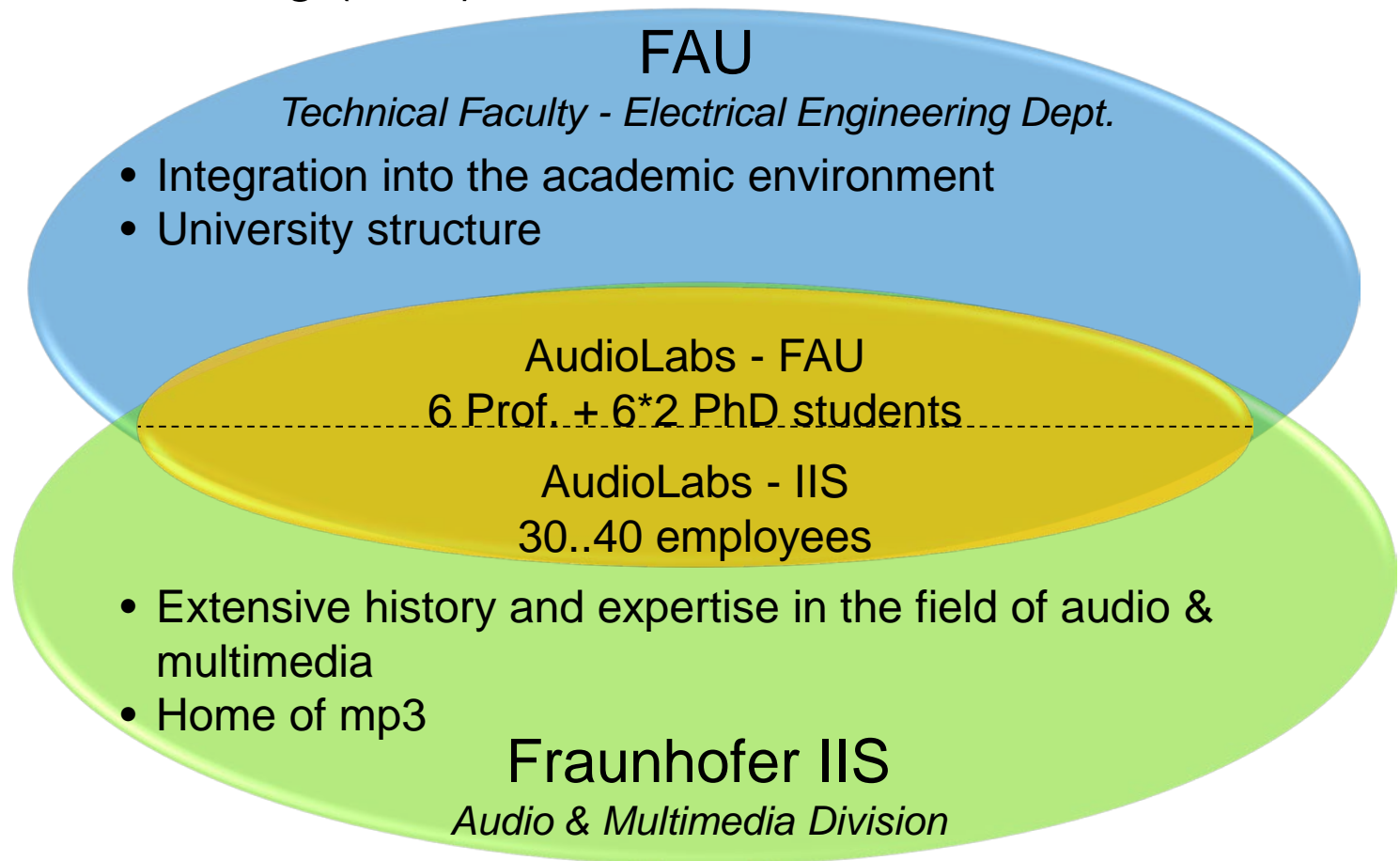
AUDIO LABS

About me

- Michael Schoeffler

- Doctoral candidate at the International Audio Laboratories Erlangen

- Background:
  Master's degree in Computer Science from Karlsruhe Institute of Technology

- Research interests:
  - Overall Listening Experience (OLE)
  - Prediction algorithms for OLE and audio quality
  - New methodologies for experiments (e.g. web-based or <u>VR-based experiments</u>)

AUDIO LABS

Internation Audio Laboratories Erlangen (AudioLabs)

- Joint research institute of Friedrich-Alexander Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS

FAU
Technical Faculty - Electrical Engineering Dept.

- Integration into the academic environment
- University structure

AudioLabs - FAU
6 Prof. + 6*2 PhD students

AudioLabs - IIS
30..40 employees

- Extensive history and expertise in the field of audio & multimedia
- Home of mp3

Fraunhofer IIS
Audio & Multimedia Division

Overview

- On the <u>Validity</u> of <u>Virtual Reality-based Auditory Experiments</u>: A Case Study about Ratings of the <u>Overall Listening Experience</u>

- Contents of this talk

  - **Validity:** 3
    Comparison between results obtained from a VR experiment and from a real-world experiment

  - **VR-based Auditory Experiments:** 2
    How to create VR environments?

  - **Overall Listening Experience:** 1
    Insights about evaluations in context of audio

Agenda

- What is overall listening experience?
- Motivation & research question
- VR System
    - Visuals
    - User Interface
    - Audio
- Experiment
    - Participants
    - Stimuli
    - Procedure
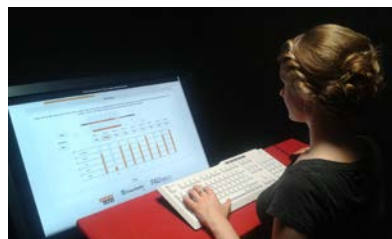   - Demo (video)
    - Results
- Conclusion & Outlook

AUDIO
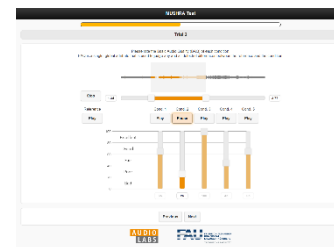LABS

WHAT IS OVERALL LISTENING EXPERIENCE?

Audio Quality

- Let's start from the very beginning...

- In context of audio, new developments are evaluated by "measuring" audio quality

- Typically, a subjective evaluation method (listening test) is applied to measure audio quality

speakers

Assessor

Audio systems (under test)

Listening test method and software

MUSHRA (ITU-R BS.1534)

- **Multi-Stimulus Test with Hidden Reference and Anchor**

- A widely used listening test methodology

- MUSHRA is recommended for assessing "intermediate audio quality"

- Assessors are asked to rate "Basic Audio Quality" of audio systems

  - BAQ: "This single, global attribute is used to judge any and all detected differences between the reference and the object [reference stimulus processed by audio system]."

[International Telecommunication Union, BS.1534 : Method for the subjective assessment of intermediate quality levels of coding systems, 2014]

MUSHRA (ITU-R BS.1534)
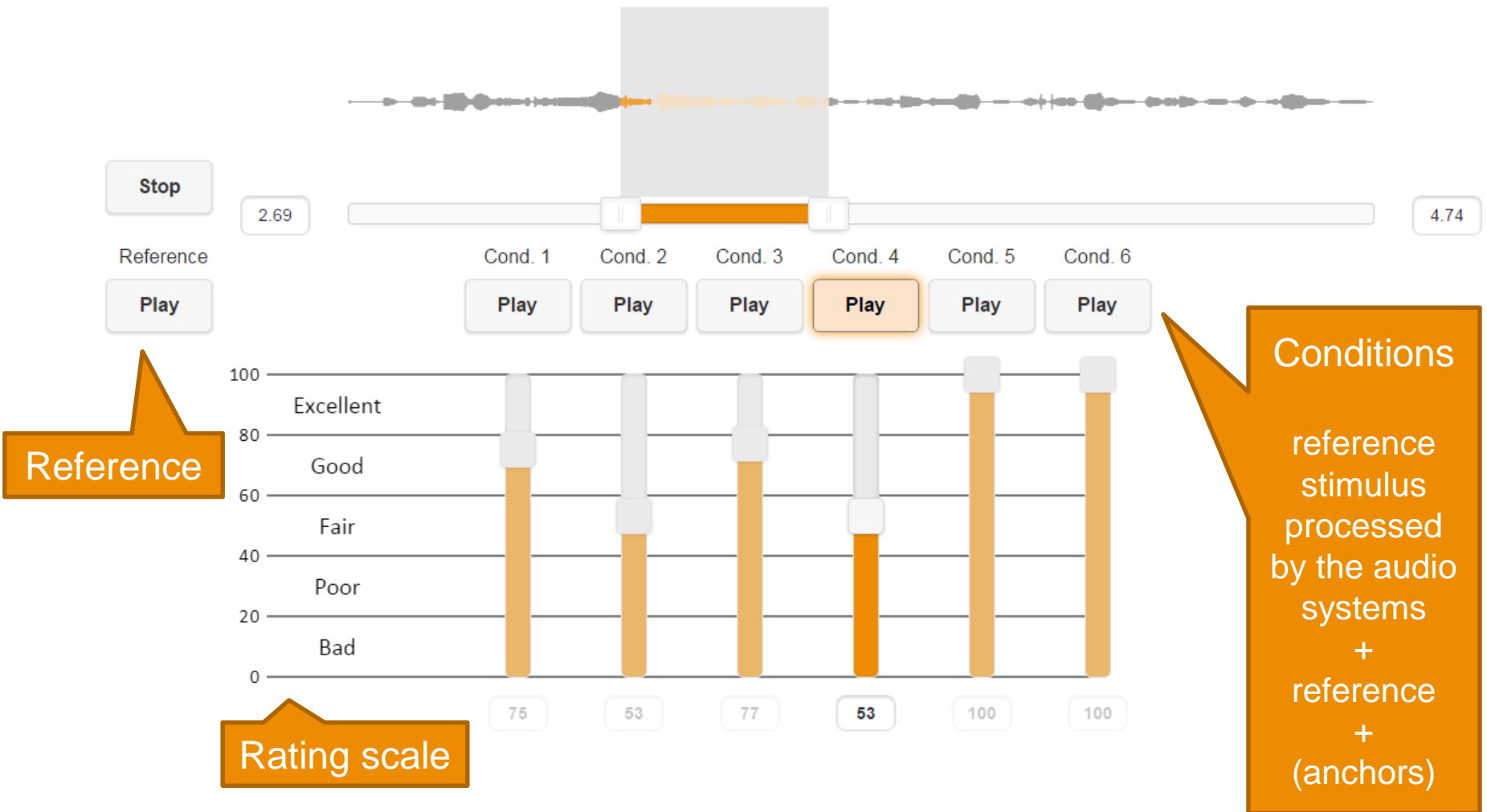
[Schoeffler et al., Towards the Next Generation of Web-based Experiments: A Case Study Assessing Basic Audio Quality Following the ITU-R Recommendation BS.1534 (MUSHRA), 1st Web Audio Conference, 2015]
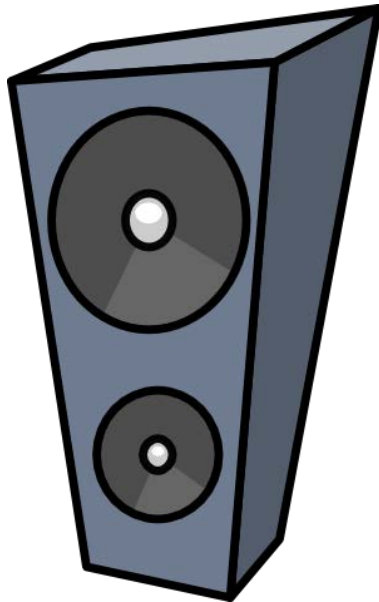
Overall Listening Experience

- ## What is the overall listening experience (OLE)?
    - The sensation, perception and cognition of sound events (e.g. music)

- ## What is a rating of the overall listening experience?
    - A rating reflects the degree of the listener's enjoyment (when listening to music)
    - When listeners are asked to rate the overall listening experience, they take everything into account that influences their enjoyment
    - E.g.: Mood, audio quality, song artist, lyrics,...

