Radiation Directivity of Human and Artificial Speech

Teemu Halkosaari

Supervisor:

Matti Karjalainen

Instructors:

Markus Vaalgamaa

Juha Backman

Agenda

Part I:	WHY?	1.	Motivation and Background
		2.	Goals
Part II:	The STUDY	3.	Measurements
		4.	Analyzes
		5.	Modelling
		6.	Results
Part III:	The OUTCOME	7.	Improvements to telephonometry
		8.	Conclusions

Part I
WHY?

1. Motivation and Background

- HATS was originally designed for narrowband speech and big handsets.
- Phone size has become smaller and smaller and headsets are used widely.
- Wideband phones are coming..
- Still
  - HATS and other artificial mouths will be used in future, although
  - we do not know exactly the directivity of HATS or
  - human mouth especially on near field locations
- Furthermore does the speech content affect the directivity?
- There is no literature on the subject
- Standardization is inadequate
  - ITU-T P.58 Head and torso simulator for telephonometry
  - ITU-T P.51 Arficial mouth
  - narrow band: 300-3400Hz, wide band: 150-7000Hz

2. Goals

- Measure frequency responses for the mouth of HATS to positions which are used by the microphones in handsets and headsets.
- Repeat the same measurements for a group of test subjects
- Most important part of the study is to **analyze the difference between mouth of HATS and an average person**.

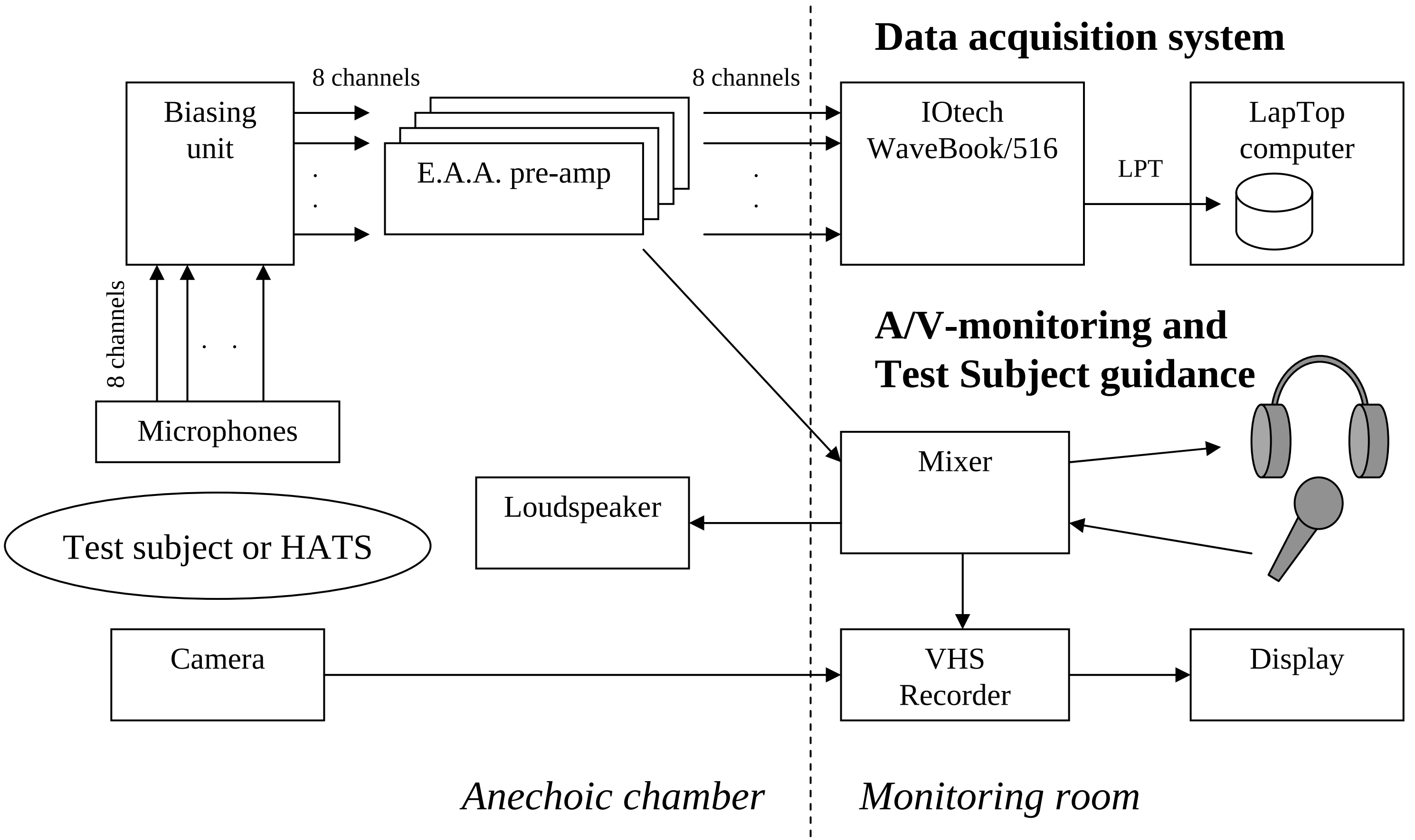
- In practice a set of transfer functions is acquired.
- In addition some quidelines if HATS or the handset/headset measurement processes should be improved.

