ITU-T Workshop on
"From Speech to Audio: bandwidth extension, binaural perception"
Lannion, France, 10-12 September 2008

OPTIMUM FREQUENCY RESPONSE CHARACTERISTICS FOR WIDEBAND TERMINALS

H. W. Gierlich, F. Kettler, S. Poschen

HEAD acoustics GmbH

A. Raake, S. Spors

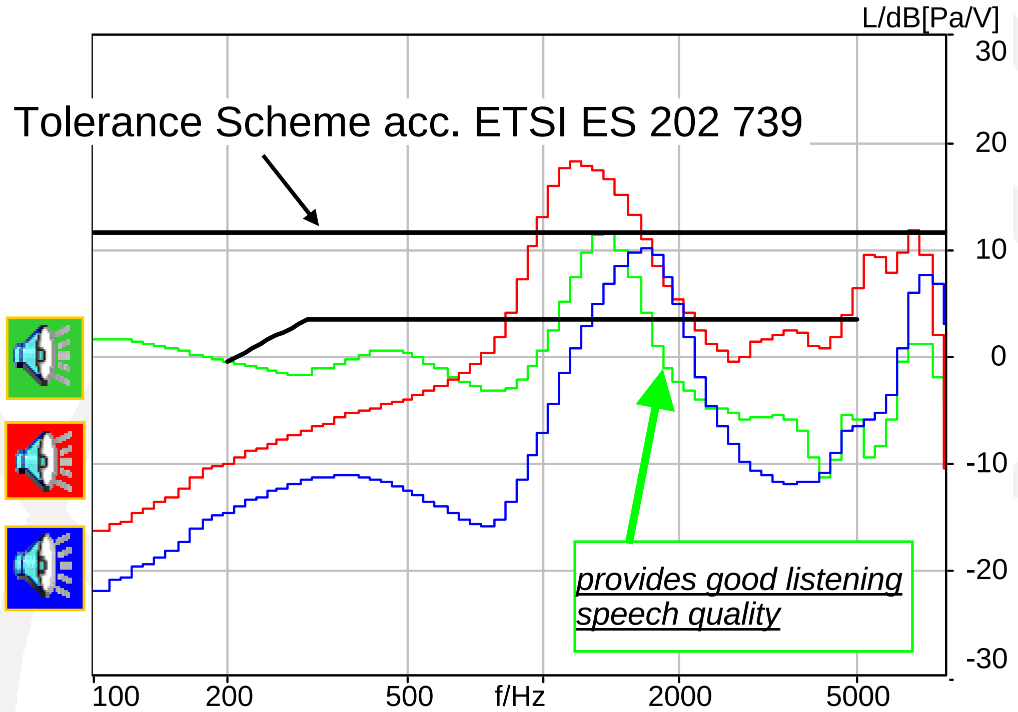
Deutsche Telekom Laboratories (Germany)

Deutsche Telekom Laboratories

HEAD acoustics

PART 1: Receiving Frequency Response Characteristics

Receiving frequency response of 3 wideband phones (handset mode, 3.4 ear, 8N, free-field)

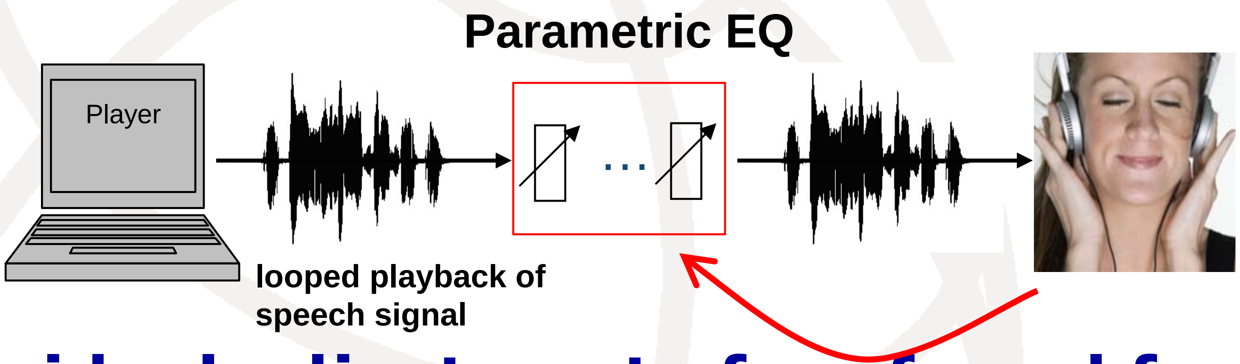


Test Conditions

- Use 3.4 or 3.3 artificial ear
- Use 8N application force
- Use artificial voice or composite source test signal
- ***NEW: use DRP to FF correction***

-> Is the tolerance scheme really desirable?

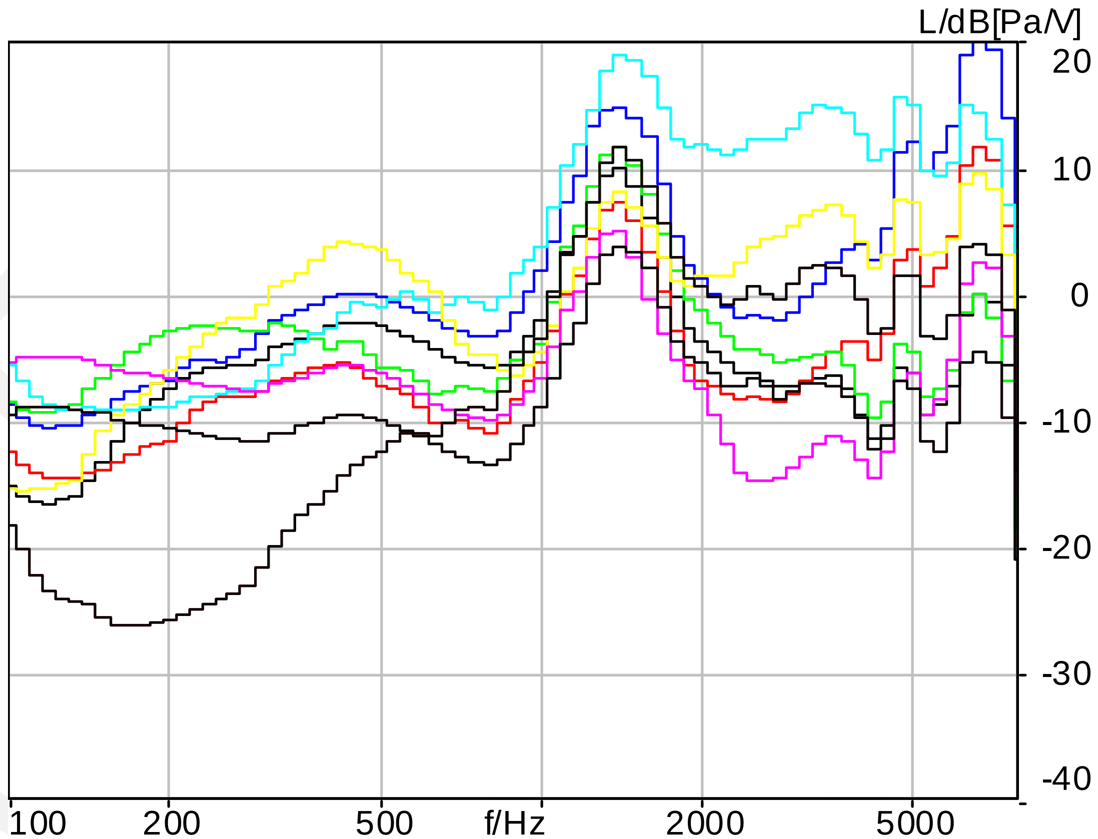
New Subjective Experiments: Pretests with Experts Listeners



