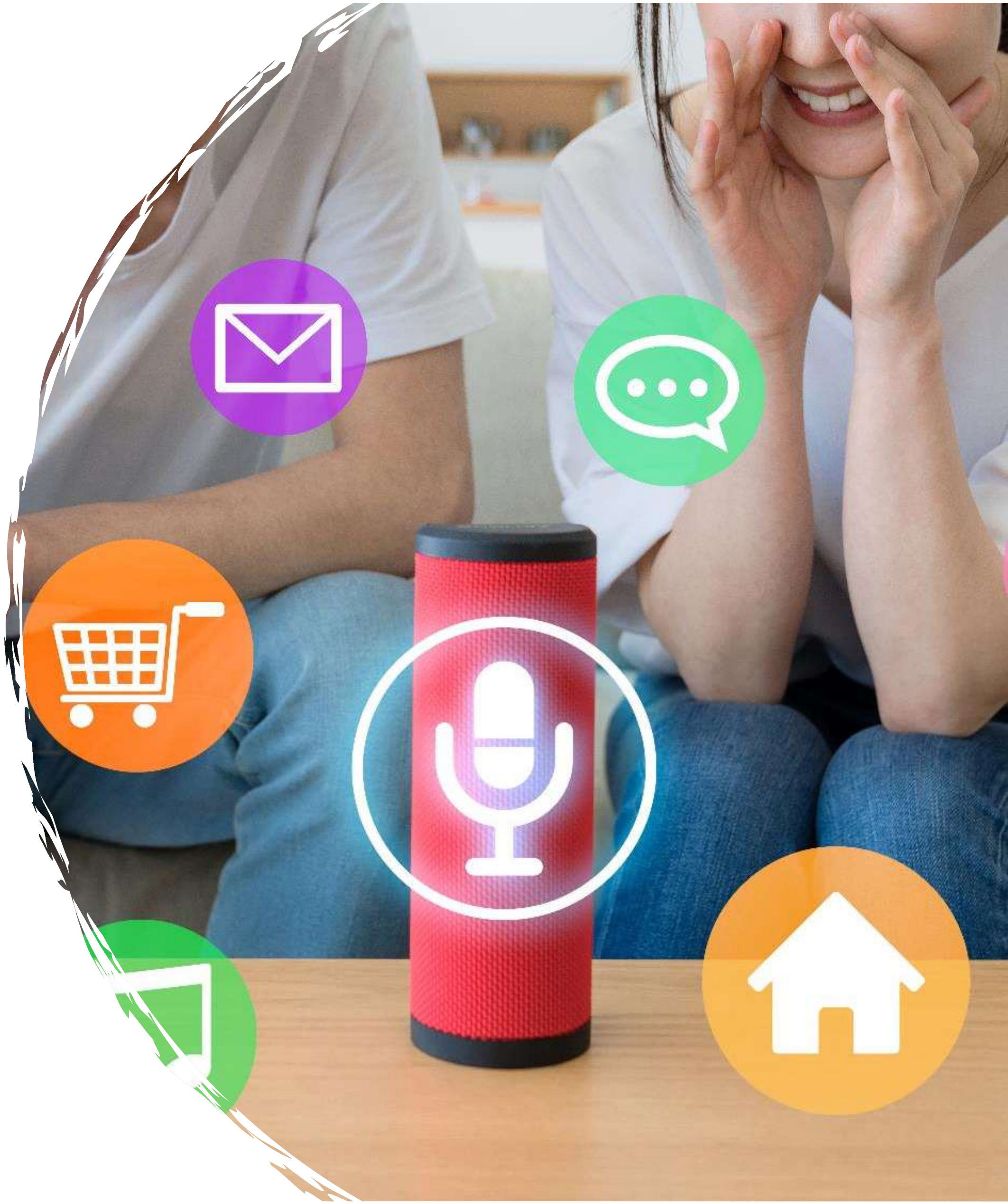
STQ Workshop

ASR Testing in Reverberant Environment

Frank Kettler, HEAD acoustics GmbH

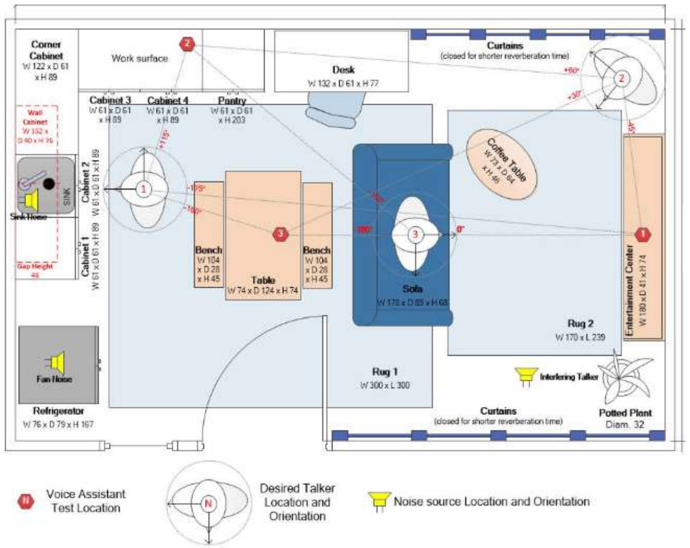
21/11/2022

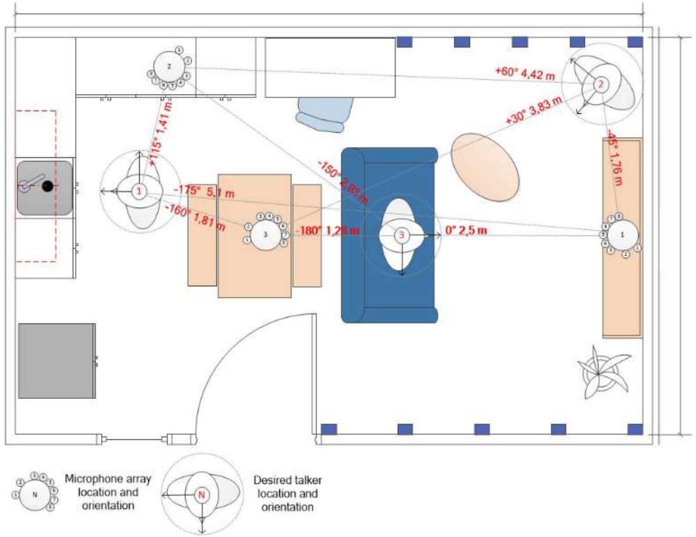


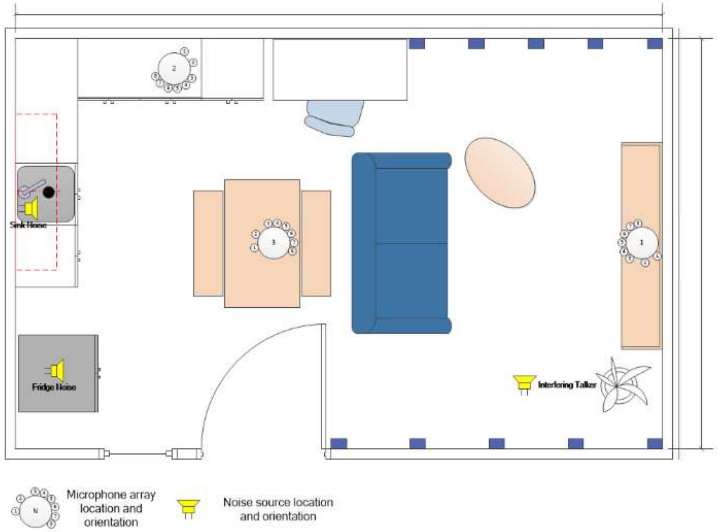
Overview

- Motivation
- Audio Material, Processing and Analysis
- WER / COR Analysis for ASR
    - Influence of Reverberation
    - Comparison of ASR Engines
- Conclusions