[Schoeffler and Herre, About the Impact of Audio Quality on Overall Listening Experience, Sound and Music Computing Conference, 2013]

Audio Quality vs. Overall Listening Experience

- Many assessment experiments focus on audio quality
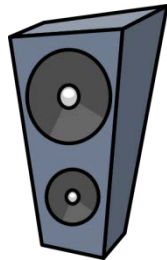- Example: Assessment of loudspeakers

Loudspeaker A
"AudioExperts S200 Pro"

Loudspeaker B
"Tinny by PoorAudio"
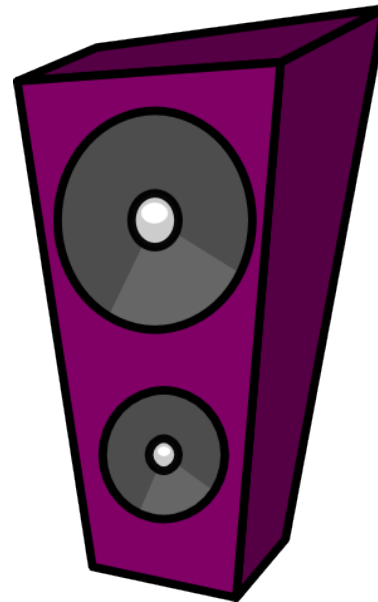
Audio Quality vs. Overall Listening Experience

- Audio Quality

Audio Quality vs. Overall Listening Experience
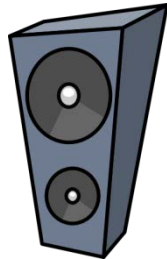
- Overall Listening Experience

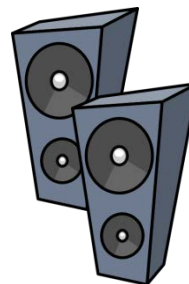# Effects on Overall Listening Experience

Audio Quality (Distortions, Signal Bandwidth)
- *Schoeffler and Herre, About the Impact of Audio Quality on Overall Listening Experience, SMC, 2013*
- *Schoeffler et al., How Much Does Audio Quality Influence Ratings of Overall Listening Experience?, CMMR, 2013*

Single-/Multi-channel System
- *Schoeffler et al., The Influence of the Single-/Multi-Channel-System on the Overall Listening Experience, 55th AES Conf., 2014*
- *Schoeffler et al., The Influence of Up- and Down-mixes on the Overall Listening Experience, 137th AES Conv., 2014*
- *Rumsey et al., Relationships between Experienced Listener Ratings of Multichannel Audio Quality and Naïve Listener Preferences, JASA, 2005*

Listening Room
- *Schoeffler et al., The Influence of the Single-/Multi-Channel-System on the Overall Listening Experience, 55th AES Conf., 2014*
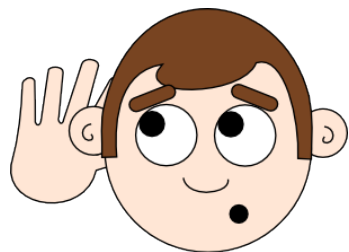
Individual Listener
- *Schoeffler and Herre, About the Different Types of Listeners for Rating the Overall Listening Experience, SMC, 2014*
- *Pearson and Dollinger, Music preference correlates of Jungian types, PID, 2004*
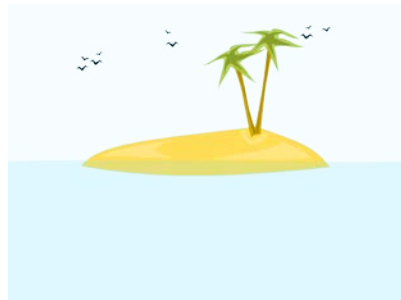
Overall Listening Experience

- Why is the overall listening experience important?

- When it comes to selling your product, the "average customer" might care more about the overall listening experience than audio quality
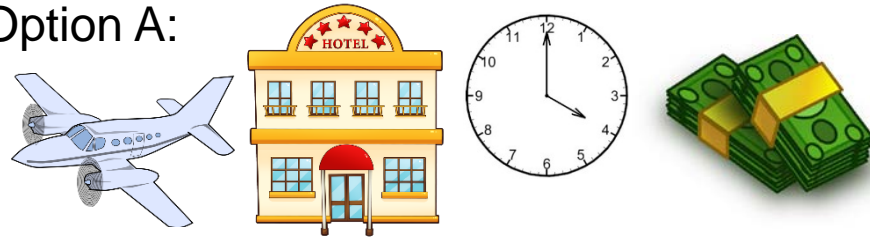
MOTIVATION & RESEARCH QUESTION

"Big Picture" Motivation

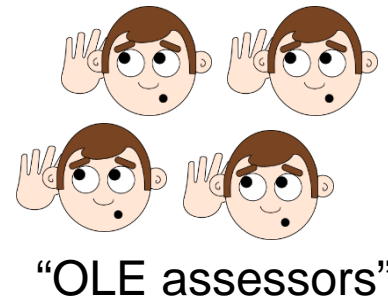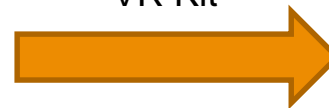- OLE evaluation of a cinema (on a island somewhere in the Pacific)

Option A:

Option B:

"VR Kit"

"OLE assessors"

Are VR experiments valid compared to real-world experiments?

Validity of VR Experiments

- **Driving Simulation by Francesco Bella**
  - Research question: Do speed measurements obtained from a real-world work zone and from a VR simulation belong to the same population?

- "…the field speeds and those from the simulation belong to the same population"

[Bella, Driving simulation in virtual reality for work zone design on highway: a validation study, 2nd SIIV, 2004]

Validity of VR Experiments

- Only a few studies have been carried out that compare real-world experiments to VR experiments

  - Especially in context to audio

- Some work has been done related to auditory virtual environments (AVEs)

- AVEs are defined as the auditory components of virtual environments which aim at creating situations in which humans have perceptions that do not correspond to their physical environment but to the virtual one.
  [Novo, Auditory virtual environments, Communication Acoustics, 2005]

  - "audio-only VR"

- Still a lot of work to do!

Research questions

- Two research questions:
  - Does the room (cinema and listening booth) have an influence on the overall listening experience?

  - **Does the environment (real-world and Virtual Reality) have an influence on the overall listening experience?**
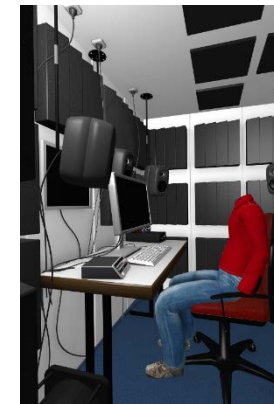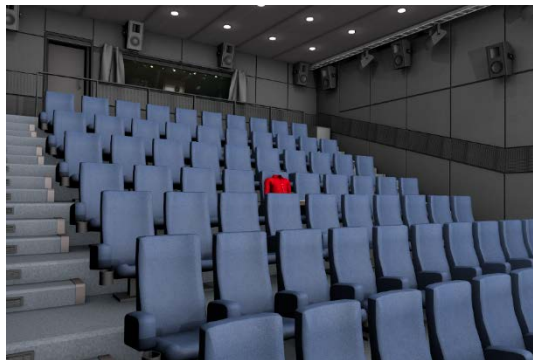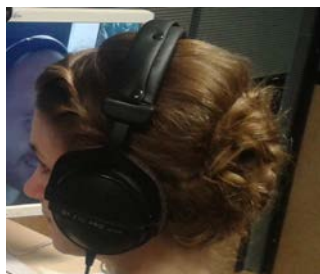
VR SYSTEM

VR System

- In order to carry out a "VR experiment", we need a VR system!

- Equipment:

Oculus Rift DK2 / Visuals
Head-mounted display
low-latency headtracking
75 Hz display
960 x 1080 pixels per "eye"

Headphones / Audio

Software

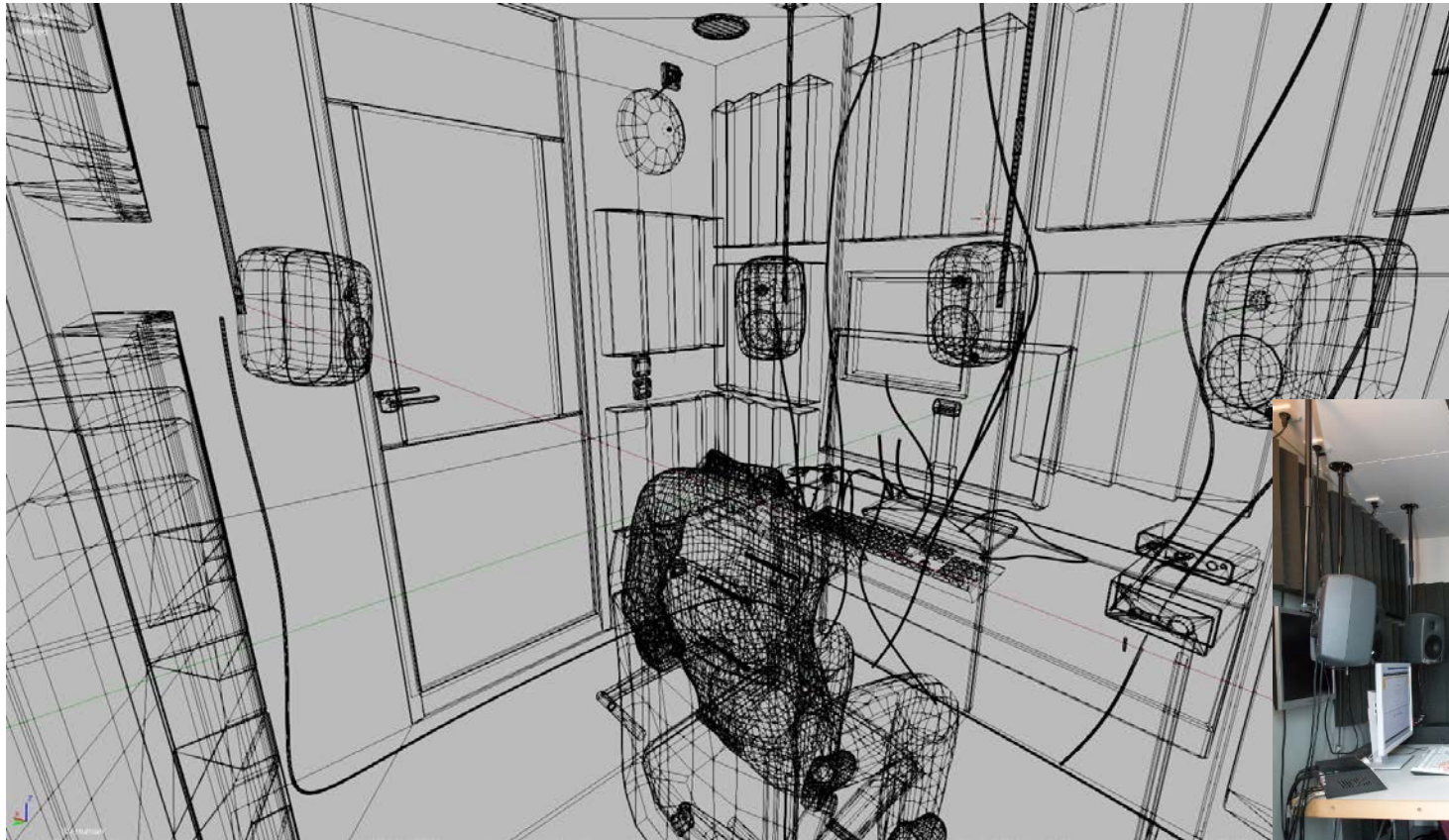How to create authentic visual stimuli for our VR environment

VISUALS

Visuals

- Authentic visual stimuli are achieved by providing photorealistic graphics

- Blueprints of the two rooms (cinema and listening booth) were used to create true-to-scale models