- individual adjustment of preferred frequency response characteristics
- conducted with different types of phones

* published DAGA 2007, March 07 and ETSI STQ, April 08

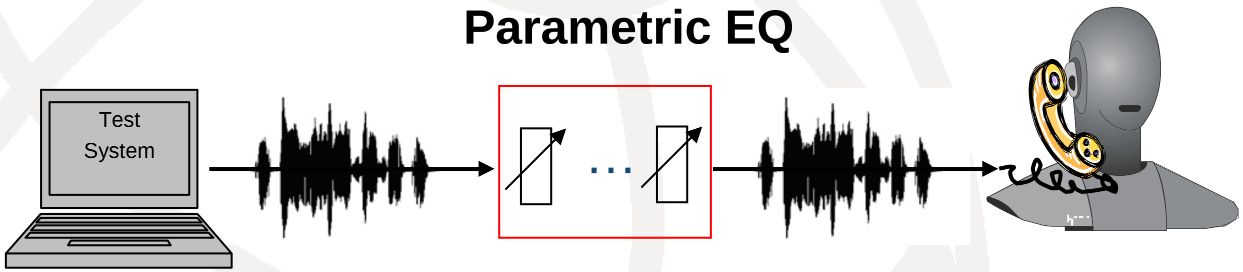
The Results - Some Examples



-> wide range of personal preferences

Step 2: Rank Ordering of the Response Characteristics

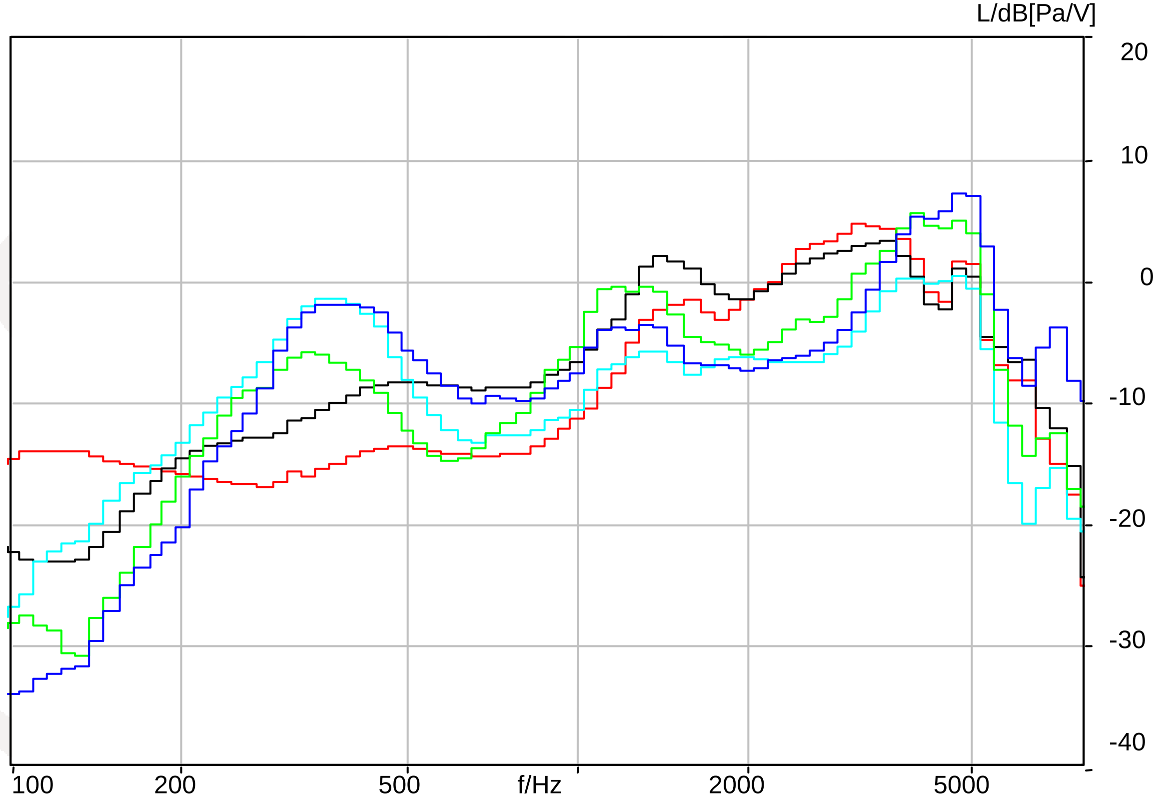
- Phones with different EQ-Settings applied at HATS



- Ranking of the different EQ-Settings by experts:
- Results:
  - clear “winner” setting for most phones
  - “favorite” frequency responses similar for all phones