ASR Testing Environment, Standardization


Corner Cabinet W 122 x D 61 x H 89
Work surface
Desk W 132 x D 61 x H 77
Curtains (closed for shorter reverberation time)
Wall Cabinet W 152 x D 40 x H 76
Cabinet 3 W 61 x D 61 x H 89
Cabinet 4 W 61 x D 61 x H 89
Pantry W 61 x D 61 x H 203
Cabinet 2 W 61 x D 61 x H 89
Cabinet 1 W 61 x D 61 x H 89
SINK
Sink Noise
Gap Height 46
+115°
-175°
-160°
+60°
+30°
-150°
-180°
0°
-55°
Coffee Table W 73 x D 64 x H 46
Bench W 104 x D 28 x H 45
Table W 74 x D 124 x H 74
Bench W 104 x D 28 x H 45
Sofa W 178 x D 83 x H 68
Entertainment Center W 180 x D 41 x H 74
Rug 2 W 170 x L 239
Fan Noise
Refrigerator W 76 x D 79 x H 167
Rug 1 W 300 x L 300
Interfering Talker
Curtains (closed for shorter reverberation time)
Potted Plant Diam. 32
Voice Assistant Test Location
Desired Talker Location and Orientation
Noise source Location and Orientation




Sink Noise
Fridge Noise
Interfering Talker
Microphone array location and orientation
Noise source location and orientation

ETSI TS 103 504 V1.1.1 (2020-07)
Methods and procedures for **evaluating performance** of voice-controlled devices and functions: far talk voice assistant devices

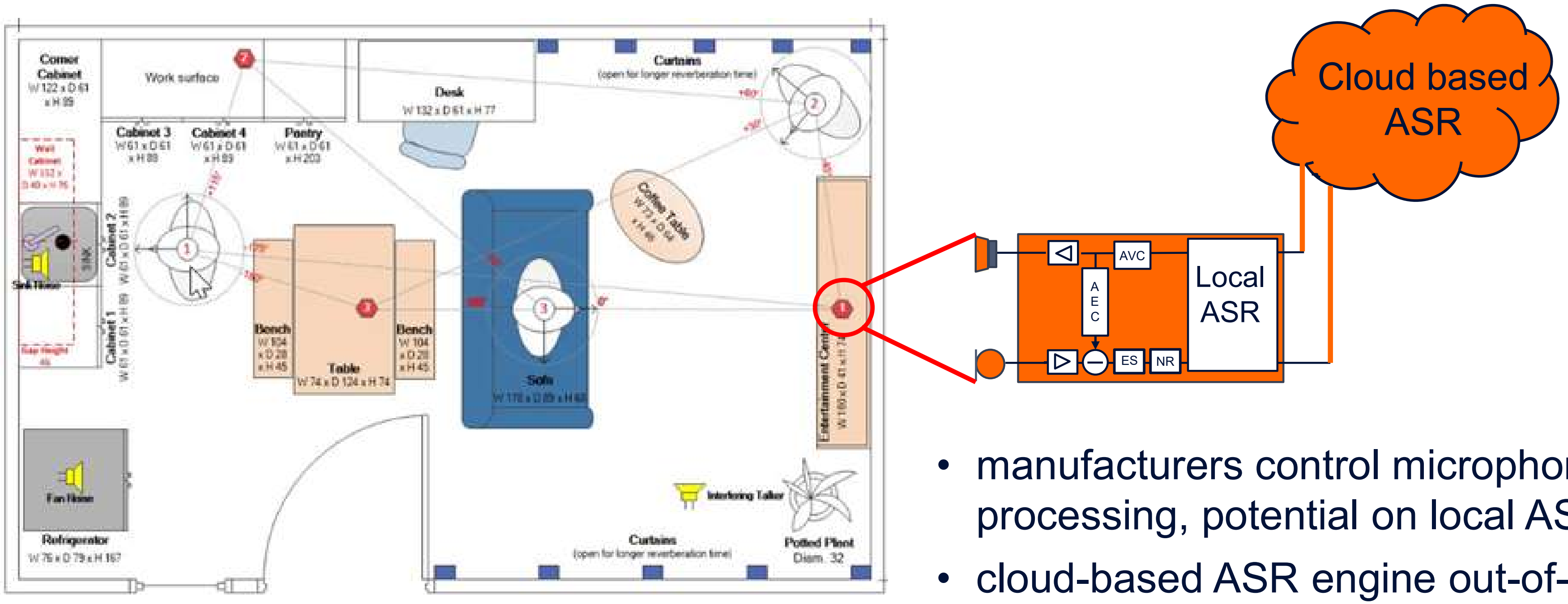
(incl. vehicular acoustics environment)

ETSI TS 103 557 V1.3.1 (2020-03)
Methods for **reproducing reverberation** for communication device measurements

ETSI TS 103 224 V1.6.1 (2022-03)
A **sound field reproduction** method for terminal testing including a background noise database

(incl. interior vehicle noise)

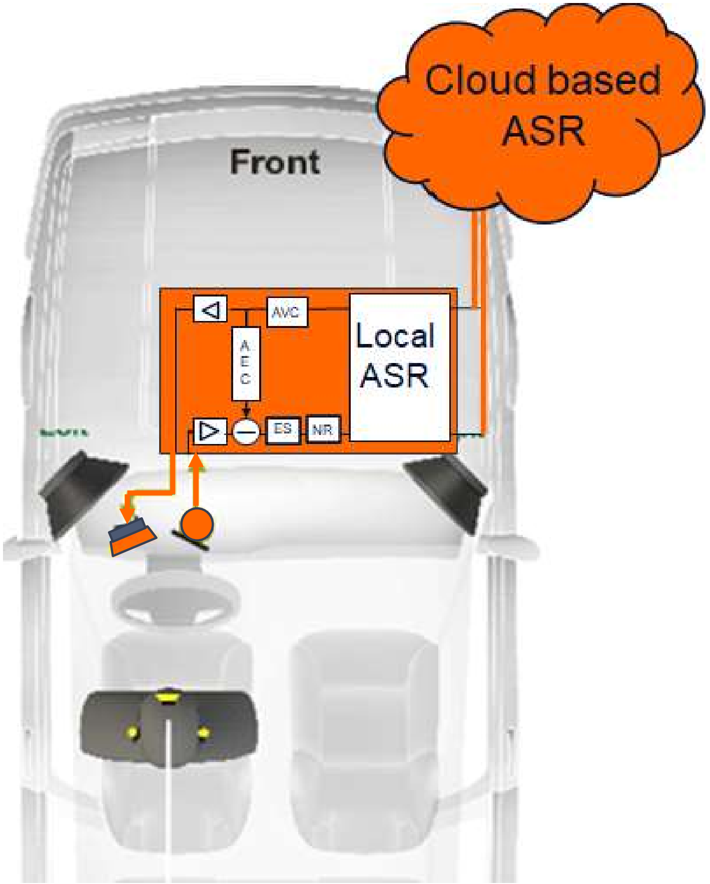
ASR Testing for Smart Home Devices


Corner Cabinet W 122 x D 61 x H 89
Work surface
Desk W 132 x D 61 x H 77
Curtains (open for longer reverberation time)
Cabinet 3 W 61 x D 61 x H 89
Cabinet 4 W 61 x D 61 x H 89
Pantry W 61 x D 61 x H 203
Cabinet 2
Cabinet 1
Coffee Table
Sink Noise
Bench W 104 x D 28 x H 45
Table W 74 x D 124 x H 74
Bench W 104 x D 28 x H 45
Sofa
Entertainment Center
Fan Noise
Refrigerator W 76 x D 79 x H 167
Interfering Talker
Curtains (open for longer reverberation time)
Potted Plant Diam. 32
Cloud based ASR
AVC
A E C
ES
NR
Local ASR