Part II

The STUDY

3. Measurements: Setup (1)

- Parallel 8 channel audio recording at 32kHz sample rate



3. Measurements: Setup (2)

- Small anechoic chamber at HUT

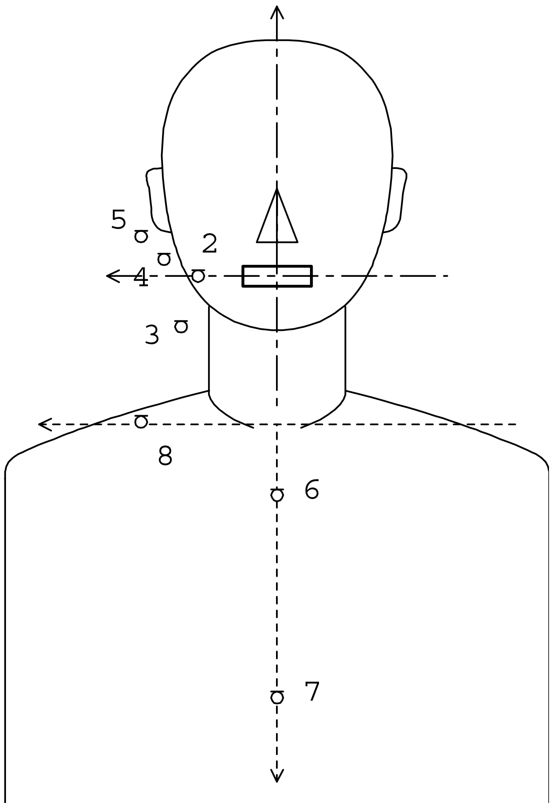
3. Measurements: Subjects

- HATS: B&K 4128
  - Compensated white noise as excitation signal (flat in MRP).
- Test subjects: 8 male, 5 female
  - Speech material:
  - ”Kaksi vuotta sitten kävimme ravintola Gabrielissa Helsingissä ja söimme siellä padallisen fasaania banaanilla höystettynä” (includes all phonemes)
  - Separate phonemes \n m η s r\ and \a e i o u y ä ö\
  - Different speech volumes: loud, normal, and silent
- In addition supplementary MLS-measurements for HATS in far field.

3. Measurements: Probes

- 8 Sennheiser KE-4-211-2 electret microphones
- Phones: 5510, 8310, 9110 and headsets: Stereo Headset HDS-3, Headset HDC-5, and Boom Headset HDB-4
- Microphone positions
  - MRP = 25mm in front of mouth
  - measurement positions:

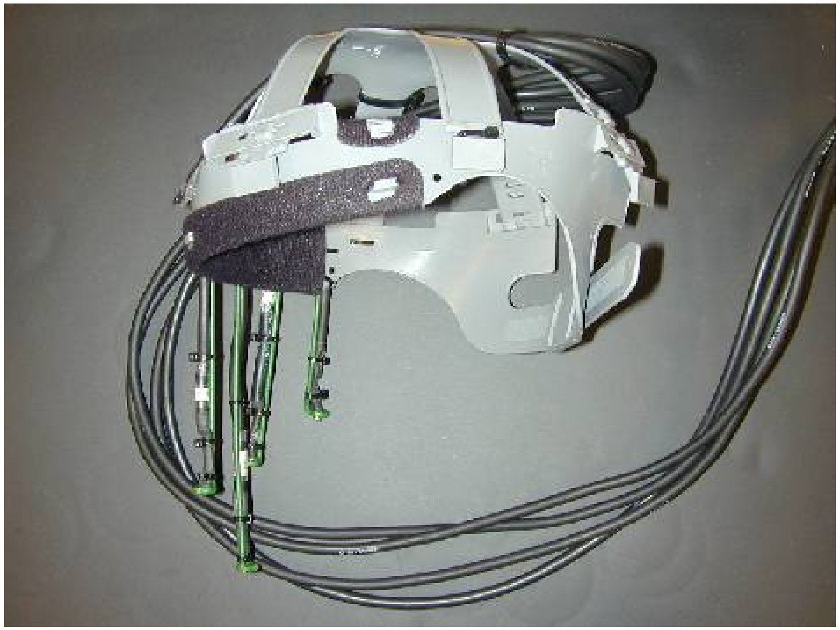
Channel no.	On-axis (mm)	Off-axis (mm)	Corresponds
1	500	0	Reference position
2	0	H 60	Large handset
3	-10	H 70, V -40	Large handset
4	-30	H 85, V 10	Small handset
5	-70	H 100, V 30	Boom headset
on chest:			
6	50	0	Headset
7	200	0	Headset
8	0	100	Headset



NOKIA

3. Measurements: Positioners

- Helmet and chest positioner were developed to keep microphones in right positions.

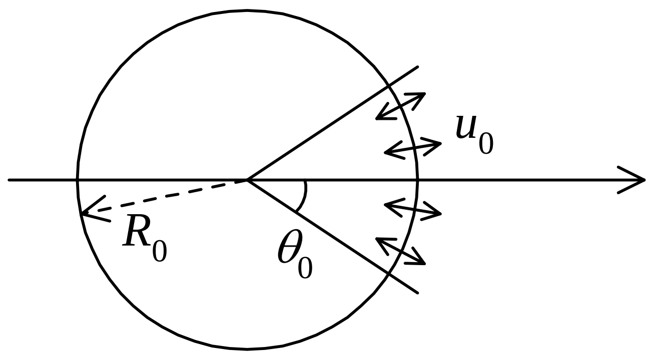


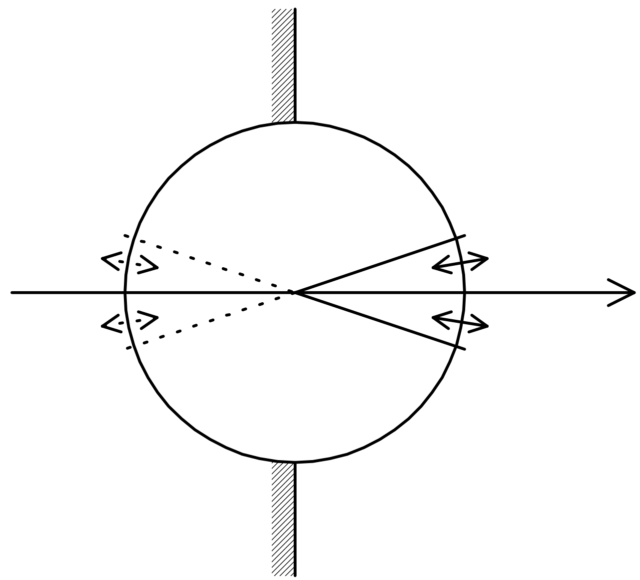
4. Analyzes

- Starting point: parallel 8-channel 32kHz 16bit raw audio data.
- Phonemes and active speech were segmented manually (CoolEdit Pro).
- All analyzes were made in Matlab environment:
  - SNR and Coherences to ensure reliability
  - Power spectra estimates
  - **Transfer function estimates**
    - Reference point in far field (0.5m in front) and MRP
  - Confidential intervals consideration for transfer functions
- FFT 1024 samples, Hanning-window, 50% overlap
- Averaging and Smoothed to 1/3-octave bands
  - Speech was averaged weighting with coherence