Example Result

- “optimum” response characteristics for 3 phones



Deutsche Telekom Laboratories

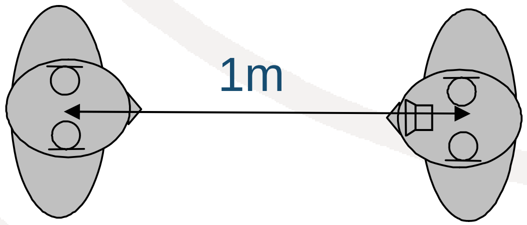
HEAD acoustics

Step 3: Listening Only Test

Test Conditions

4 speakers: 2 female, 2 male

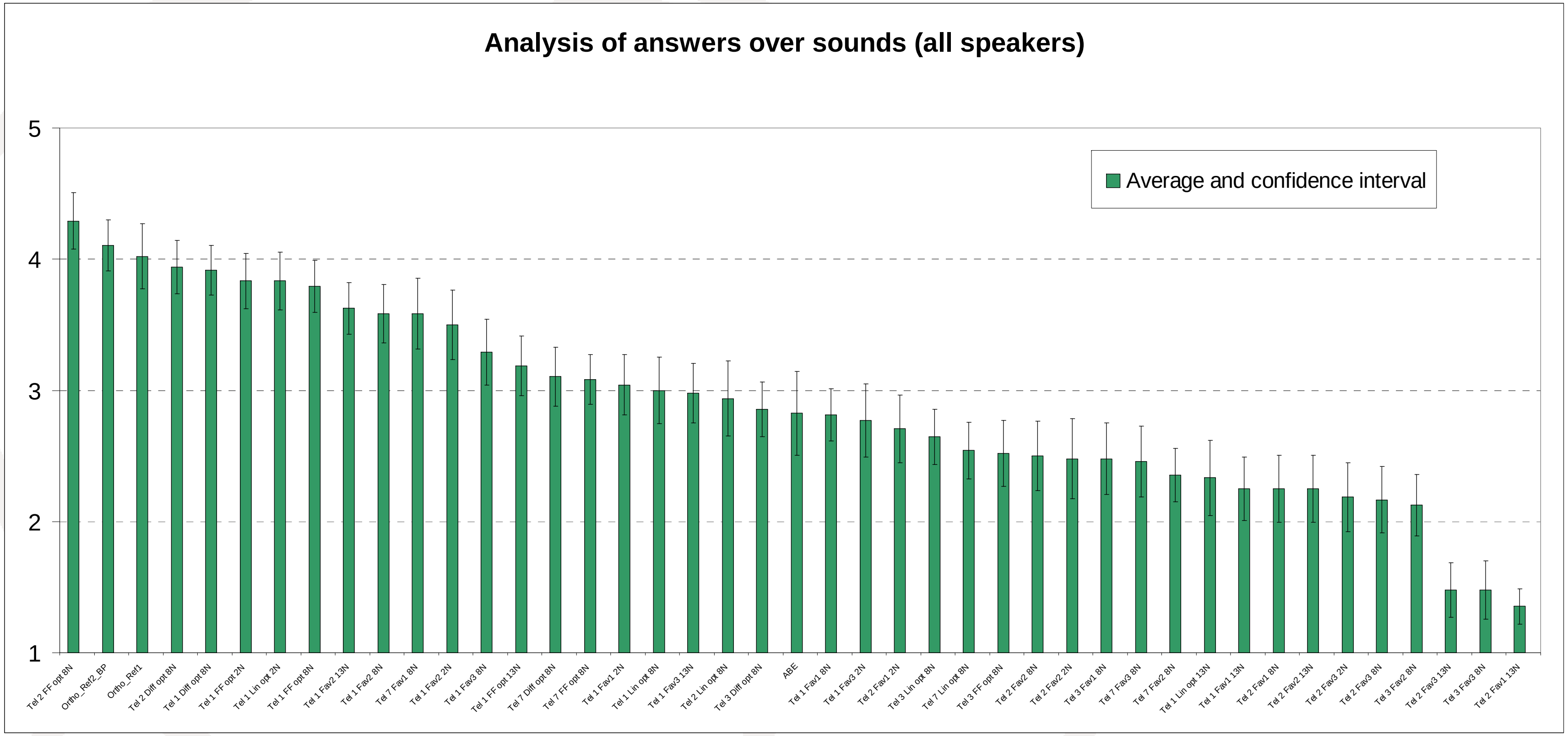
- 4 wideband phones, each combined with several frequency responses
- 1 artificial bandwidth extension
- 1 ortho-telephonic reference



Representation & Assessment

- 43 conditions x 4 speakers, each assessed by 12 naïve test persons
- Assessment of “speech sound” on a five point MOS-scale:
  - 5 – excellent
  - 4 – good
  - 3 – fair
  - 2 – poor
  - 1 – bad

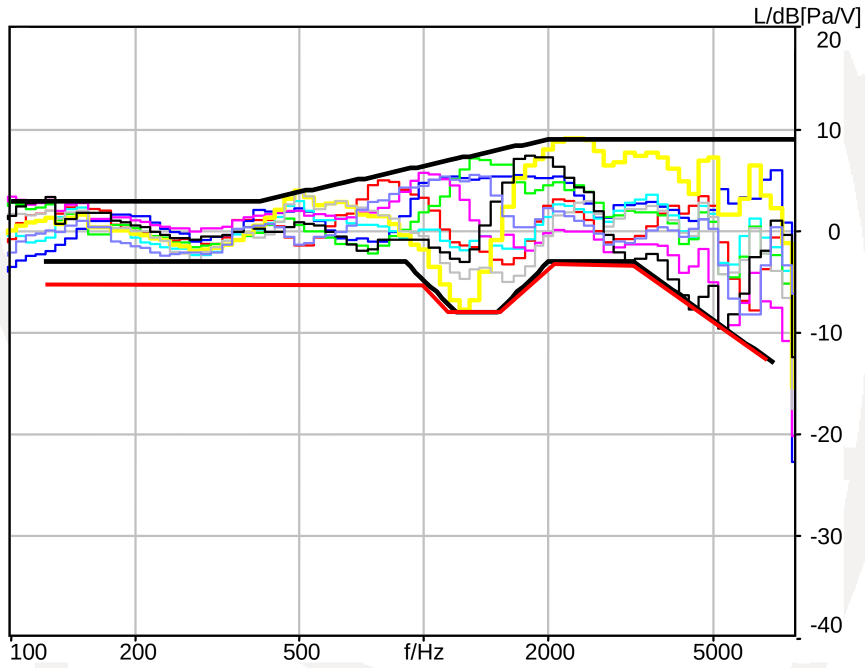
Results of LOT



- whole quality range is covered
- “winners”: ortho-telephonic reference, flat frequency response with DF or FF EQ
- -> extract for new tolerance scheme

Conclusion - New Tolerance Scheme for Receiving

Possible new tolerance scheme with diffuse field reference:

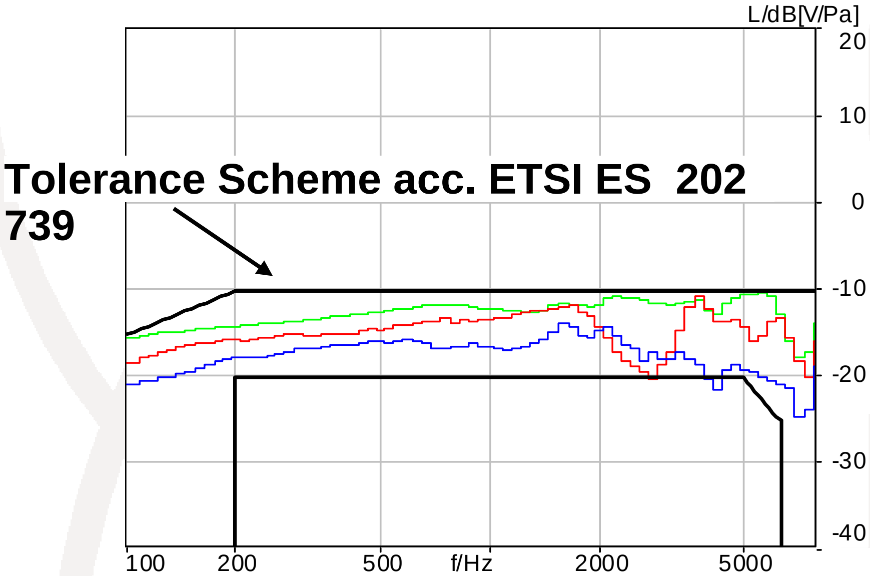


- **9 frequency responses with MOS-LQS ≥ 3.6 (average over all speakers)**
- **black lines: "most tight" tolerance possible**
- **red line: possible "smoothed" version of lower tolerance line**
- ***Results discussed in ETSI STQ and ETSI NG DECT, to be discussed in ITU***

PART 2: Frequency Response in Sending

Sending frequency response of 3 wideband phones

(handset mode, 3.4 ear, 8N)



- Tolerance scheme given by ETSI ES 202 739 is mostly meet for all
- All phones provide good listening speech quality ($TMOS_{wb}$ = 4.2)

Problem:

- Today's and future phones are more and more used in noisy environments
- Transmission of wideband noise
- People start talking with "Lombard Effect"

Desirable Frequency Response Char. in Noisy Environments

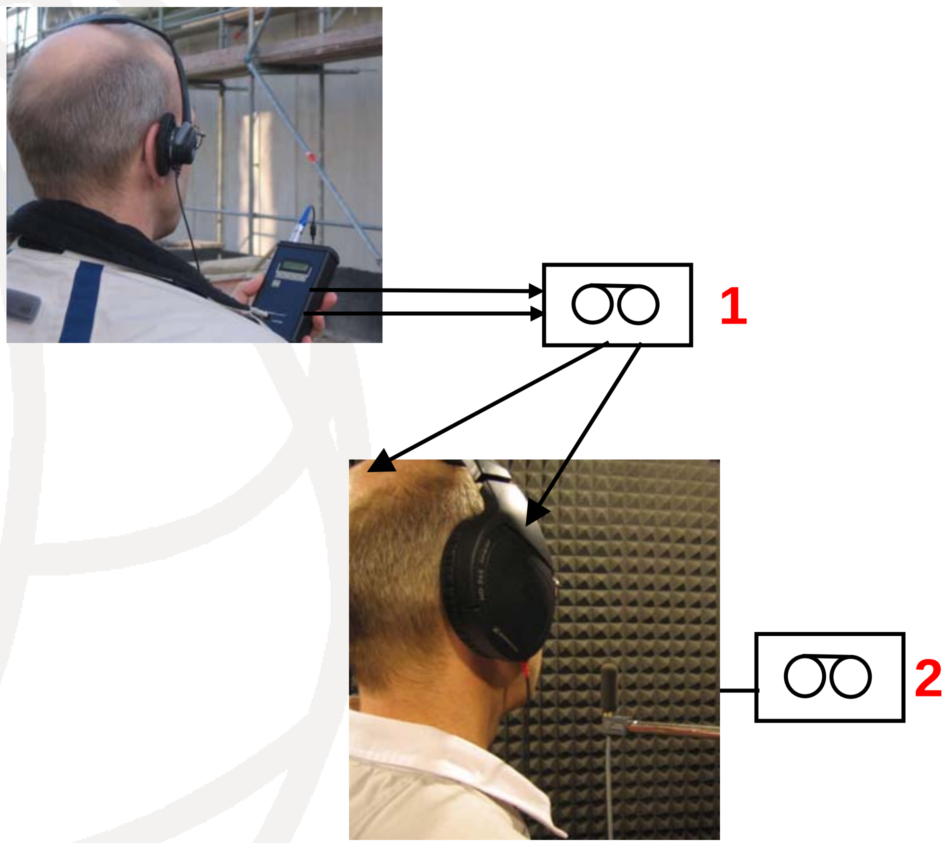
Setup of the subjective experiment (Step 1 & 2):

- Binaural recording of different realistic background noises (here: car, pub, café, living room, call centre) → **STEP 1**