- manufacturers control microphone(s), signal pre-processing, potential on local ASR engine
- cloud-based ASR engine out-of-control
- optimization often iterative process, “try-and-error”…

➢ desirable tuning hints for acoustic pre-processing

ASR Testing for Vehicle-mounted Applications

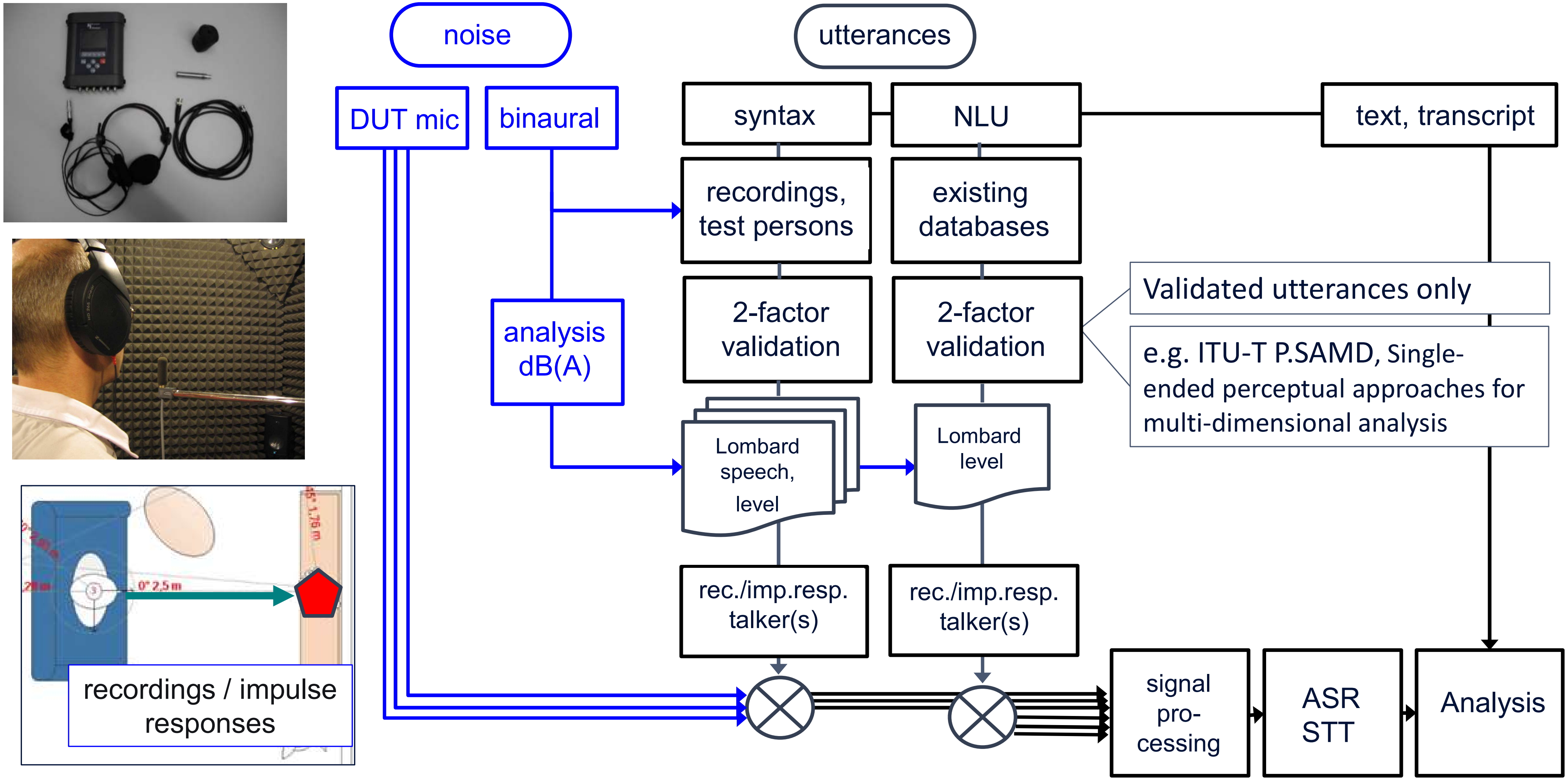


- car manufactures control microphone(s), signal pre-processing, potential on local ASR engine
- cloud-based ASR engine out-of-control
- certification tests often require test repetitions, sometimes “try-and-error”…

➢ desirable tuning hints for acoustic pre-processing

<u>before certification tests</u>

Principle



recordings / impulse responses

noise

DUT mic

binaural

analysis dB(A)

utterances

syntax

NLU

text, transcript

recordings, test persons

existing databases

2-factor validation

2-factor validation

Validated utterances only

e.g. ITU-T P.SAMD, Single-ended perceptual approaches for multi-dimensional analysis

Lombard speech, level

Lombard level

rec./imp.resp. talker(s)

rec./imp.resp. talker(s)

signal pro-cessing

ASR STT

Analysis

ASR Analysis and Terminology

Edit distance (Levenshtein distance)

Word Error Rate (WER) / Correct Word Rate (COR)

Insertion (INS): “I AM NOT GOING TO CHARGE YOU ANYTHING NOW SHE SAID”
“I AM NOT GOING TO CHARGE YOU ANYTHING NOW IF SHE SAID”

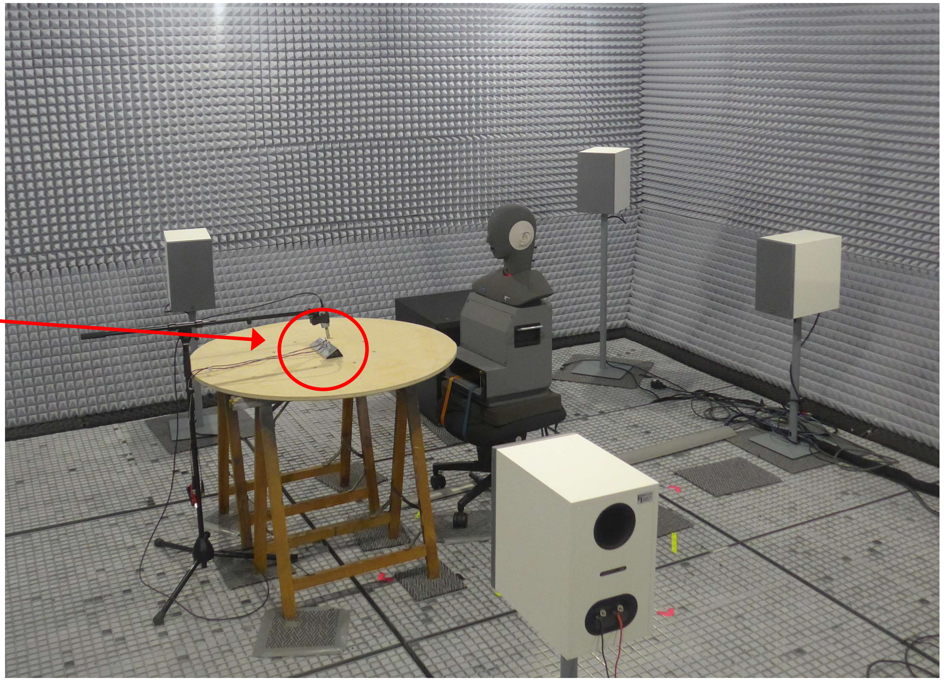
Substitution (SUB): “JUST BY LOOKING AT THEM”
“JUST ME LOOKING AT THEM”