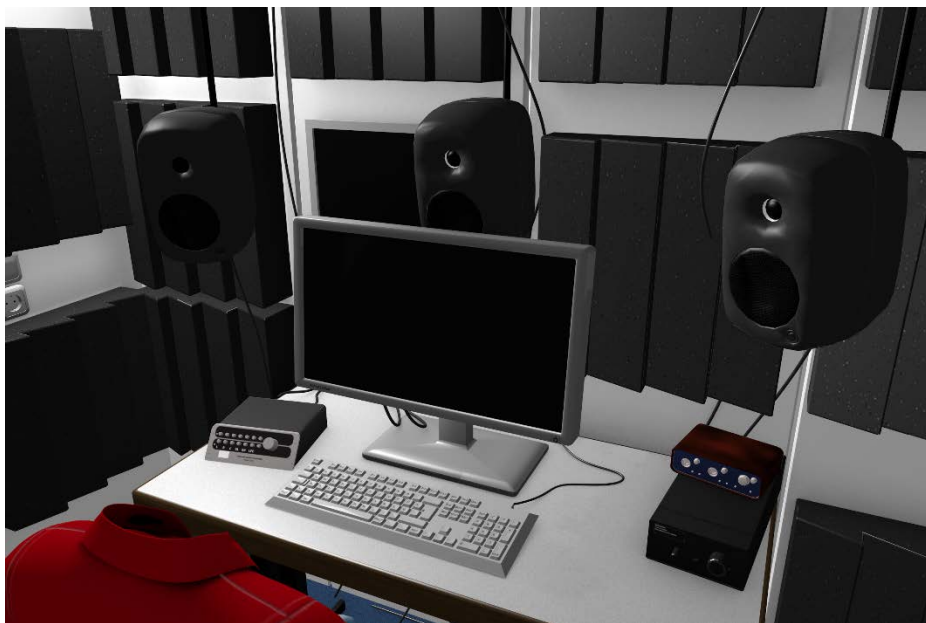- Textures of the objects (of the two models) were partly photographed
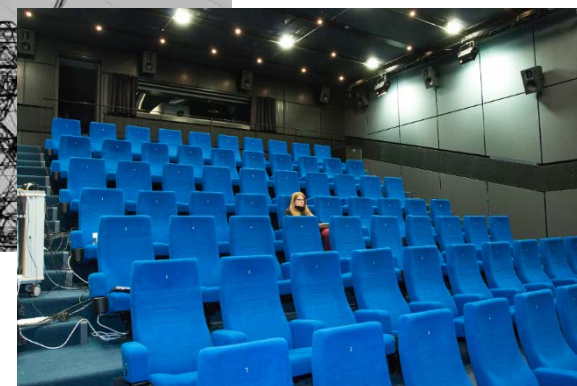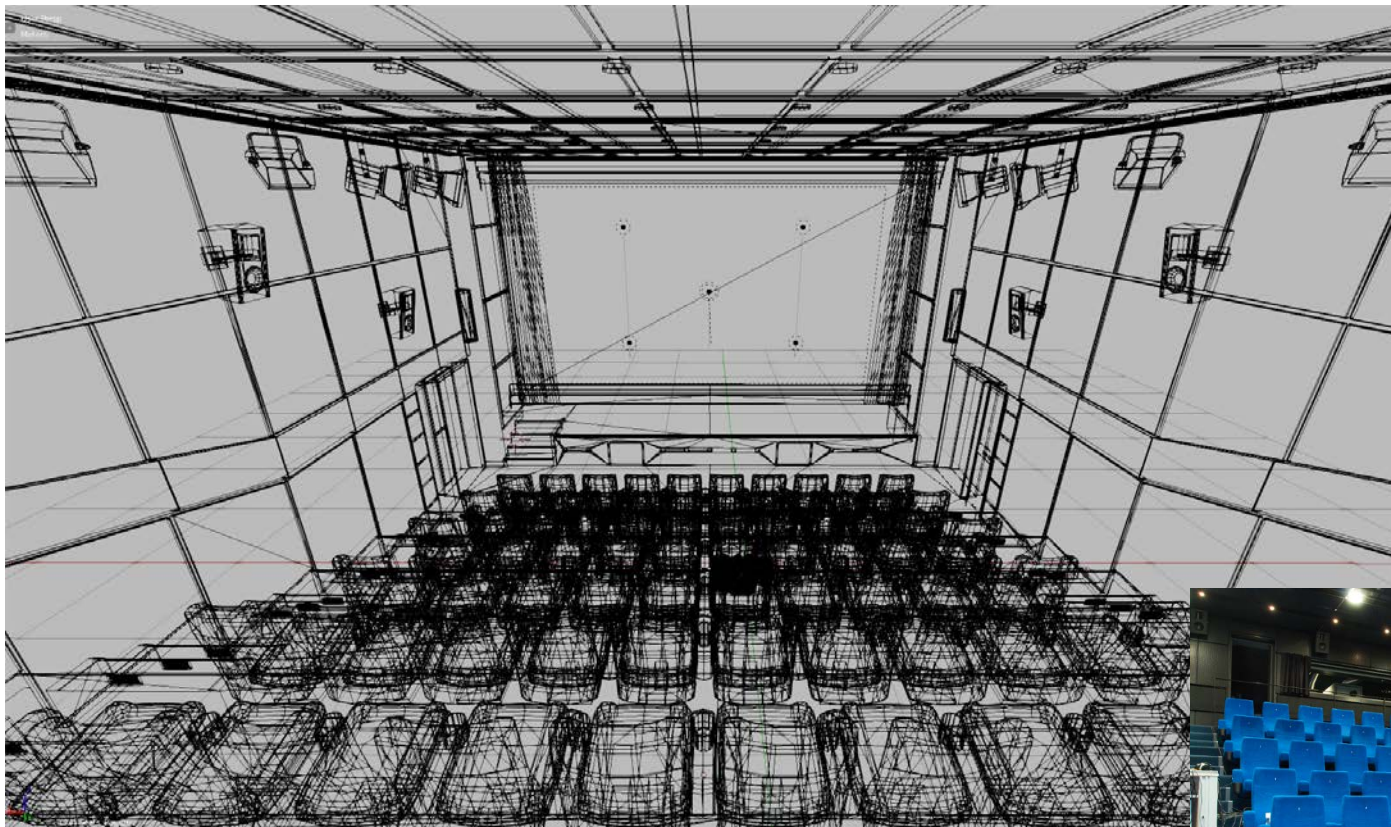
Room Models

- **Listening Booth**

Listening Booth

- Even though the resulting graphics are not 100% photorealistic, we are satisfied with the results.
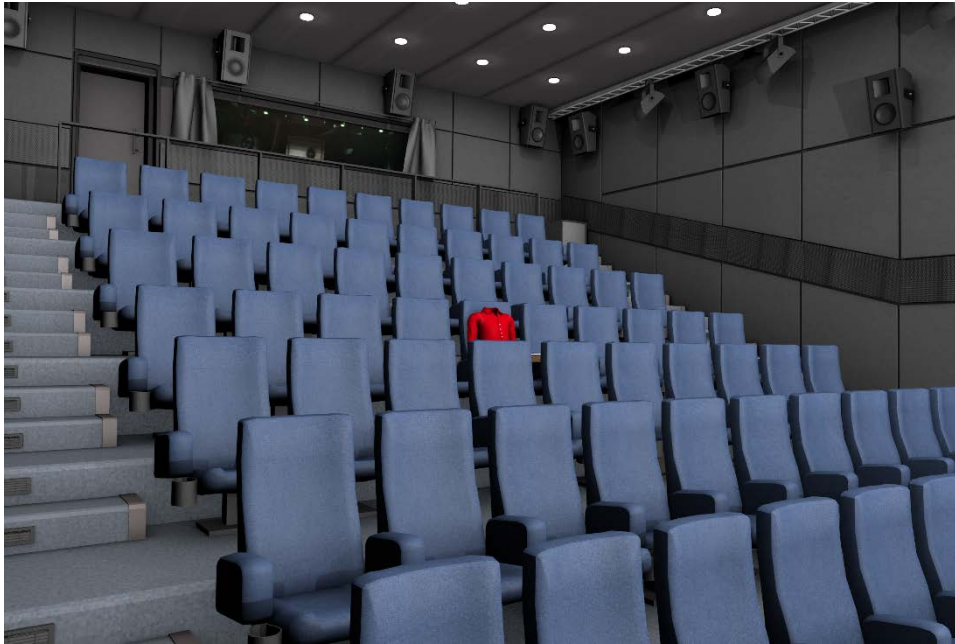
Room models

- Cinema

Cinema

On the Validity of Virtual Reality-based Auditory Experiments
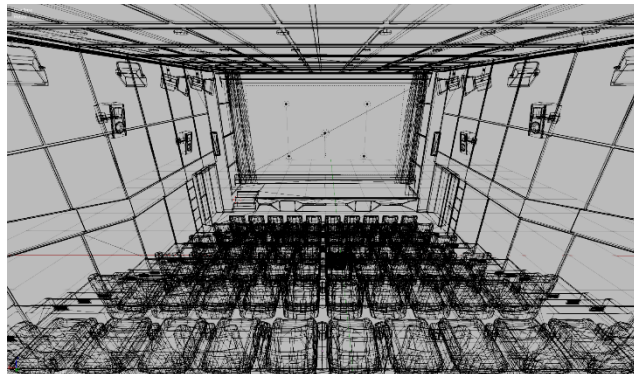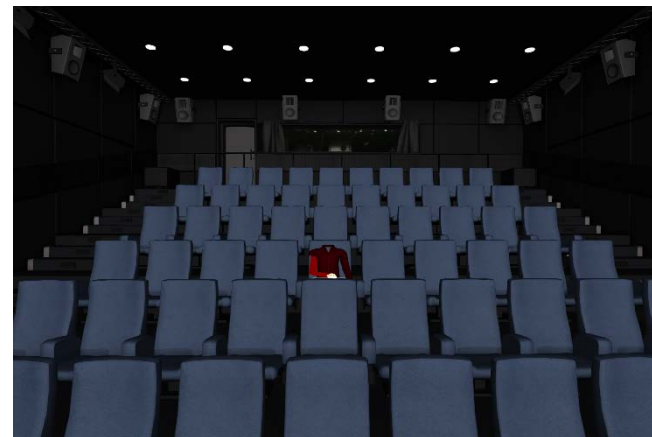
Game Engine

- Open source
- Fully "hackable"

OGRE renderer modifications

Visuals – what could be improved?

- Quality of graphics turned out good, however, they are not perfect!

- In order to further improve the graphics we have to...

  - Create all textures by ourselves
    - very-time consuming

  - Use a different game engine
    - Ogre engine has its limits (Unity and CryEngine have more features and look much better)
    - When we started the project, game engines had no full support for the Oculus Rift => fully hackable engine was needed

  - Higher resolution
    - Limited by the Oculus Rift DK2
    - Screen-door effect is still present (lines separating the display's pixel are visible)
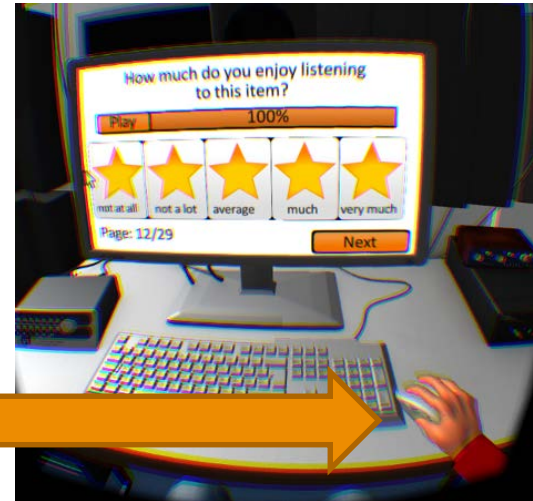
USER INTERFACE

User Interface

- Participants gave their responses by a graphical user interface (GUI) displayed on a screen

- We implemented a framework that enables to easily configure an "experiment GUI"

Feedback

- In a pilot test, many participants reported that it feels unnatural to move the "real mouse" but don't having any visual feedback in the VR environment

- To solve this issue, we added a virtual mouse to the VR environment
  - We retrieve the mouse pointer coordinates (pixels) from the operating system and map them to a "mm-scale"
  - The virtual mouse moves according to changes of the mouse pointer coordinates
  - The whole (virtual) arm moves when the virtual mouse changes its position

User Interface – what could be improved?