- Speaker in anechoic conditions, noise playback via closed headphone

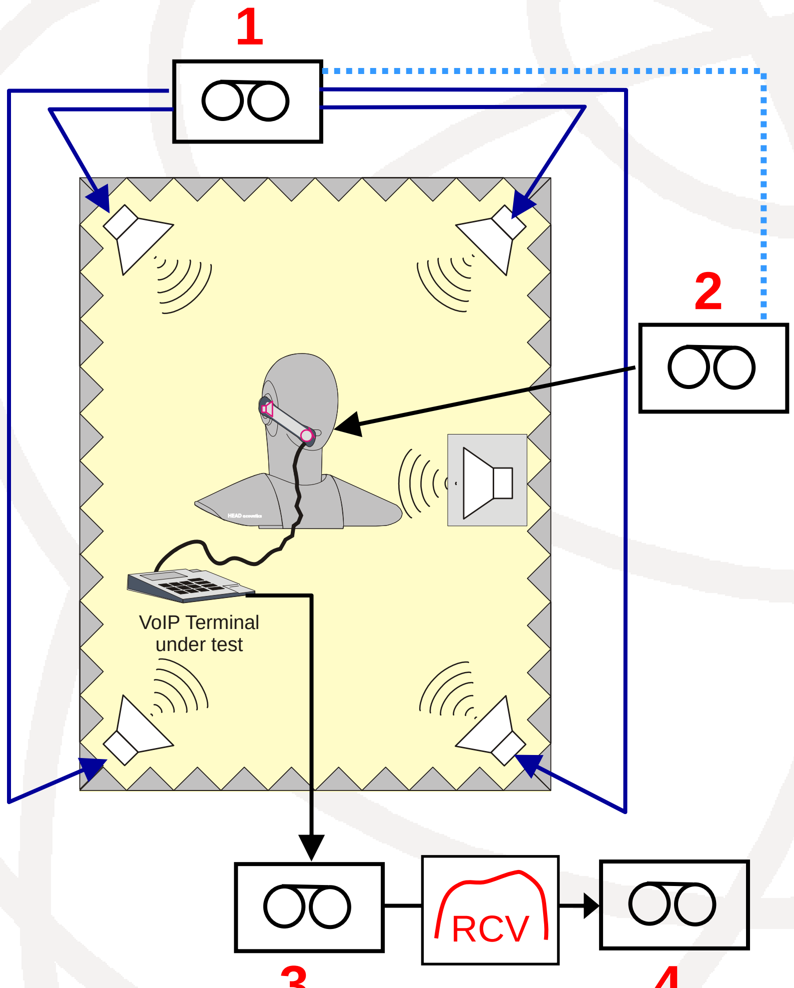
  → initiation of Lombard effect

  - recording with omni-directional microphone
  - recording synchronously to the background noise → **STEP 2**
  - (recording of Lombard speech for one female and one male speaker, each for 5 background noises)



Desirable Frequency Response Char. in Noisy Environments

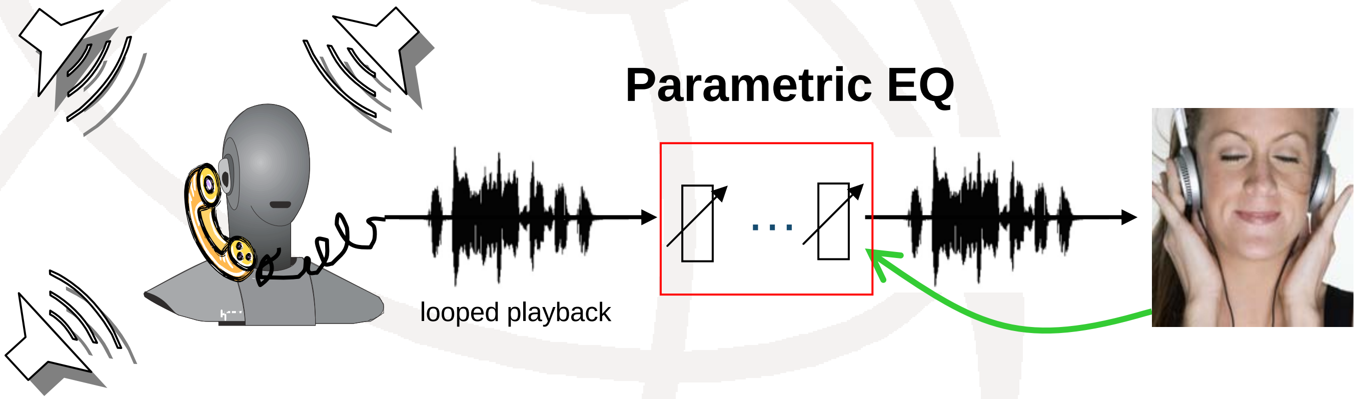
Step 3: Recording of Speech plus Background Noise



- Synchronous playback of corresponding Lombard speech (2) via HATS with background noise
- Transmission of Lombard speech and noise via 3 wideband phones (G.722) (3)
- Use "optimum" receiving frequency response to simulate receiving side (4)

Pretests with Experts Listeners

- **Equalizer adjustment by 7 expert listeners for 3 wideband phones with 5 background noises each**