Deletion (DEL): “IT WAS SEEN EARLY IN THE MORNING RUSHING OVER EASTWARD”
“ ……… EARLY IN THE MORNING RUSHING OVER …… “

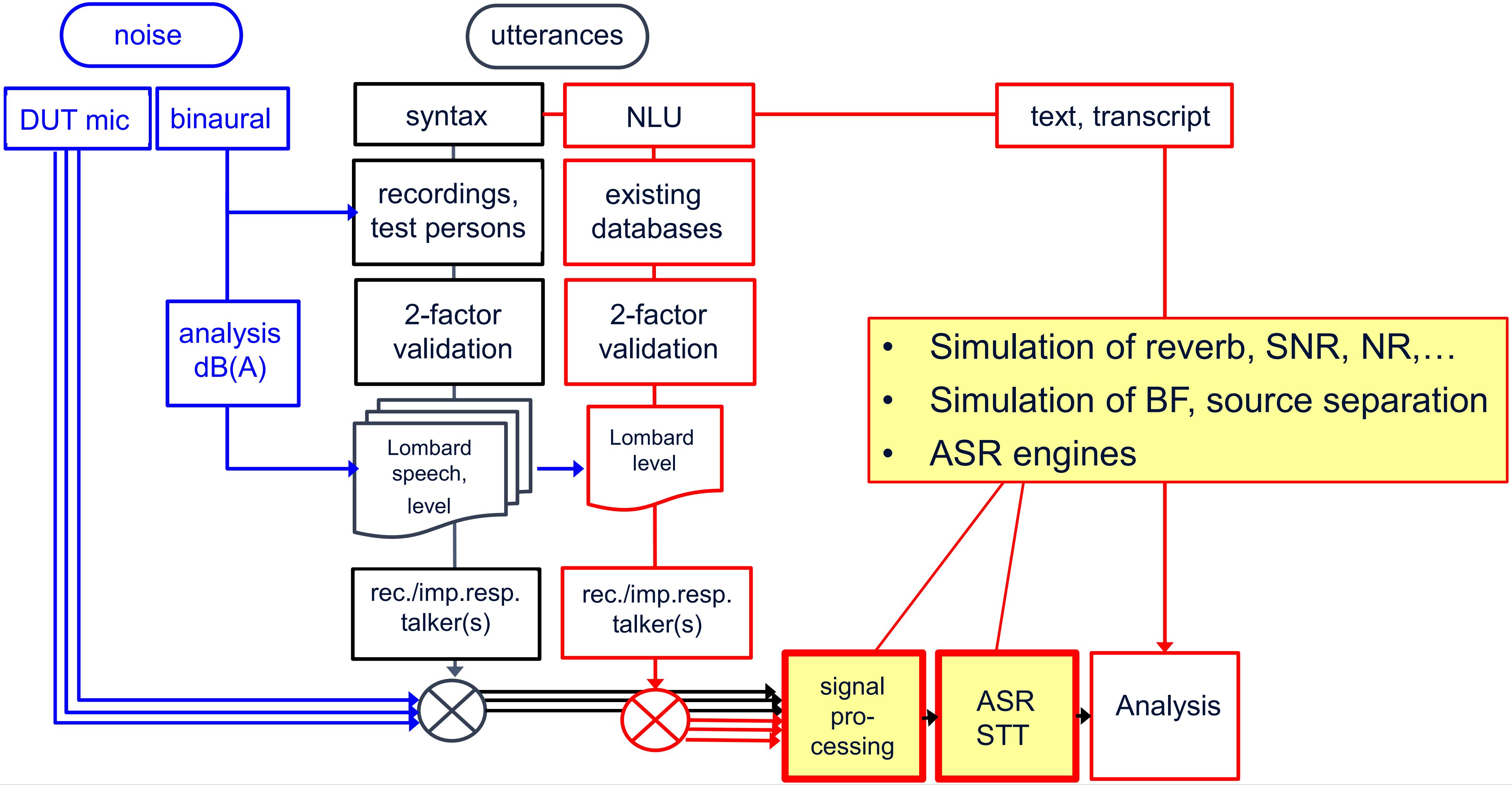
Audio Material, Example NLU

- utterances from Mozilla Common Voice Project, English language
- Mozilla validation, i.e. no contradictions
- ITU-T P.SAMD scores > 3.5 (on 5-point MOS scale)
- 269 utterances, 46 speaker, appr. 2,800 single words

- speech / reverb recordings acc. TS 103 557
- single channel DUT mic
- room noise acc. TS 103 224
- Lombard level considered
- fullband audio files, downsampled 16 kHz



Principle



noise

utterances

DUT mic

binaural

syntax

NLU

text, transcript

recordings, test persons

existing databases

analysis dB(A)

2-factor validation

2-factor validation

- Simulation of reverb, SNR, NR,…
- Simulation of BF, source separation
- ASR engines

Lombard speech, level

Lombard level

rec./imp.resp. talker(s)

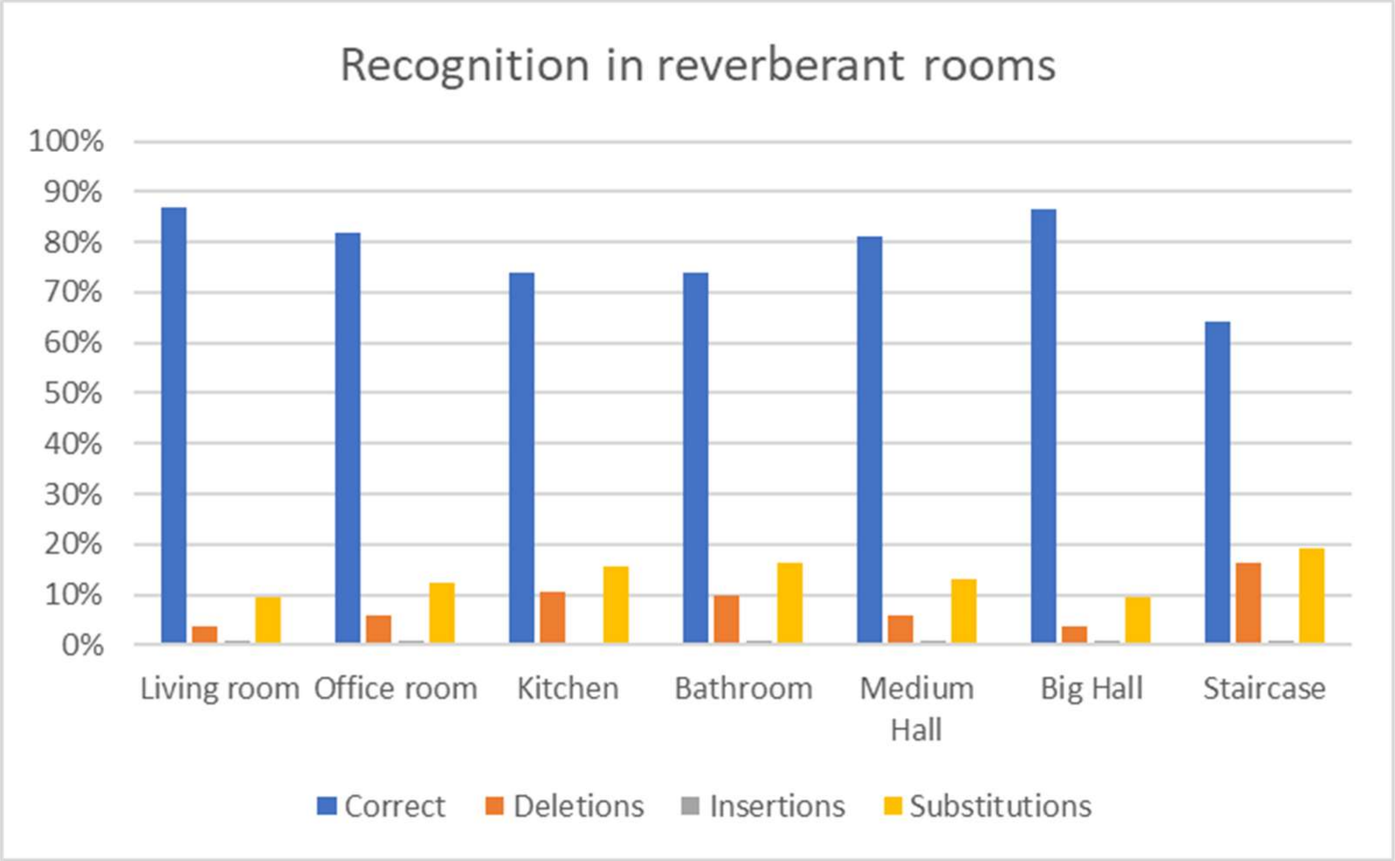
rec./imp.resp. talker(s)