- In another pilot test, some participants started to move both hands around, when they realized the right hand and mouse is moving

  - Position tracking of the participant's body parts might improve the quality of experience

- The control elements of the "experiment GUI" looked "pixelated"

  - Oculus Rift has a resolution of 1920 x 1080 pixels

  - As the display is split, the maximum resolution for rendering the virtual scene is only 960 x 1080 pixels

  - The actual resolution for rendering the "experiment GUI" is about 500 x 300 pixels
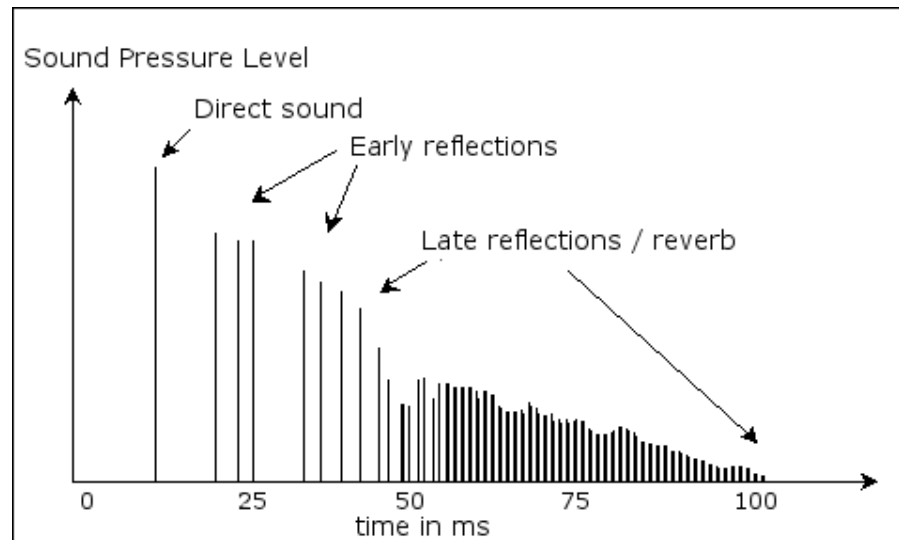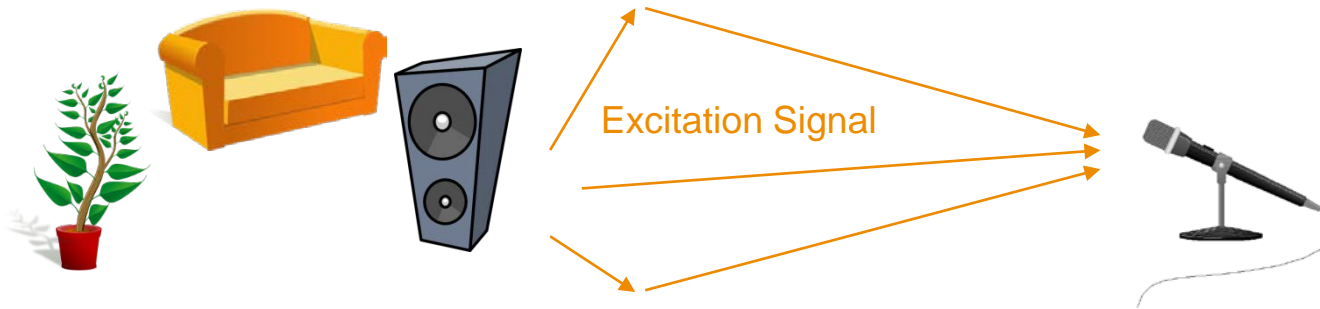
1080 px

300 px

AUDIO LABS
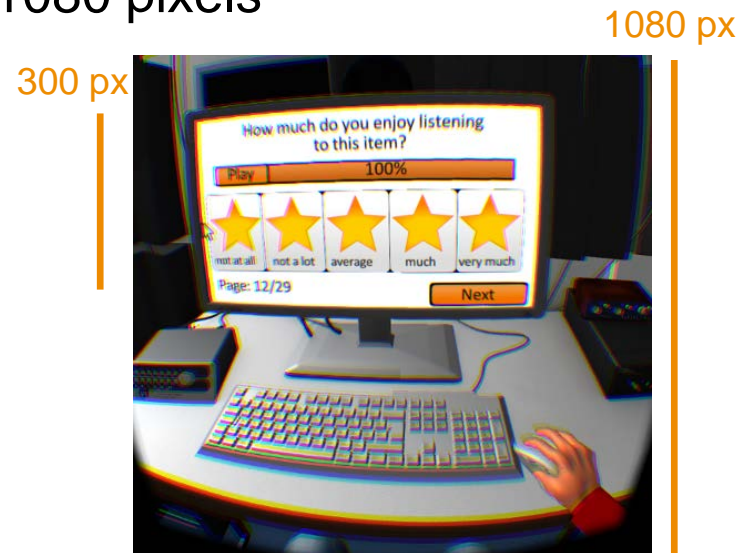
How to create authentic auditory stimuli for our VR environment

AUDIO

Room Impulse Responses

- How do you capture the "acoustical characteristics" of a room?
    - => by recording a "room impulse responses"

Excitation Signal

Room Impulse Responses

- **Excitation Signal**

  Ideal impulse (Dirac)
  - Infinitely short
  - Infinite power
  - Only theoretical

  ➡ Exact impulse response

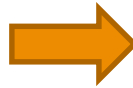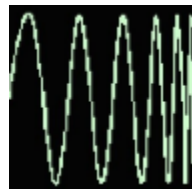 **Approximated impulse**
  - e.g. pistol, clapping hand, paper bag etc.
  - Poor reproducibility

  ➡ Approx. impulse response

  Sweep signal
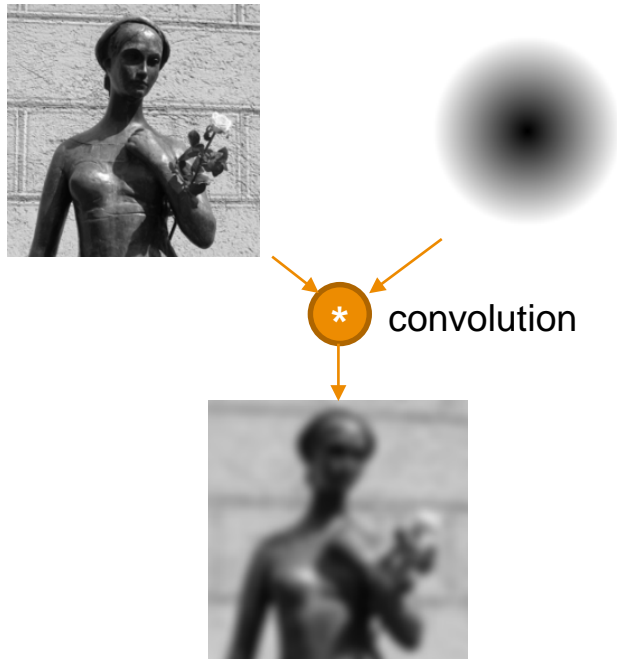 - Periodically swept sine

  ➡ Sweep response
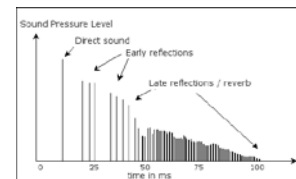 post-processing needed: deconvolution with source signal

[Acoustics Engineering, Measuring Impulse Responses using DIRAC, Technical Note, 2007]
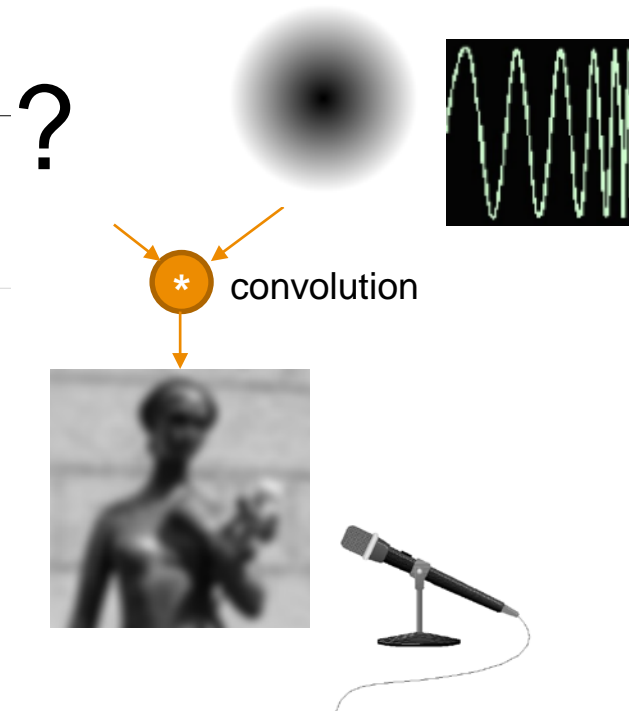
Room Impulse Response

- Deconvolution of sweep signals

- Image processing as an example

Point spread function (PSF)

convolution

Deconvolution:

?

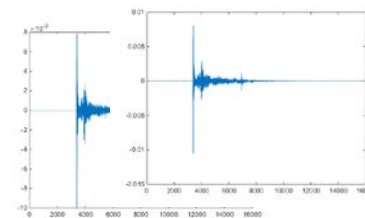convolution

More advanced: Binaural Room Impulse Responses (BRIRs)

Dummy head with two "in-ear" microphones

[Satongar et al., The Salford-BBC Spatially-sampled Binaural Room Impulse Responses dataset, http://rdata.salford.ac.uk/sbsbrir/, CC BY-NC-SA 4.0, 2014]
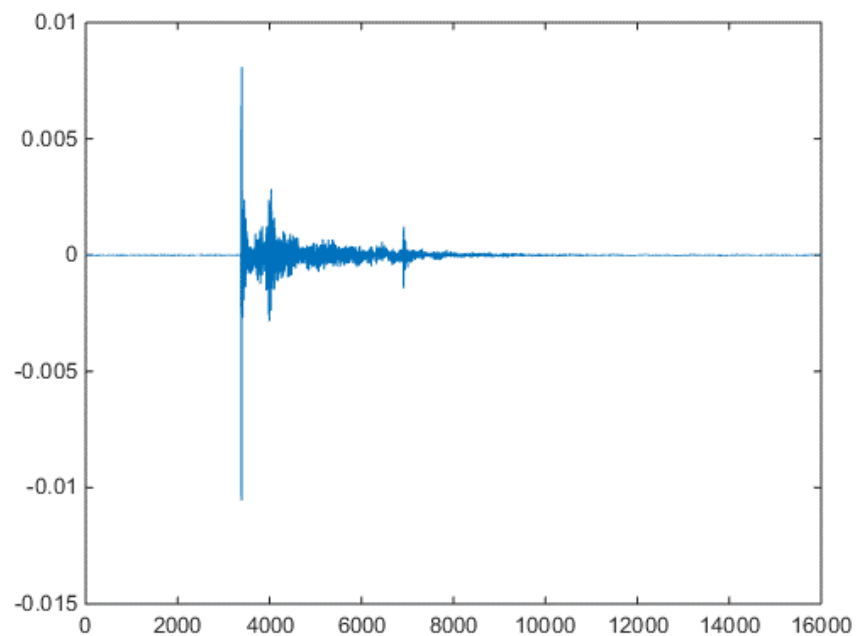
- The BRIR "contains information" about the room, ear, head, body etc.

- Each recording results in two BRIRs (left ear and right ear)