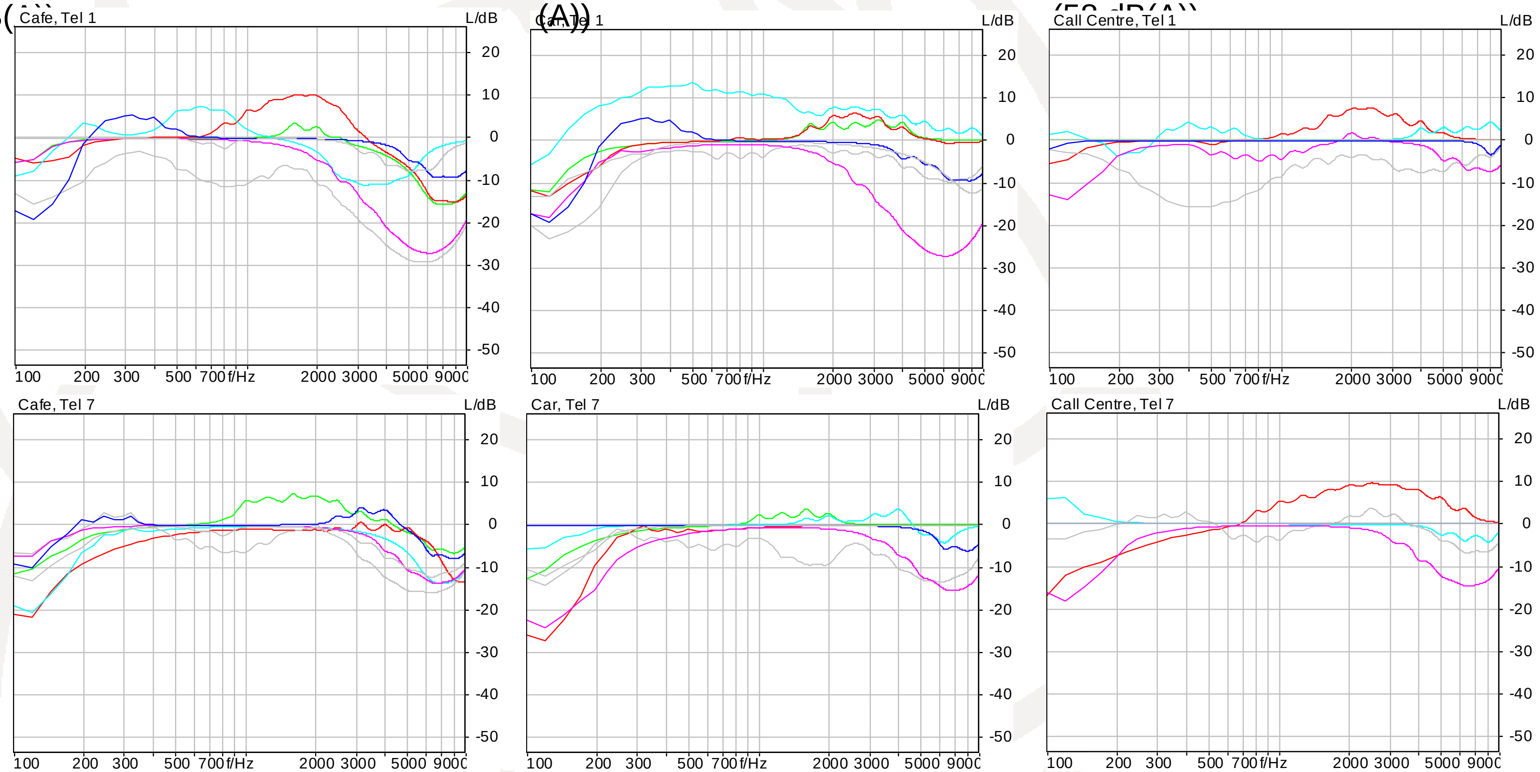


- **Adjustment of preferred frequency response to achieve best overall sound quality (speech and noise)**

Results of Pretests

Expert‘s equalizer settings for …

non-stationary cafeteria noise (68 dB(A)) | stationary car noise (69 dB (A)) | non-stationary call centre noise



phone 1

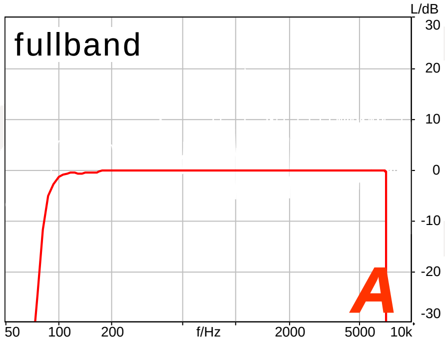
phone 7

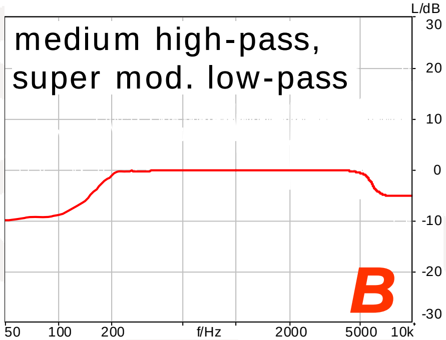
Results and Assumptions

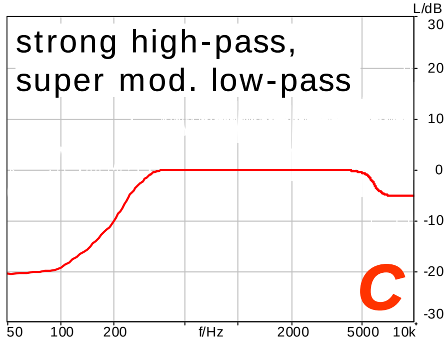
- **Results:**
  equalizer settings ...
  - show similar characteristics for all experts for the different noise-telephone combinations
  - are noise level dependent: high and low passes are inserted
  - vary only slightly for one type of noise in combination with different phones

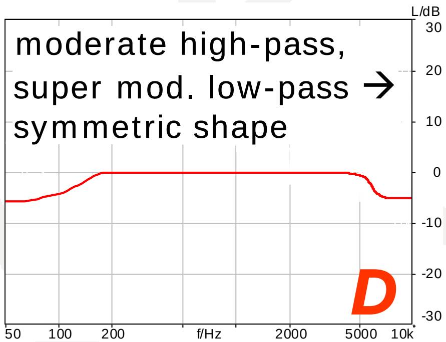
- **Assumptions:**
  - Depending on the noise level more or less strong high and low passes are inserted in order to limit the annoyance due to the background noise – although also the speech sound is moving towards narrowband speech!
  - For further improvement of the speech intelligibility an emphasis around 2...3 kHz is applied.

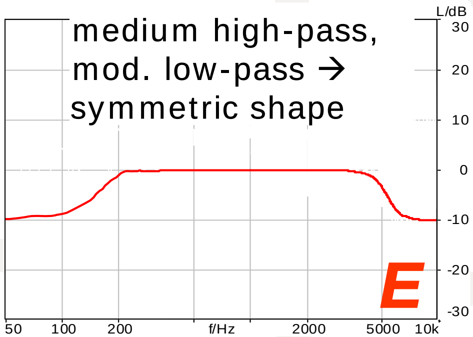
Listening Test: Filters Applied

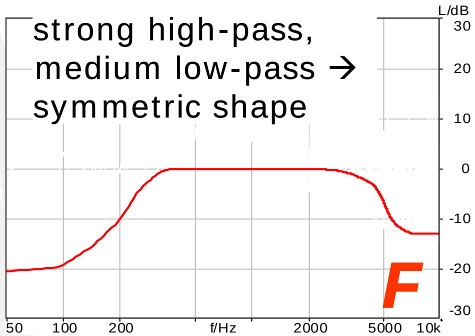




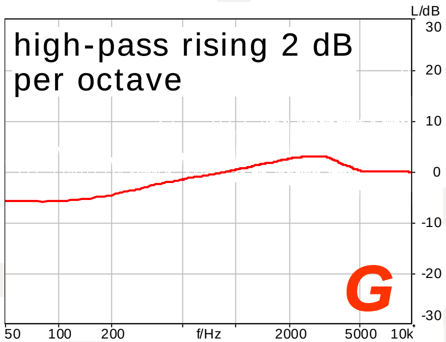


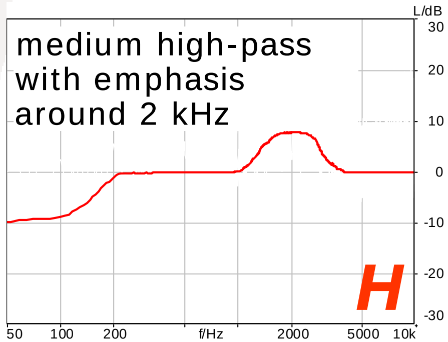






Depending on the results of the expert's pretest, the following 8 filters were created and used for the listening tests





Listening Test 1

Test conditions – LOT 1

- speech transmitted via 1 wideband phone (flattest sending frequency response, “traditional big” handset, no noise reduction)
- 5 background noises
- filtered with 8 filters extracted from expert pre-test
- all filtered with optimum frequency response of receiving frequency response test
- Due to Lombard effect up to 15 dB level difference between listening samples -> loudness adjustment to provide “same” loudness for all samples