signal pro-cessing

ASR STT

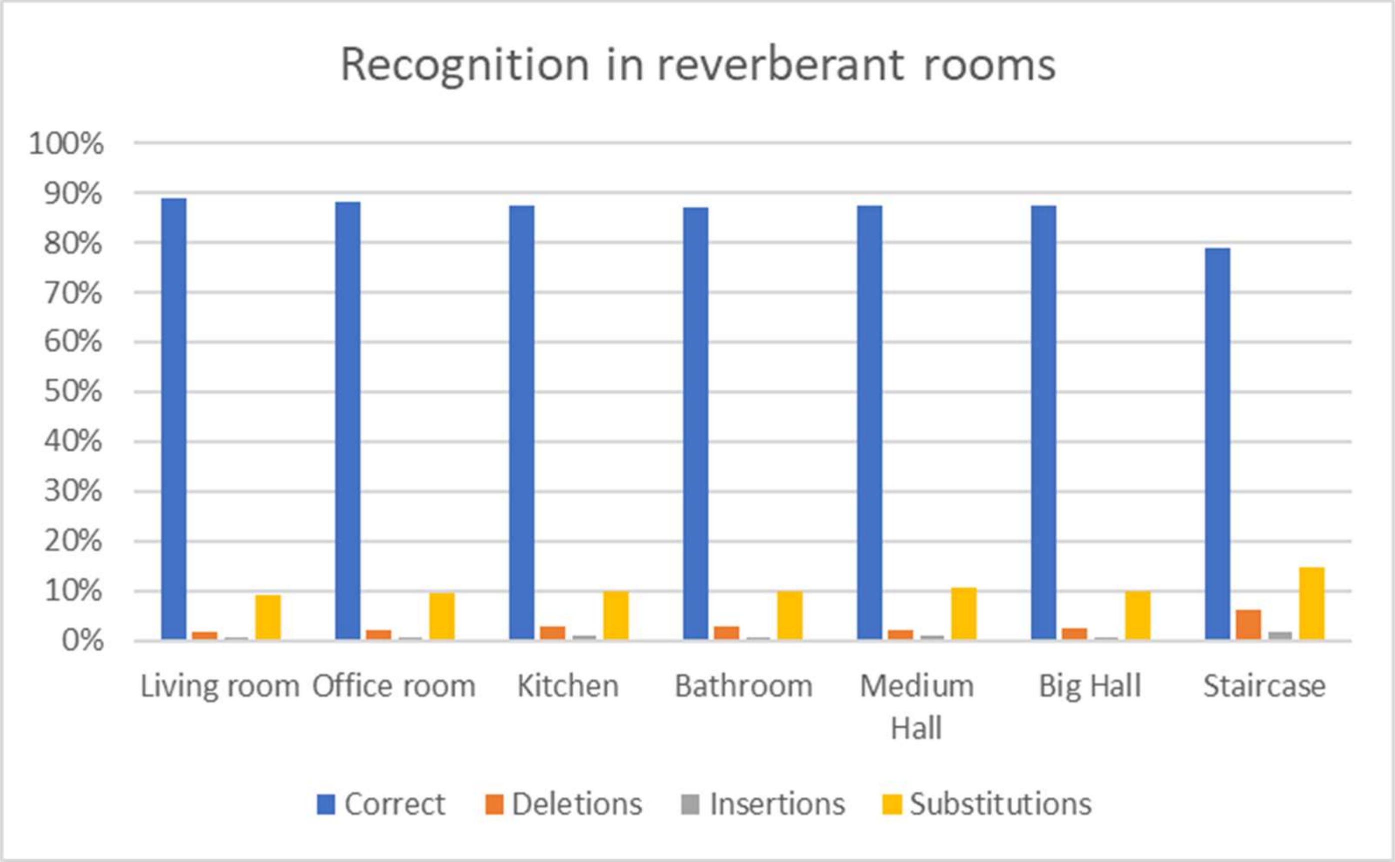
Analysis

Recognition ASR Engine I, Single Talker



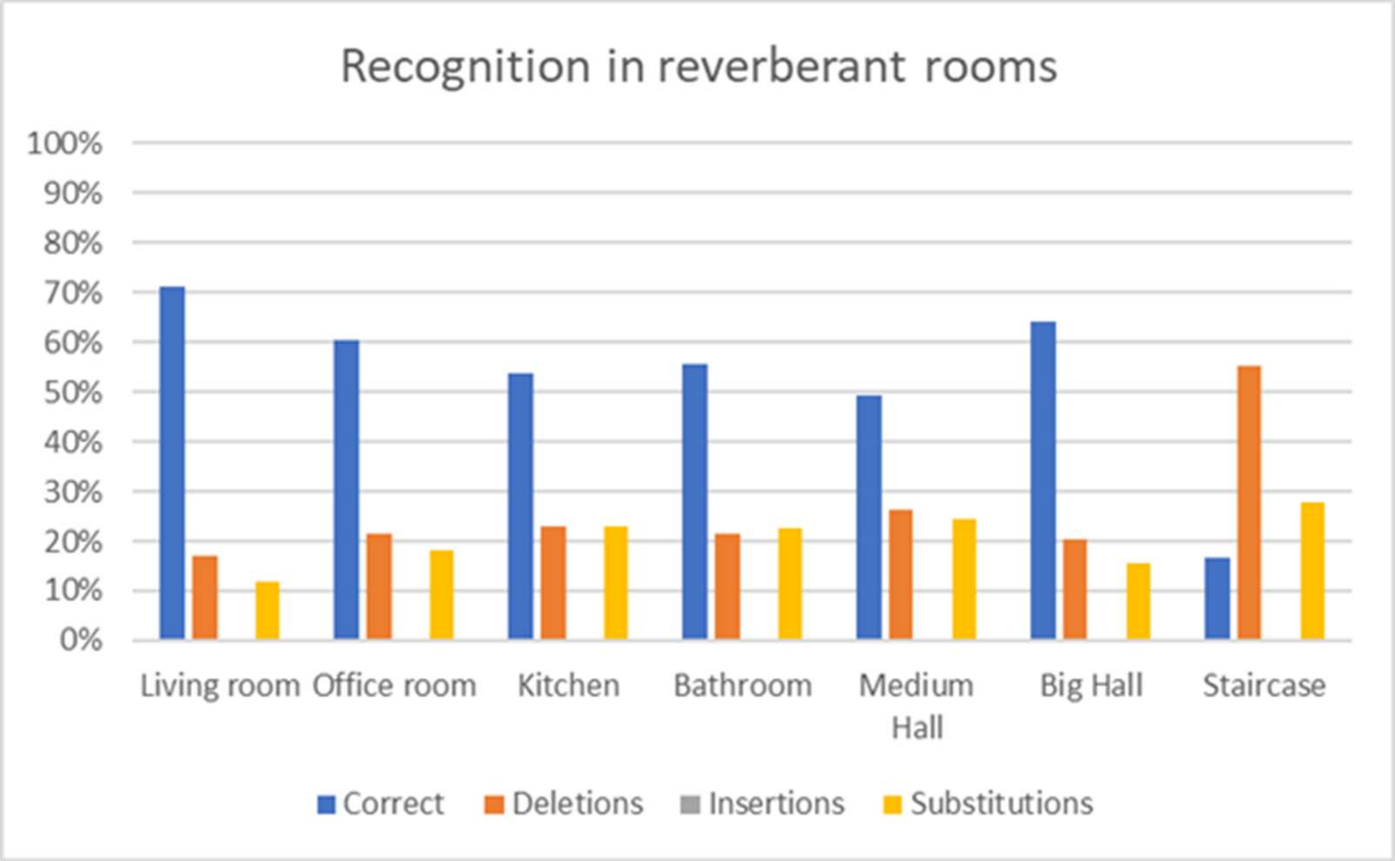
- COR between 87% and 65%
- minimum for Kitchen and Bathroom characteristics, obviously not only RT_{60} relevant
- SUB dominate the WER (…20%)
- DEL relevant (10 – 16%)
- INS irrelevant, no erroneous insertions (single talker scenario)

Recognition ASR Engine II, Single Talker