[Müller and Massarani, Transfer-function measurement with sweeps, J. Audio Eng. Soc., 49(6), 2001]
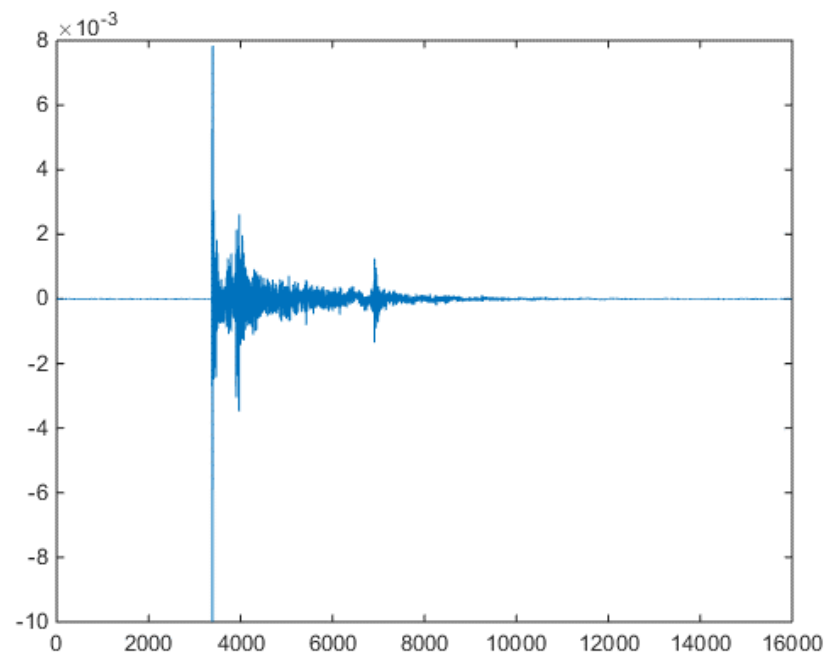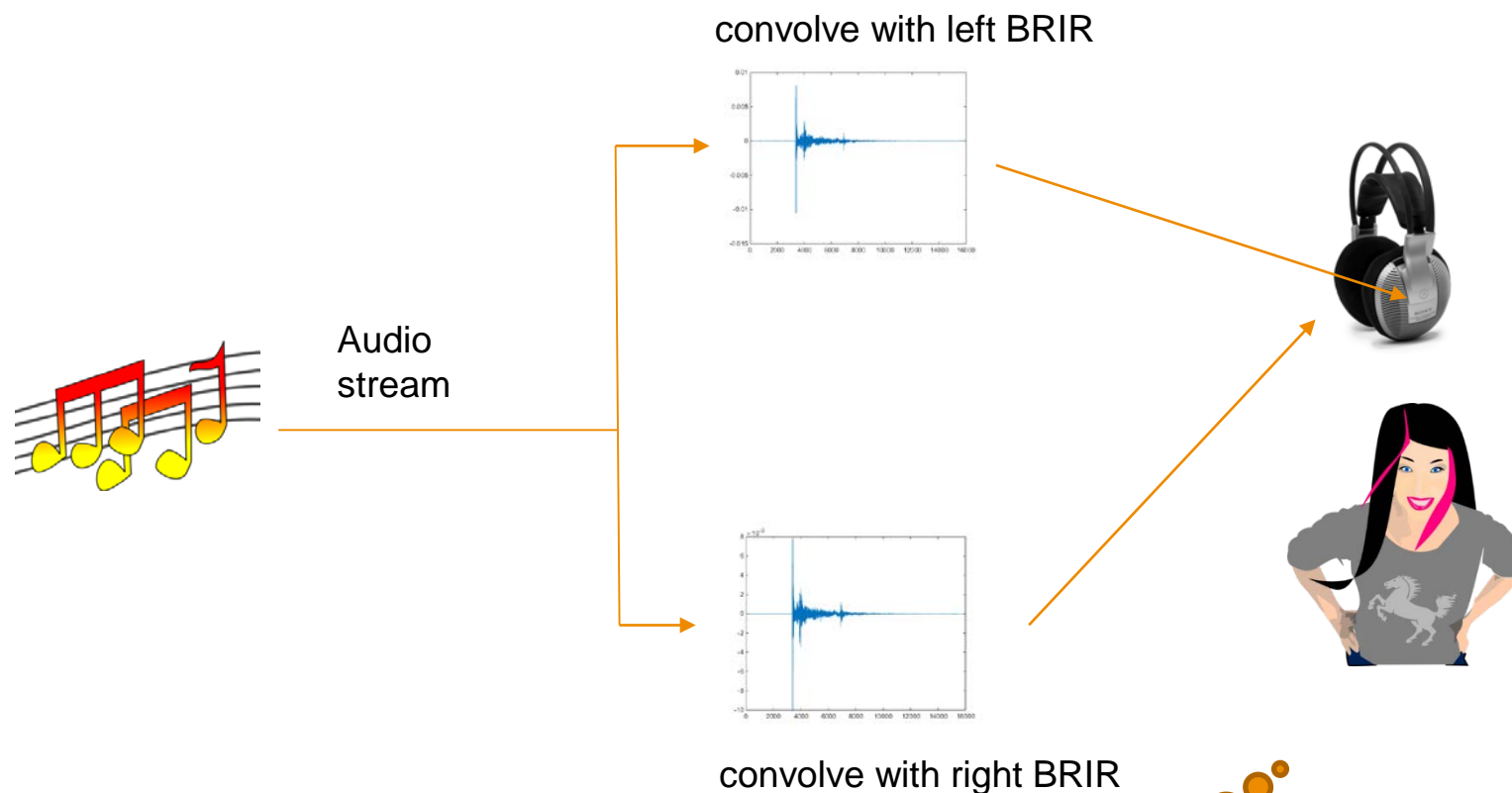
Example of BRIRs

- Left ear

- Right ear

AUDIO LABS

Convolution of Binaural Room Impulse Responses

convolve with left BRIR

Audio stream

convolve with right BRIR

Wow, sounds like I'm listening to this music while sitting in a room.

BRIR measurement

- If users are allowed to move their head, multiple BRIRs must be measured

  - A BRIR measurement is needed for each head position
  - Step sizes of less than 3° should provide a binaural synthesis free from resolution artefacts [Lindau et al., Minimum grid resolution for dynamic binaural synthesis, Acoustics, 2008]

- In each room, we measured 1215 different head rotations

  - Yaw: from -40° to +40° / step size of 1°
  - Pitch: from -6° to +6° / step size of 3°
  - Roll: from -3° to +3° / step size of 3°
  - No translational movements!

BRIR measurement

- We measured BRIRs of both rooms (cinema and listening booth) with a custom-made dummy head
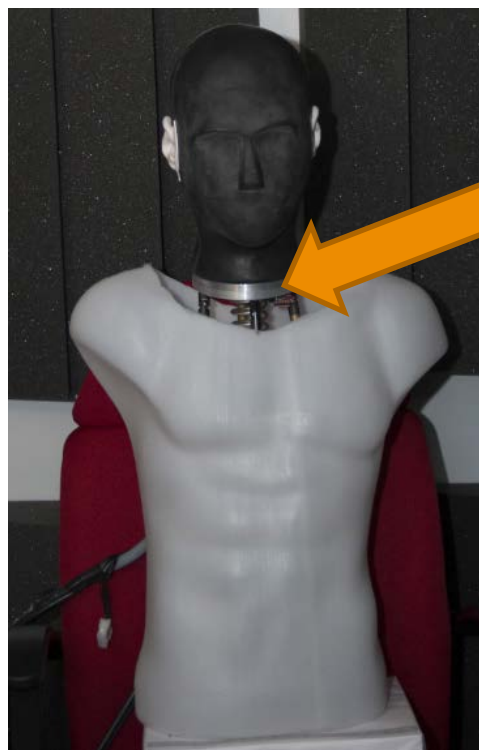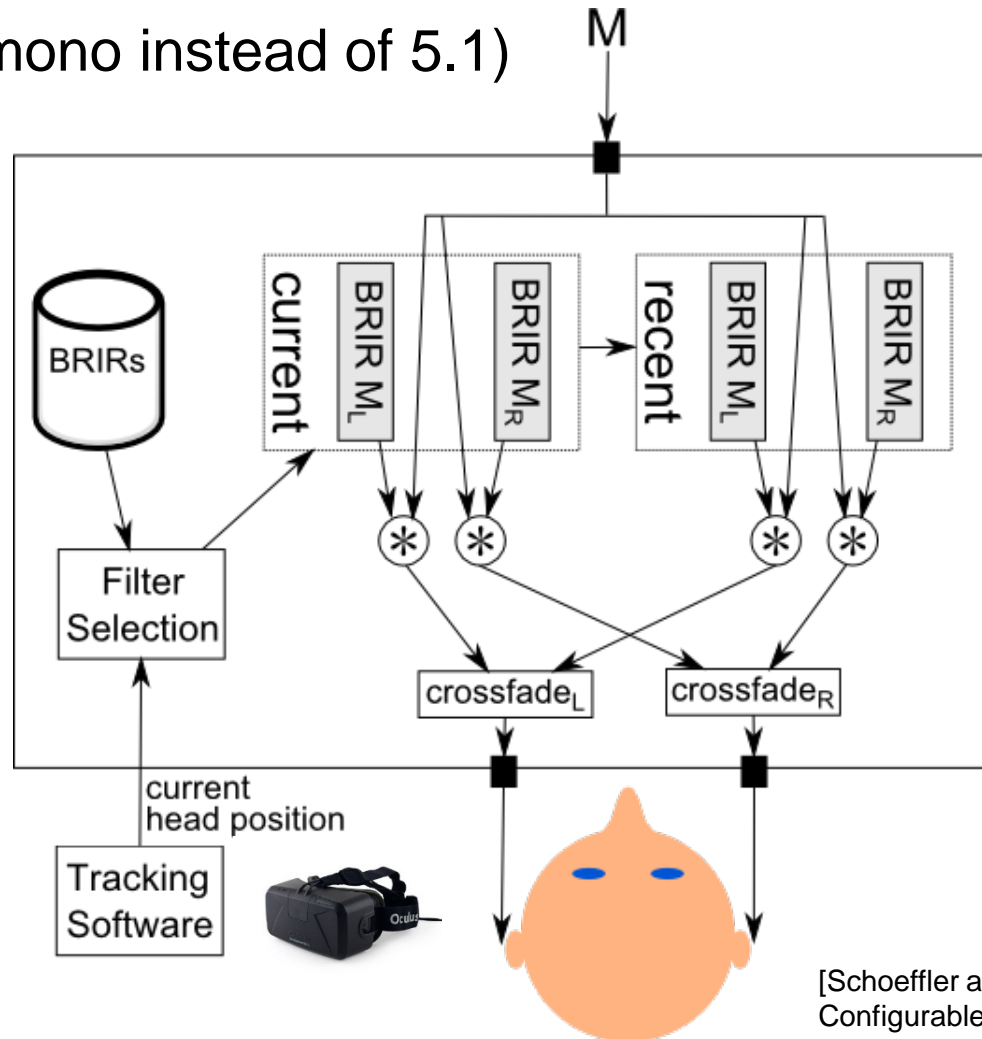

The dummy head has several motors that enable to measure multiple BRIRs with different head rotations.

We measured 1215 head rotations multiplied by 6 loudspeakers (5.1 surround sound).

In total, 10h measurement time for one room

AUDIO LABS

"Convolution Engine"

- Simplified (mono instead of 5.1)

[Schoeffler and Hess, A Comparison of Highly Configurable CPU-and GPU-Based Convolution Engines, 133rd AES Conv., 2012]

AUDIO LABS

Equalization of Headphones

- Nowadays, headphones are "diffuse-field equalized"

- To put it simple: "Diffuse-field equalized headphone try to mimic a loudspeaker in a reverbant room"

- The measured BRIRs were post-processed to compensate the headphone's transfer function by inverse filtering

- "Nonetheless, it was clearly shown that in all cases even a non-individual equalization will yield more plausible simulation results than using no equalization at all."
[Schärer and Lindau, Evaluation of Equalization Methods for Binaural Signals, 126th AES Conv, 2009]

Audio – what could be improved?

- "Individual BRIRs"
- The dummy head has not *your* ears, head and torso!

- Individual BRIRs must be measured without a dummy head
- Every participant would have to sit in each room for about 10 hours ☹
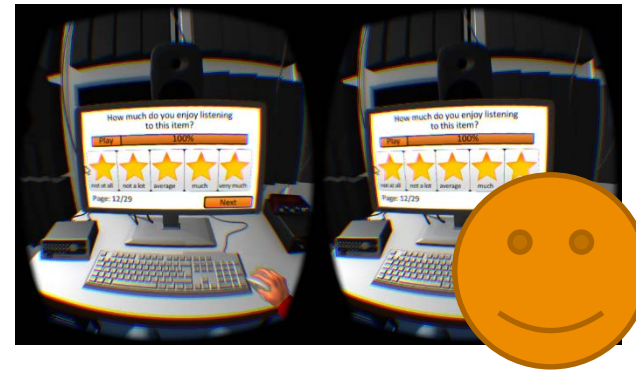
VR System (Recap)

- Equipment:

Oculus Rift DK2 / Visuals

Headphones / Audio


Software

EXPERIMENT

Participants

- 30 participants (24 males, 6 females) volunteered to participate in the experiment

- 20 participants identified themselves as professionals in audio (audio researchers, audio engineers, Tonmeisters etc.)

- 27 participants were familiar with listening tests and indicated that they had volunteered for at least one listening test before

- 20 participants reported that they regularly play computer games for more than one hour a week

Stimuli

- Fifteen music excerpts of songs of various genres
    - You're Beautiful (James Blunt)
    - She Drives Me Crazy (Fine Young Cannibals)
    - Symphony Nr. 4 (Peter I. Tschaikowsky)
    - Shout (Tears For Fears)
    - Tonight (Alex Max Band)
    - ...
- Duration about 10-15 seconds
- Mainly covered the most recognizable part of the song (e.g. the refrain)
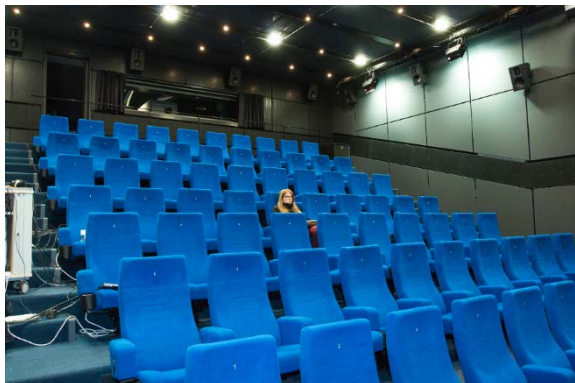- Mixed in 5.1 surround sound

AUDIO LABS

Procedure

- **5 Sessions**

- **"Initial session" (first session)**

  - Participants rated the excerpts in a multi-stimulus comparison

    - "How much do you enjoy listening to each item?"