5. Modelling

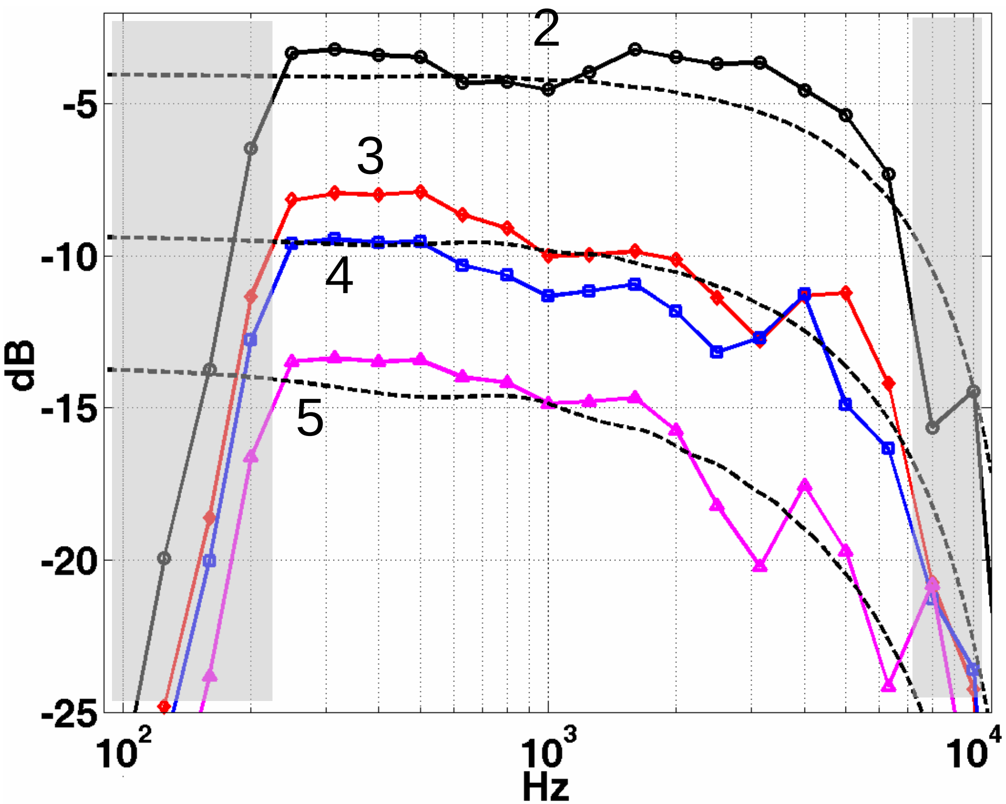
- Results were validated by modelling: Sphere and piston source
  - Head = sphere
  - Mouth = radially vibrating round surface of the sphere (piston).
- An infinite baffle can be added to simulate the upper body.
  - Mirror source method
- Numerical implementation in Matlab