- COR between 89% and 79%
- very robust against reverb
- SUB dominant
- DEL play minor role
- INS irrelevant (single talker)

Recognition ASR Engine III, Single Talker



- COR between 71% and 17%
- least reliable ASR engine
- SUB and DEL relevant
- DEL get dominant
- INS = 0 (single talker)

Robustness of ASR Engines

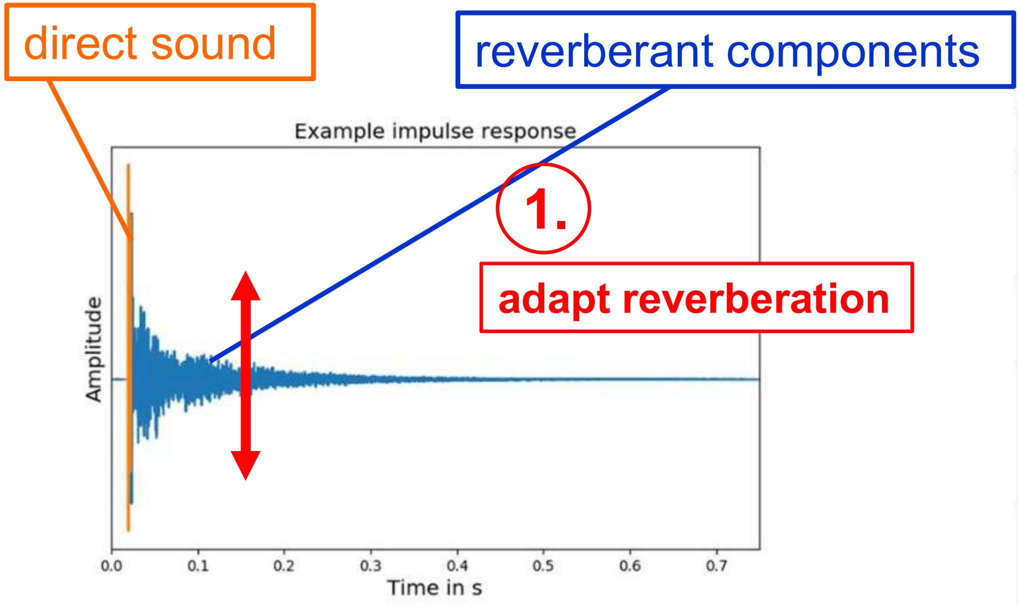
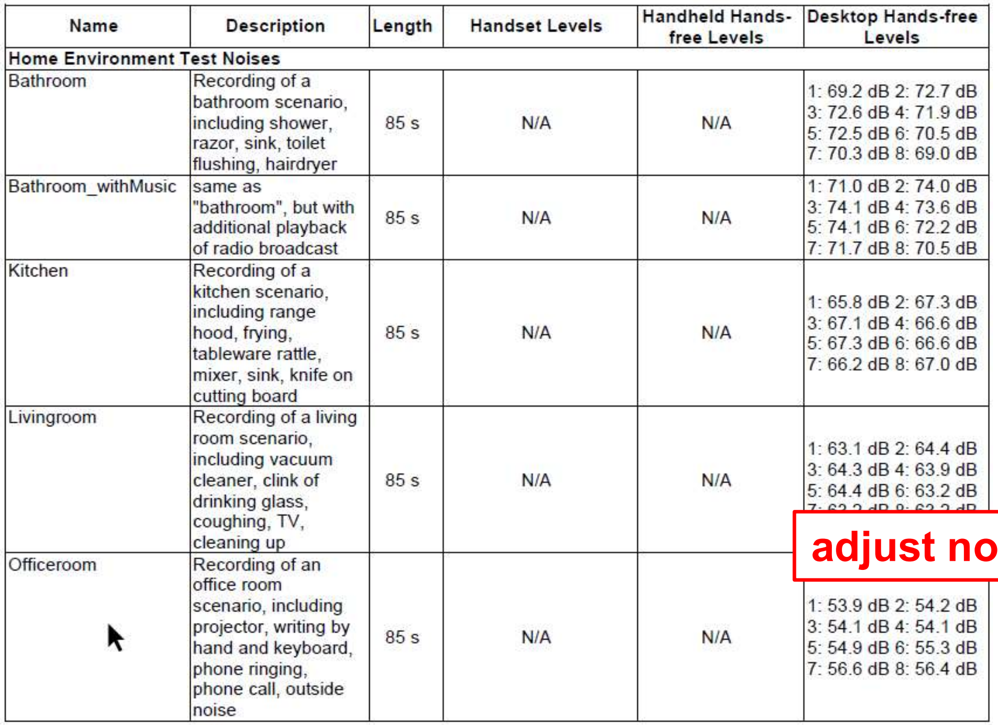