    - => "Basic Item Rating"

  - Headphones

  - Professional listening room

  - Purpose:

    - Familiarization with the stimuli and the rating scale

    - To find out how much a participant liked each song without being influenced by the room (and environment)

  - At the end of the session, participants tried out the Oculus Rift
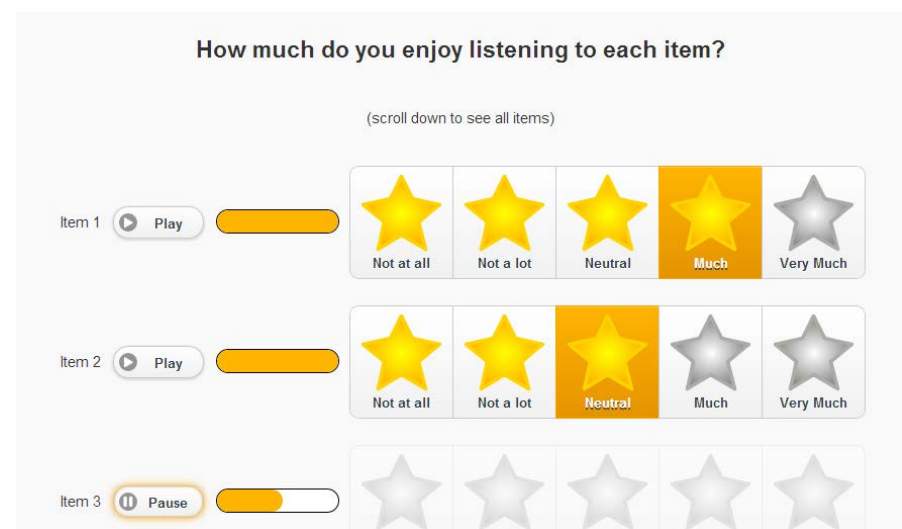
Procedure (cont.)

- 4 sessions (in random order)
- Participants rated the stimuli in different environments and rooms
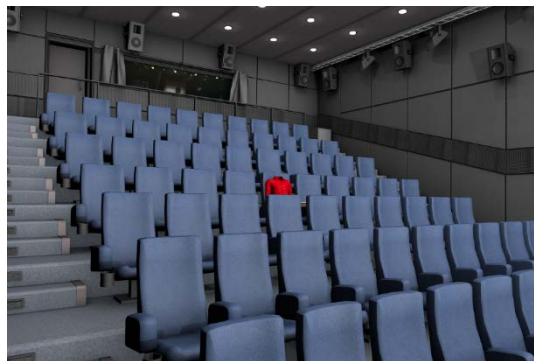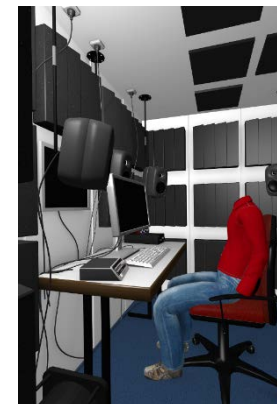

Real-world cinema
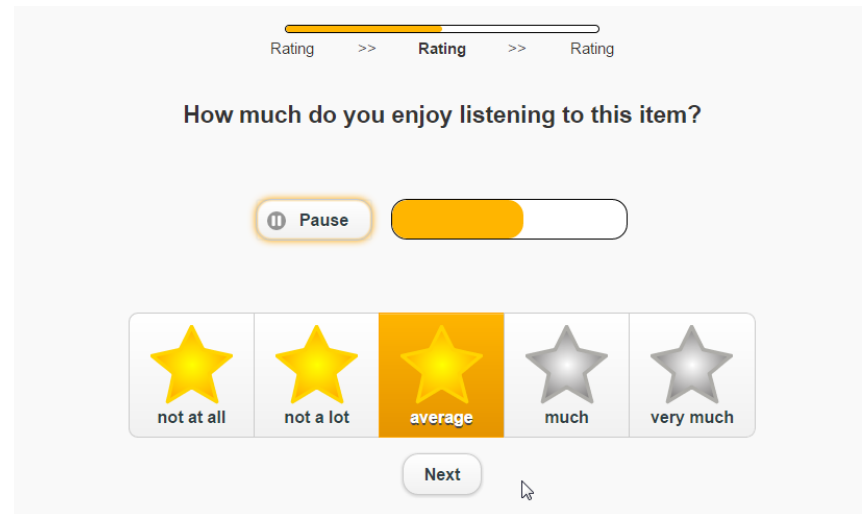

Real-world listening booth

VR cinema

VR listening booth

Procedure (cont.)

- **Second to fifth session**
  - Single-stimulus rating
    - "How much do you enjoy listening to this item?"
  - 5.1 surround sound
  - Room: Listening booth or cinema
  - Environment: "Real world" or VR

  - Additional questions about the room acoustics:
    - Loudness
    - Reverb
    - Distance
    - Liking of bass
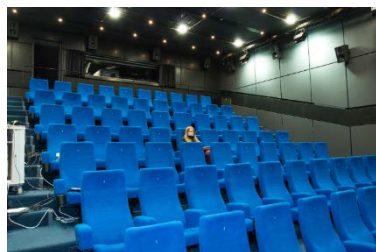   - Liking of treble

Procedure - Overview
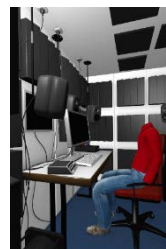
Initial session
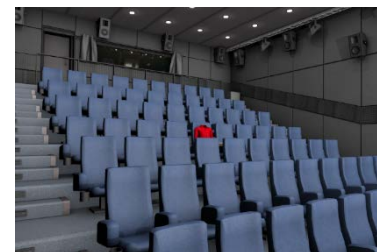
Random order

| Real-world Listening Booth | Real-world Cinema | VR Listening Booth | VR Cinema |

RESULTS

AUDIO LABS
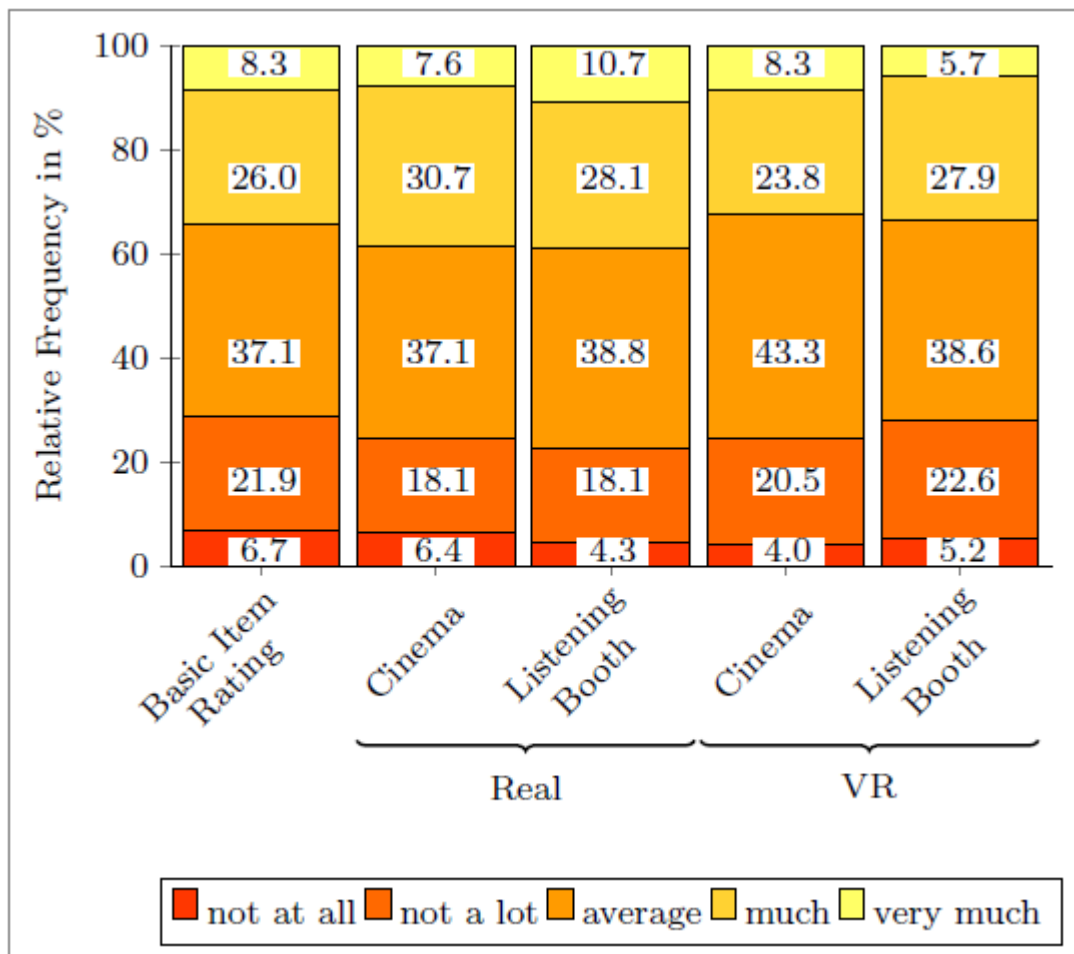
Results

- Statistical analysis: The ratings obtained from the VR sessions are significantly lower than ratings obtained from the real-world sessions
- However, the effect size is very weak

# ADDITIONAL RESULTS

Asking about the following attributes was not the main focus of the experiment. Please see them as results of a pilot study.

Loudness

- **Stimulus: pink noise from center loudspeaker**
  - Remark: Loudness was calibrated to the same level

How loud is this item?

| | very quiet | quiet | normal | loud | very loud |
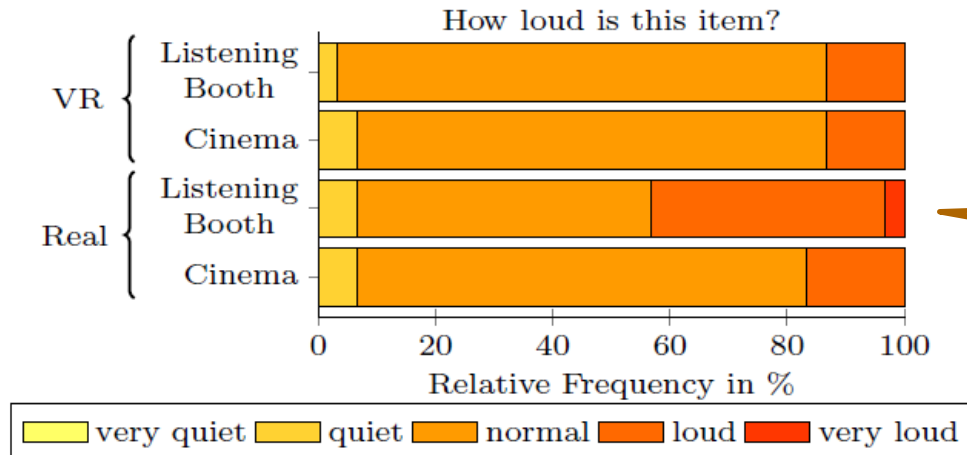
Real-world listening booth was perceived louder