Presentation & Assessment

- 2 speakers: 1 female, 1 male

→36 conditions x 2 speakers, each assessed by 16 naïve test persons

- Diotic representation with open, free-field equalized headphones
- Assessment of “overall quality” on a five point MOS-scale:

  5 – excellent

  4 – good

  3 – fair

  2 – poor

  1 – bad

Results of LOT 1

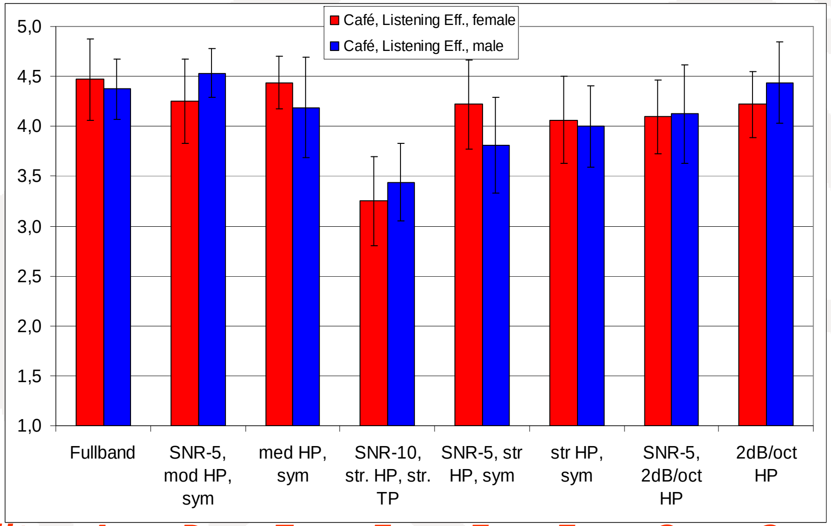
- Findings:
  - The higher the background noise level is the stronger the Lombard effect is. -> Lower MOS scores
  - Some results for males better than for females (females tend to sound shrill and sharp for higher background noises in conjunction with the Lombard effect)
  - No clear preference for one response characteristics, **only tendencies**
  - Full bandwidth is preferred for most noises
  - Strong high-pass characteristics is not preferred

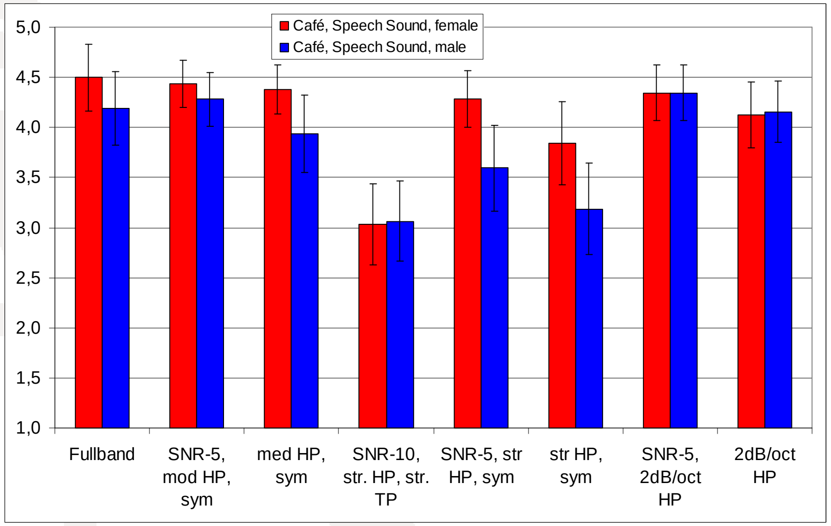
Listening Test 2

Test conditions – LOT 2

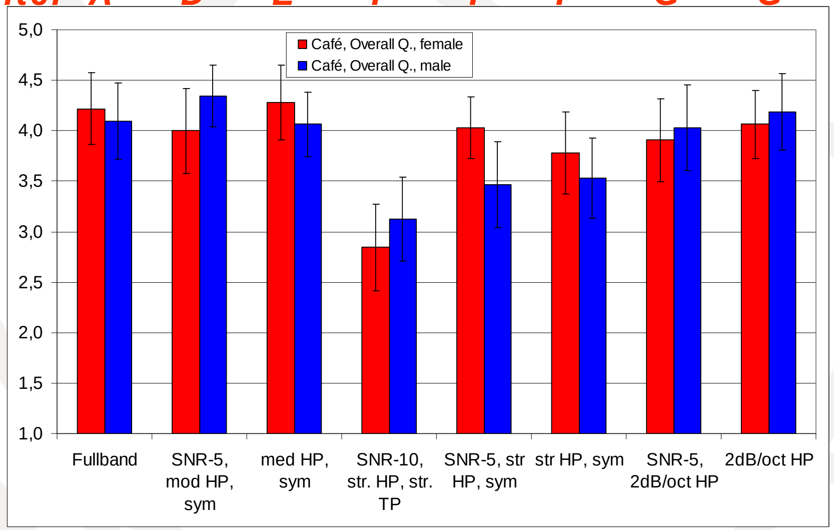
- Separate assessment for 3 noises (café, car, living room)
- Reduced number of filters per noise
- Additional "simulated" conditions with 5/10 dB worse SNR for each noise and filter
- Listening level adjusted to 73 dB SPL active speech level
- Diotic representation with free-field equalized head phones
- 3-fold assessment by naïve persons (similar to P.835):
  - listening effort (complete relaxation possible, no effort required – attention necessary, no appreciable effort required – moderate effort required – considerable effort required – no meaning understood with any feasible effort)
  - speech sound quality (excellent – good – fair – poor – bad)
  - overall quality (excellent – good – fair – poor – bad)
- Assessment in steps of 0.5 MOS

Results LOT 2 (café noise).