Figure 2: Time domain representation of an example impulse response

Impulse responses from ETSI TS 103 557



Name	Description	Length	Handset Levels	Handheld Hands-free Levels	Desktop Hands-free Levels
Home Environment Test Noises					
Bathroom	Recording of a bathroom scenario, including shower, razor, sink, toilet flushing, hairdryer	85 s	N/A	N/A	1: 69.2 dB 2: 72.7 dB 3: 72.6 dB 4: 71.9 dB 5: 72.5 dB 6: 70.5 dB 7: 70.3 dB 8: 69.0 dB
Bathroom_withMusic	same as "bathroom", but with additional playback of radio broadcast	85 s	N/A	N/A	1: 71.0 dB 2: 74.0 dB 3: 74.1 dB 4: 73.6 dB 5: 74.1 dB 6: 72.2 dB 7: 71.7 dB 8: 70.5 dB
Kitchen	Recording of a kitchen scenario, including range hood, frying, tableware rattle, mixer, sink, knife on cutting board	85 s	N/A	N/A	1: 65.8 dB 2: 67.3 dB 3: 67.1 dB 4: 66.6 dB 5: 67.3 dB 6: 66.6 dB 7: 66.2 dB 8: 67.0 dB
Livingroom	Recording of a living room scenario, including vacuum cleaner, clink of drinking glass, coughing, TV, cleaning up	85 s	N/A	N/A	1: 63.1 dB 2: 64.4 dB 3: 64.3 dB 4: 63.9 dB 5: 64.4 dB 6: 63.2 dB 7: 63.2 dB 8: 63.2 dB
Officeroom	Recording of an office room scenario, including projector, writing by hand and keyboard, phone ringing, phone call, outside noise	85 s	N/A	N/A	1: 53.9 dB 2: 54.2 dB 3: 54.1 dB 4: 54.1 dB 5: 54.9 dB 6: 55.3 dB 7: 56.6 dB 8: 56.4 dB

2.

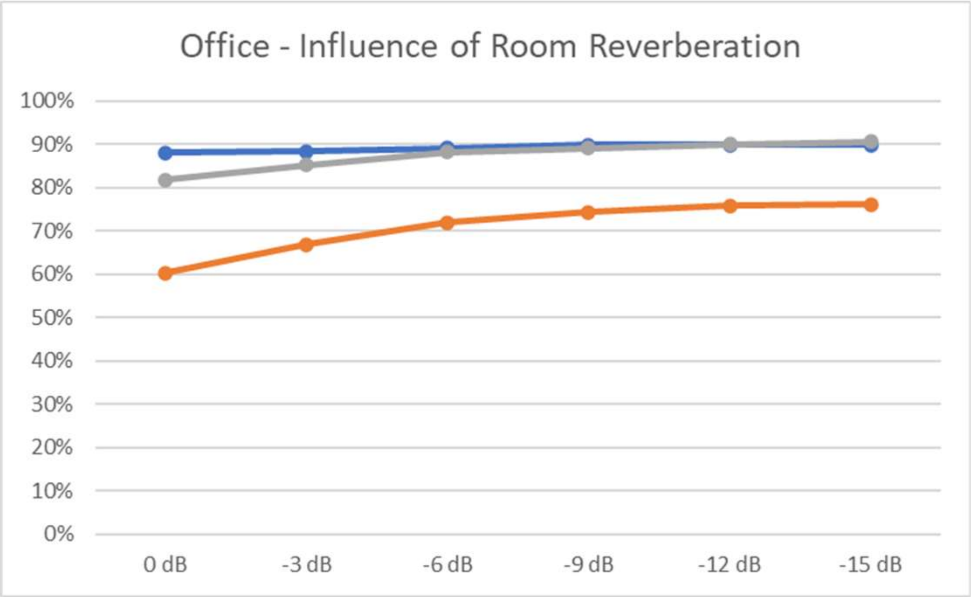
adjust noise level

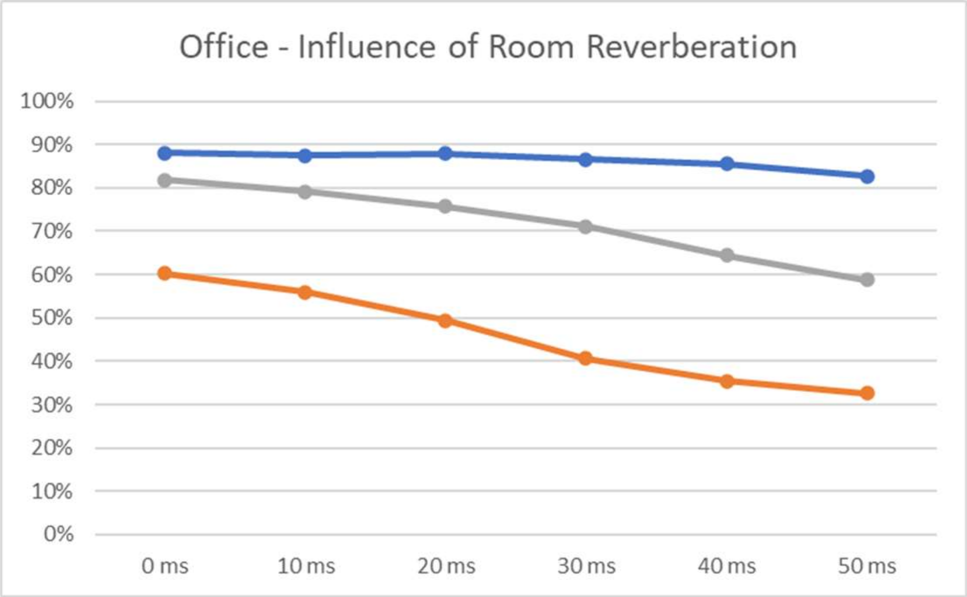
Background noise scenarios from ETSI TS 103 224

Robustness of ASR Engines I, II, III

ASR I ASR II ASR III

Example: Office room





- sensitivity on reverb component

  ASR III > ASR I > ASR II

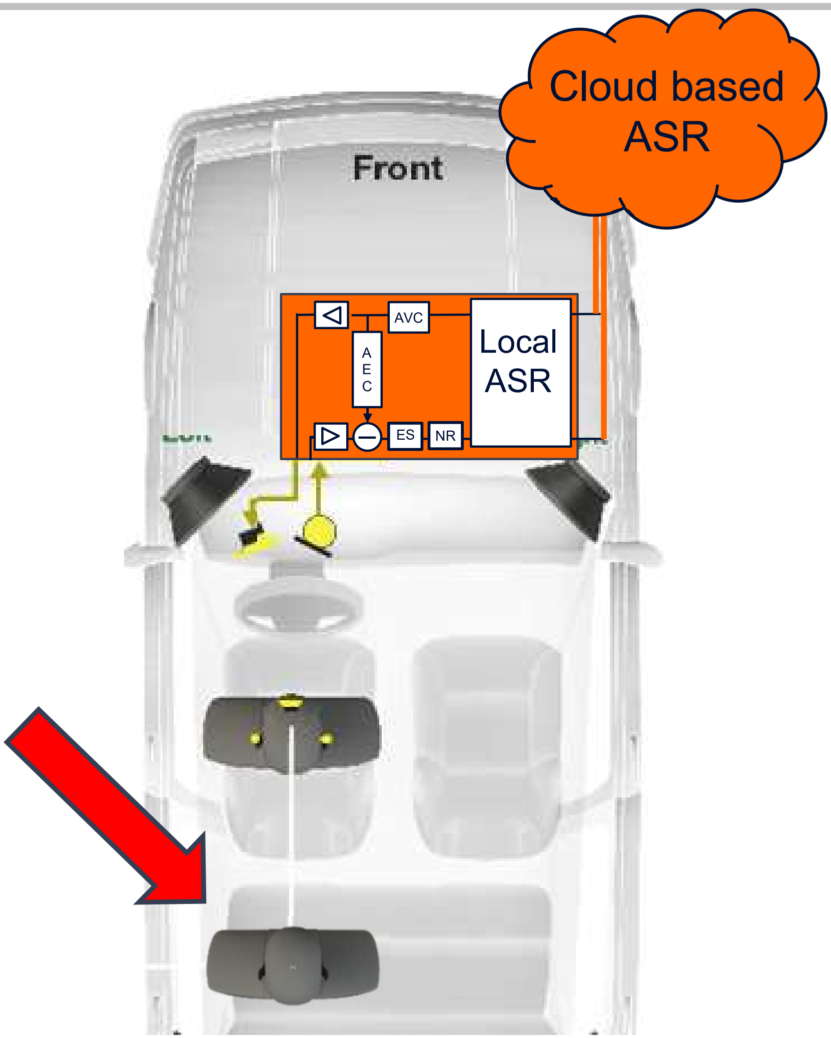
  estimate benefit of preprocessing de-reverb