AUDIO LABS
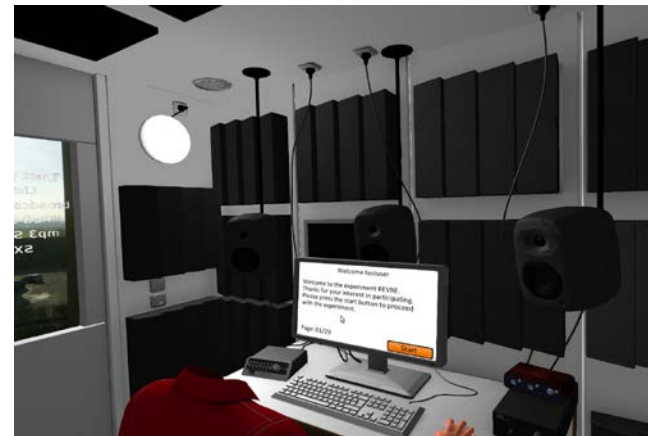
Distance Perception

- Stimulus: pink noise from center loudspeaker

How far away do you perceive the sound?
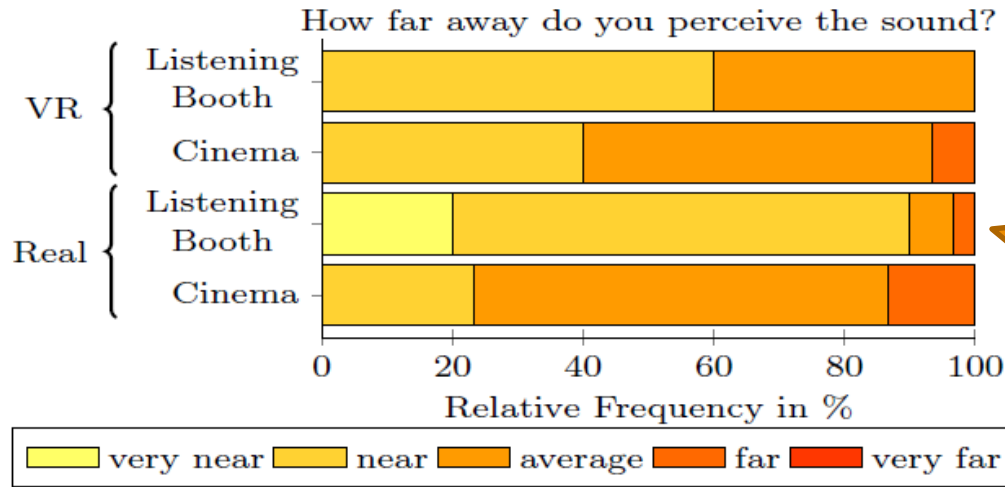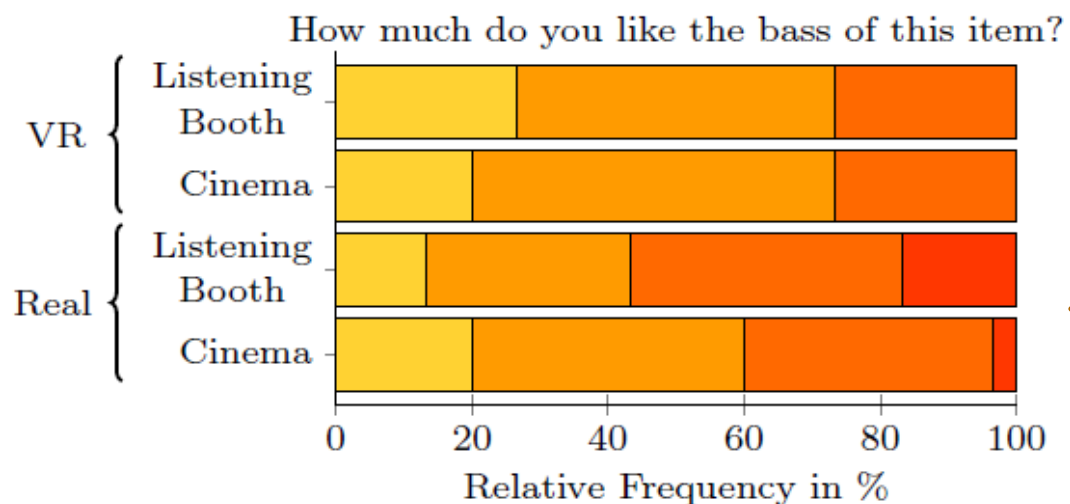
In the Real-world listening, the stimulus was perceived much closer
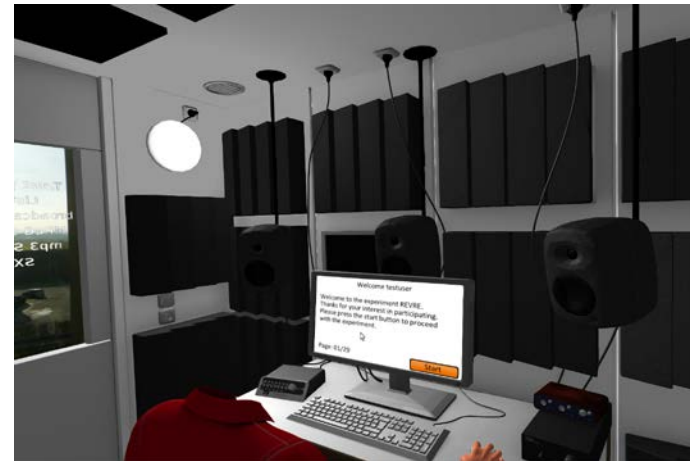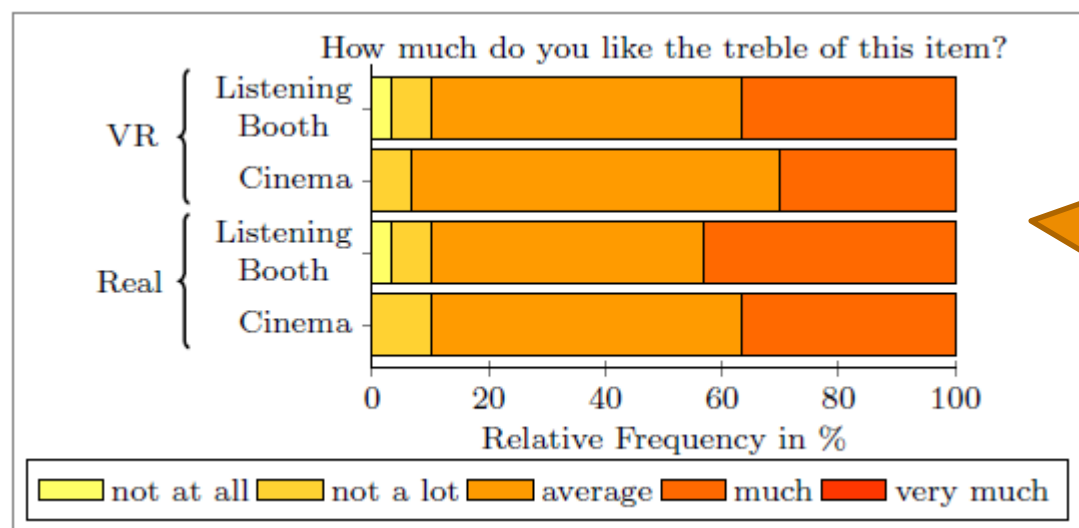
Liking of Bass

- Stimulus: Drums recording (stereo)

Bass is reproduced more "intense" by loudspeakers than by headphones

Liking of Treble

- Stimulus: Drums recording (stereo)

Nice outcome!
There are significant differences between the rooms but no significant differences between the environments.

Conclusion

- A VR system was developed that allows to create virtual scenes for experiments

- Using the system, an experiment was implemented in which ratings given in a VR environment were compared to ratings given in a real-world environment

- In context of overall listening experience, the comparison of the results indicates that the ratings retrieved from the VR experiment are consistent (only very slightly lower) to the ratings retrieved from the real-world experiment

  - Liking of treble seems also very consistent

- For other attributes (loudness, distance,....), there are significant differences between the VR- and real-world enviroment

Outlook

- In order to find out whether Virtual Reality is suited for auditory experiments, more experiments must be carried out

- We focused on overall listening experience => what about other attributes?

- Some attributes turned out to be not perfectly consistent between the VR and real-world environment
  - How can we bring VR experiments to a new level?

Thank you!

DO YOU HAVE ANY QUESTIONS?

More information about the study:
Schoeffler, Gernert, Neumayer, Westphal and Herre: "On the validity of virtual reality-based auditory experiments: a case study about ratings of the overall listening experience", Springer Virtual Reality 19:181-200, 2015.

AUDIO
LABS

Towards a model for predicting overall listening experience

BACKUP SLIDES

Experimental Setup Model and Listener Model
(Schoeffler and Herre, Towards a Listener Model for Predicting the
Overall Listening Experience, Audiomostly, 2014)
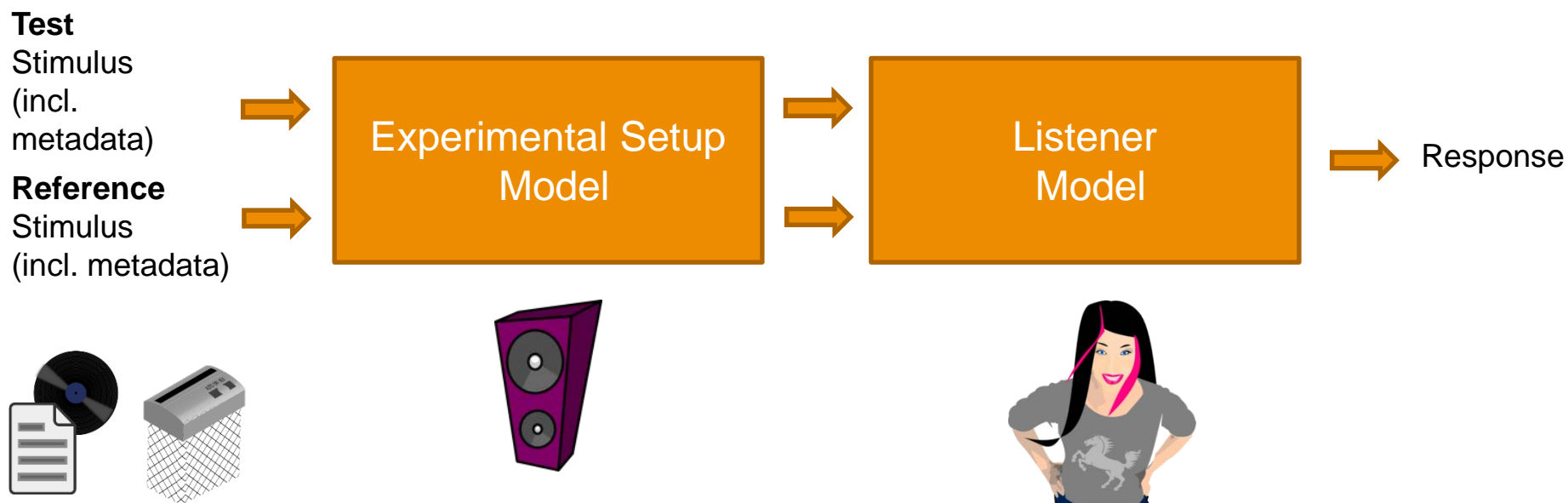
GENERIC LISTENER MODEL

Generic Listener Model

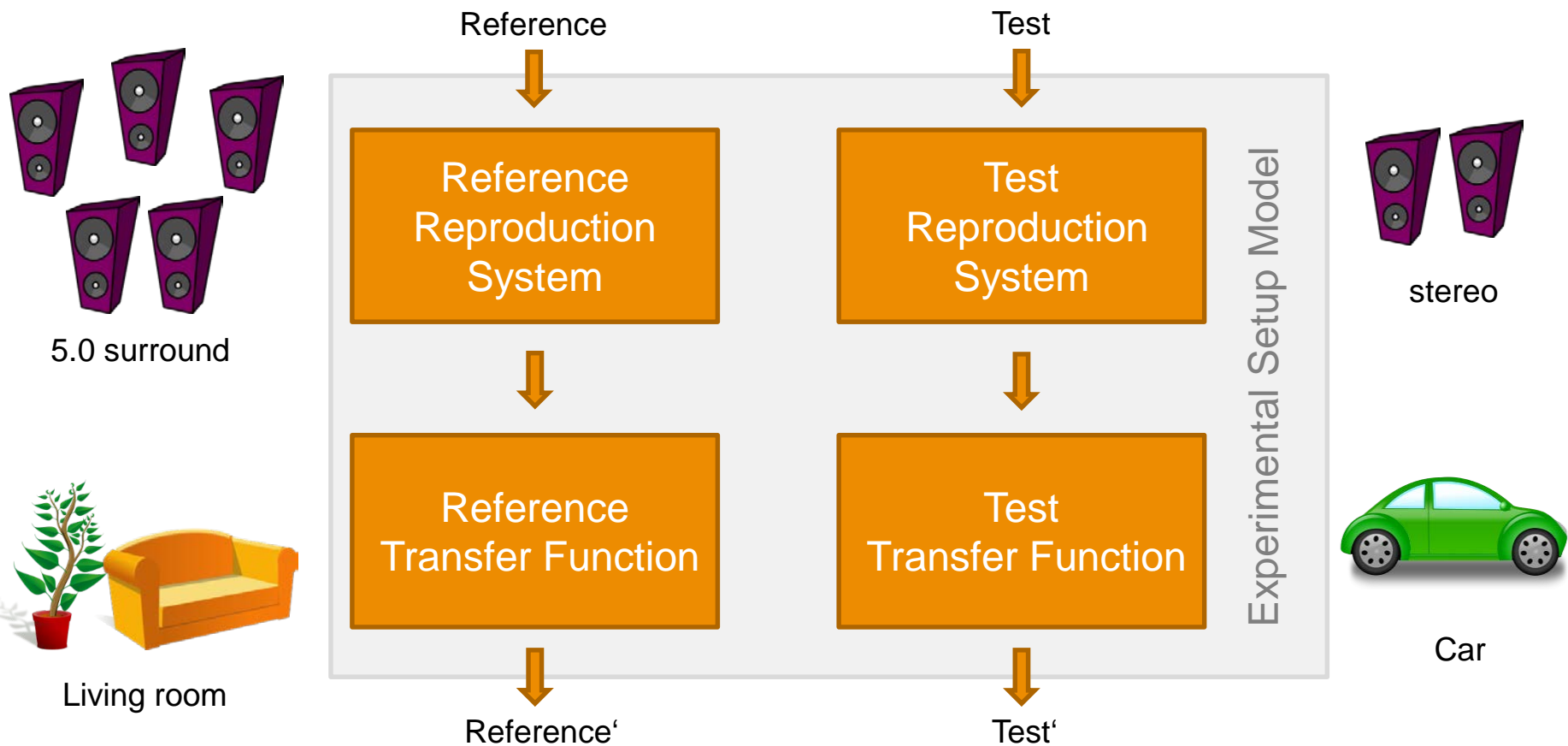- "Experiment workflow" as an example for the model