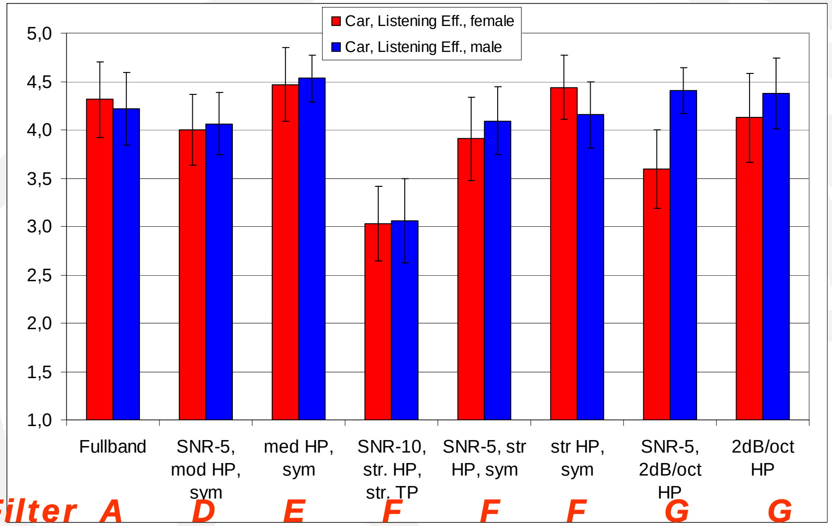
Filter A D E F F F G G

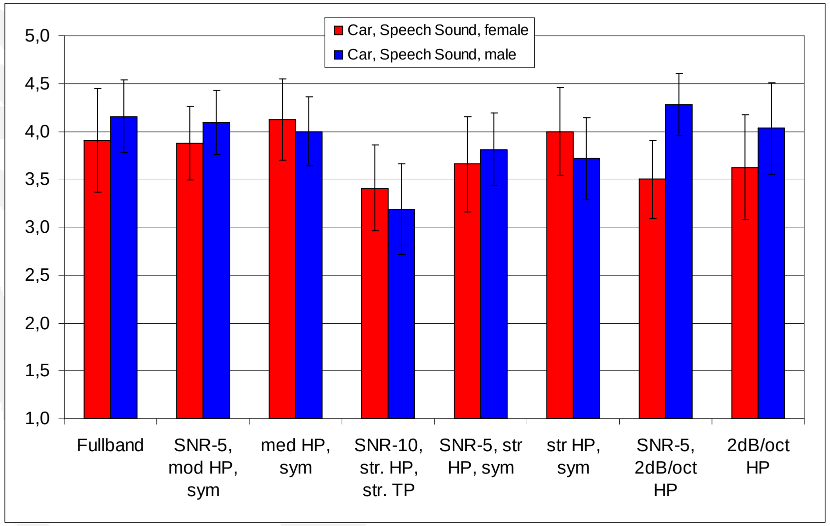


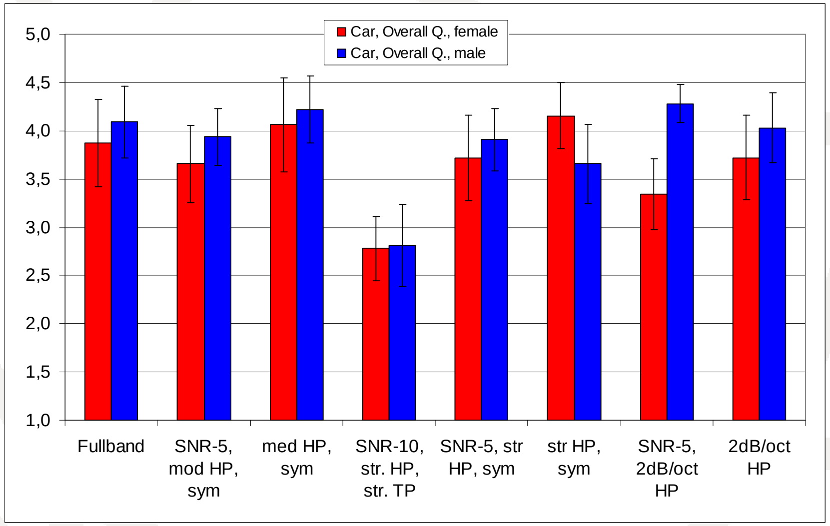
Female / male voice:

- For all parameters and both SNRs: only slightly different MOS scores for full-band, moderately band-pass and 2 dB / oct. filtered versions
- strong band-pass filtered version leads to significantly lower MOS scores for the male voice than for the female voice. Since this background noise has no dominant low frequency components, the limited bandwidth impairs the speech quality instead of reducing the annoyance due to the background noise. Furthermore the MOS score of the example with the lower SNR is about 0.5 MOS higher.

Results of LOT 2 (car noise)







Male voice:

- no significant difference for all parameters between full-band and filtered versions for original and 5 dB reduced SNR;
- moderately band-pass and 2 dB/oct. high-pass filtered versions tend to be slightly better than full-band version

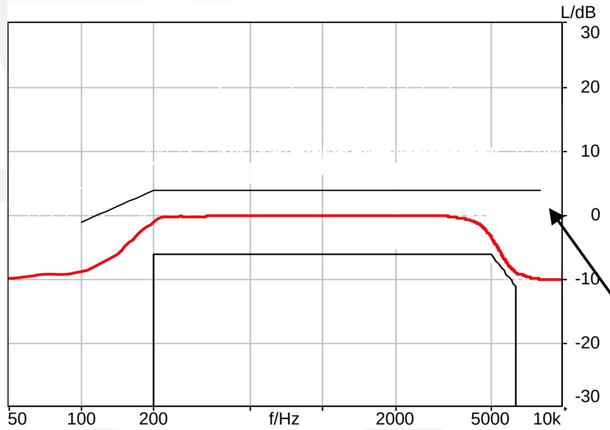
Female voice:

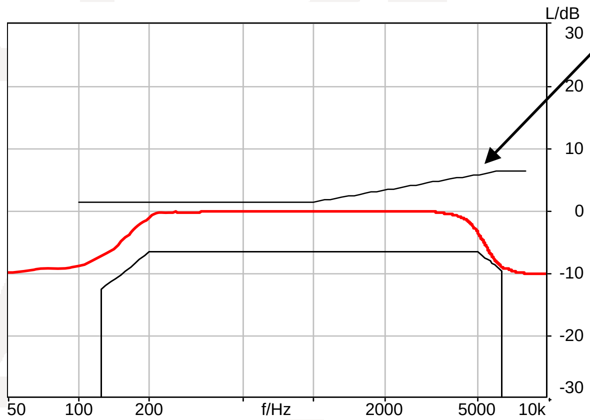
- moderately and strong band-pass filtered versions result in slightly higher MOS scores than full-band

Sending Frequency Response in Noisy Environments - Final Conclusions

Proposal for frequency response shape in sending direction under noisy conditions

- **"acoustical noise reduction" by applying a moderate or – depending on the noise – even medium band-pass filter may be helpful:**
  - **acoustical noise reduction with only slightly affect the speech sound**
  - **noise reduction algorithms may be adjusted less "aggressively"**
  - **Filter should be applied adaptively - only if a background noise is detected at the terminal's microphone**
- **Investigation should be repeated for different wideband noise reduction algorithms**





Recommended filter still matches ES 202 739 and 740 tolerance schemes!