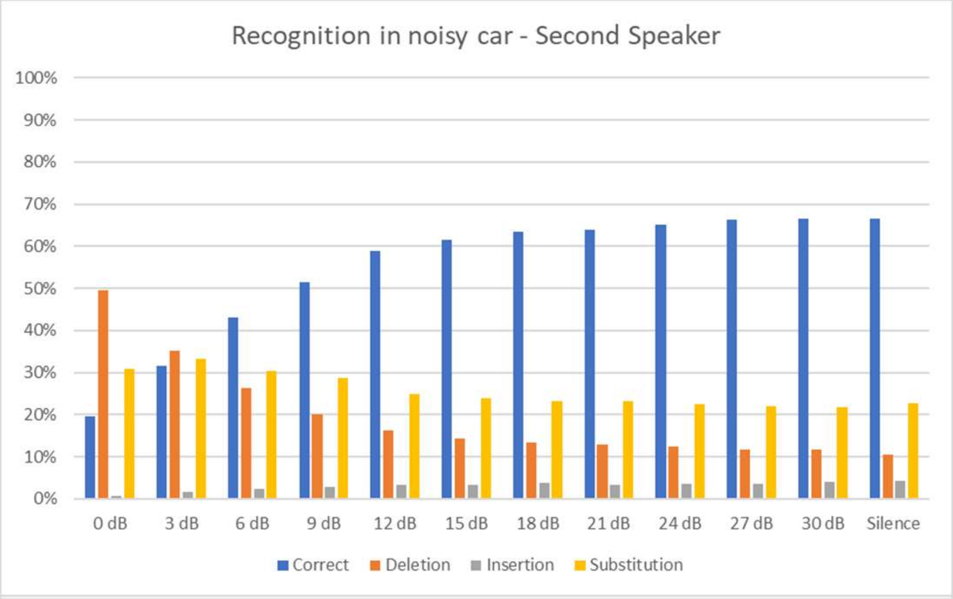
- …on time-shift of reverb component

  ASR III > ASR I > ASR II

  …..

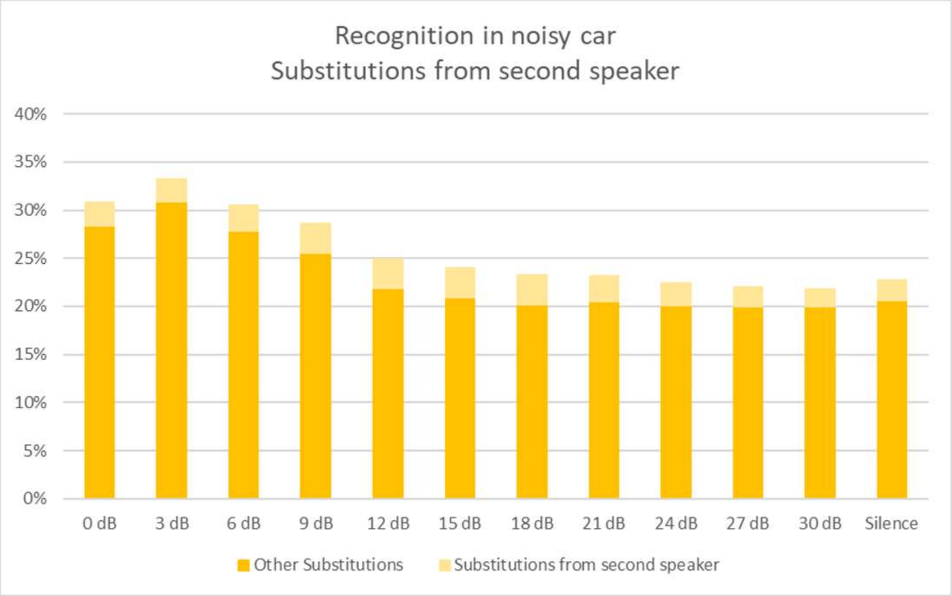
ASR in Vehicle, Concurrent Talker Scenario

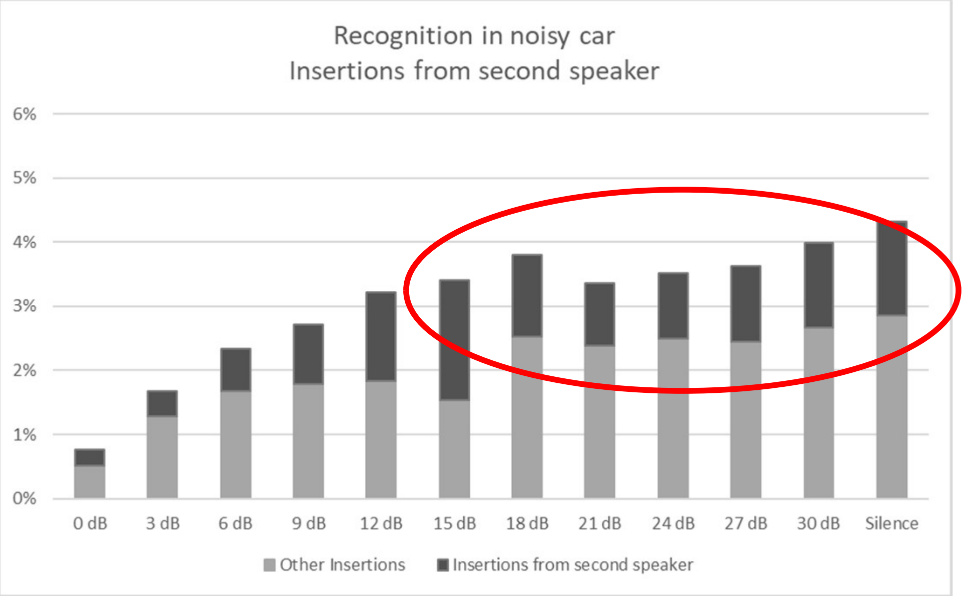




- 2nd talker (backseat, -6 dB) reduces COR 90% → < 70%
- Contributions of DEL and SUB increase accordingly
- Low SNR: DEL dominant, high SNR: SUB dominant
- INS increase (2nd talker scenario)

ASR in Vehicle, Concurrent Talker Scenario





- SUB caused by 2nd talker negligible

- INS caused by 2nd talker only arround 30%

Conclusions

- highlights **interaction between acoustic pre-processing and ASR** engine
- provides additional numbers (→ DEL, INS, SUB)

- allows
  - **comparison** performance testing of ASR engines
  - the **adaptation of acoustic pre-processing on specific ASR engines**
  - helps to detect limits and thresholds and balance implementation effort
  - …thus, can also help to **steer development process**

- beneficial ahead **of certification tests**

HEAD acoustics GmbH
Ebertstraße 30a
52134 Herzogenrath
Germany

info@head-acoustics.de
www.head-acoustics.com

Follow us on