Audio Quality Degradation
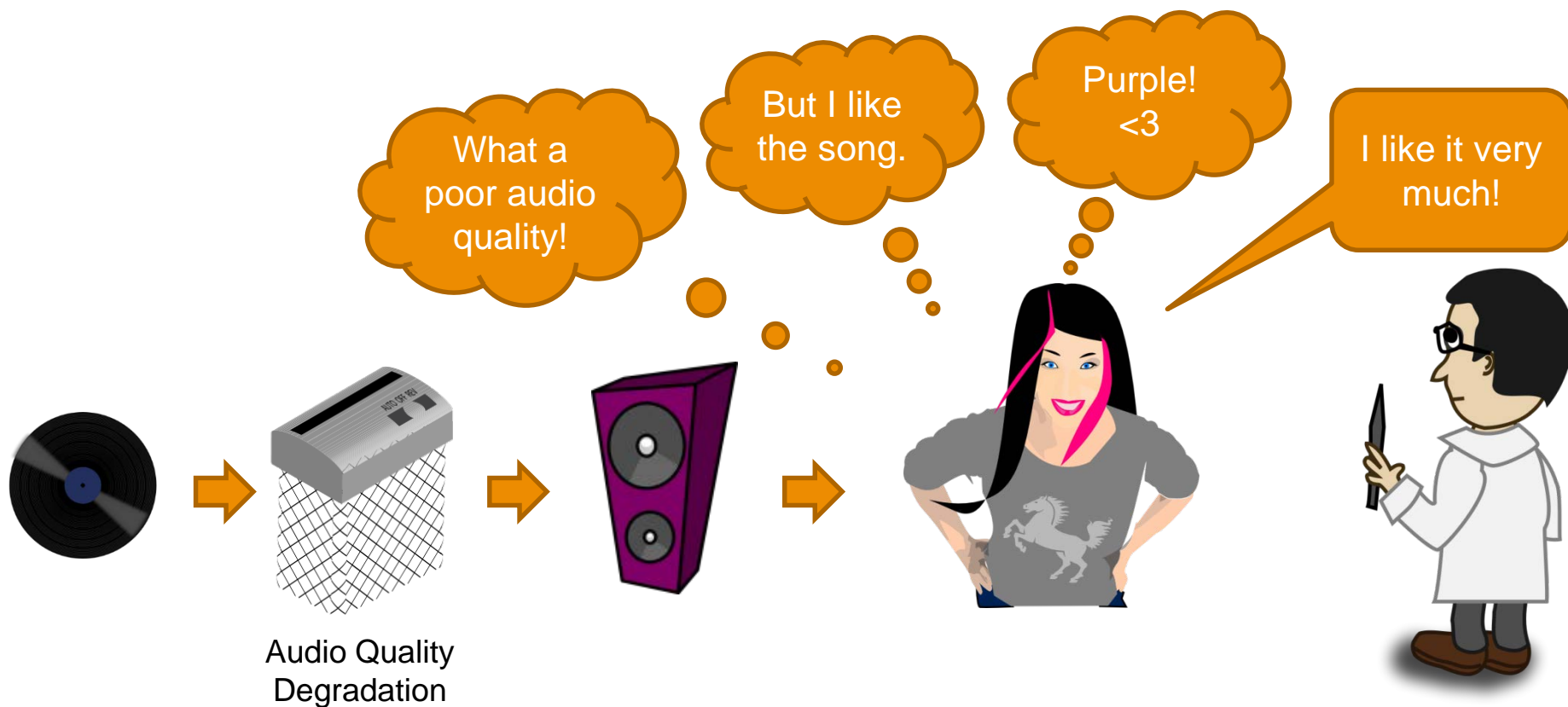
Generic Model

- The generic model covers many use cases of auditory experiments (not only rating the OLE)
- Listener rates relative to a reference
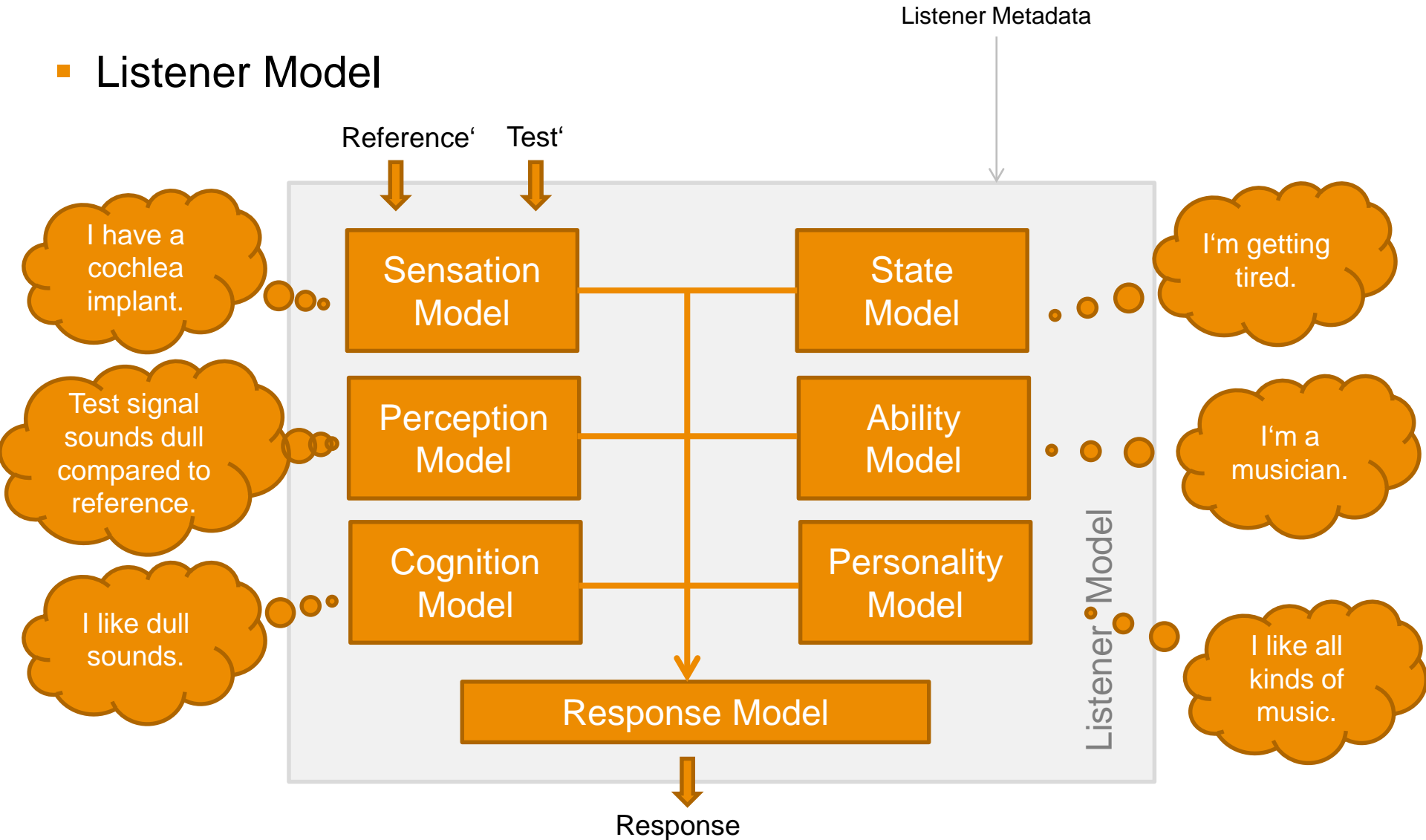
Test
Stimulus (incl. metadata)

Reference
Stimulus (incl. metadata)

Experimental Setup Model

Listener Model

Response

Generic Model

- **Experimental Setup Model**

5.0 surround

Living room

Reference

Test

| Reference Reproduction System | Test Reproduction System |

| Reference Transfer Function | Test Transfer Function |

Experimental Setup Model

stereo

Car

Reference'

Test'

AUDIO LABS

Generic Model

- **Listener Model**

Generic Model

- A closer look into the sub-models...

Model X
e.g. Perception

Model Y
e.g. Cognition
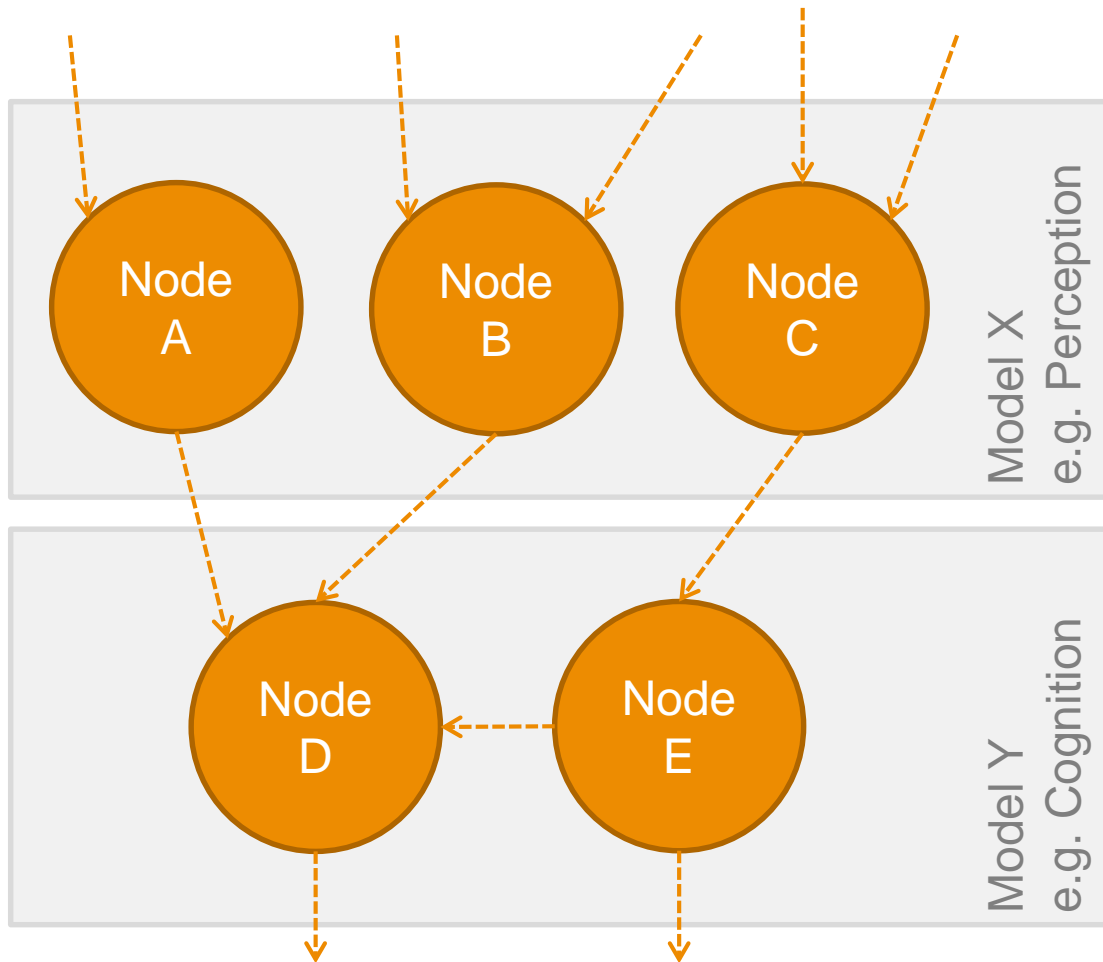
A node processes an output by using none or many inputs.

Example
Node C:
Input: *wave form of Test and Reference*
Output: *spectral centroids*

Node E:
Input: *spectral centroids*
Output: *Do I care about the difference in signal bandwidth [yes, no]*

Nodes can be everything (regression models,... etc)