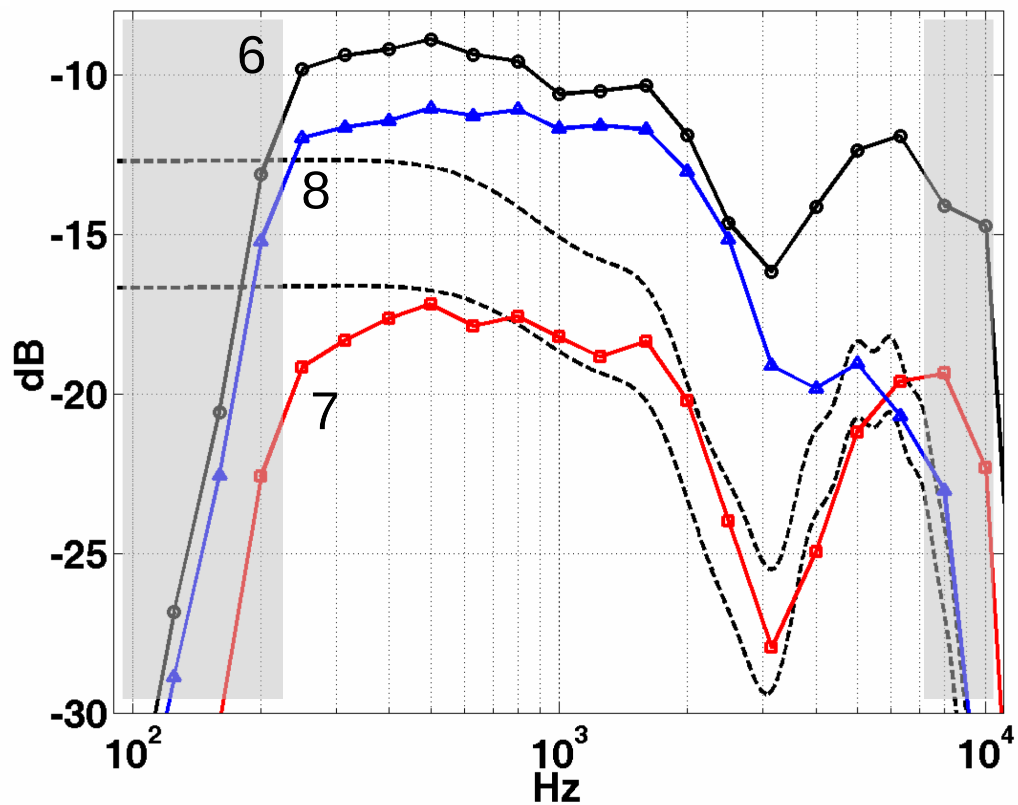




6. Results: Nearfield

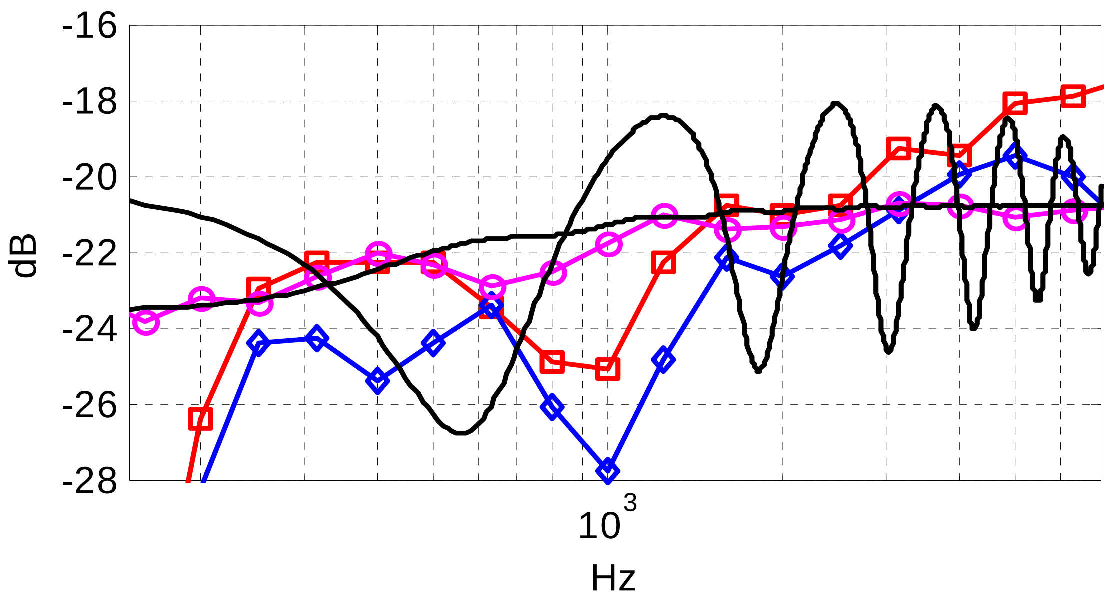
- Transfer functions for pos 2-8 referred to MRP (calculated using power-spectra)
- Models for pos. 2, 4, and 5 (left) and baffled models for pos. 6 and 7 (right)





6. Results: Far field

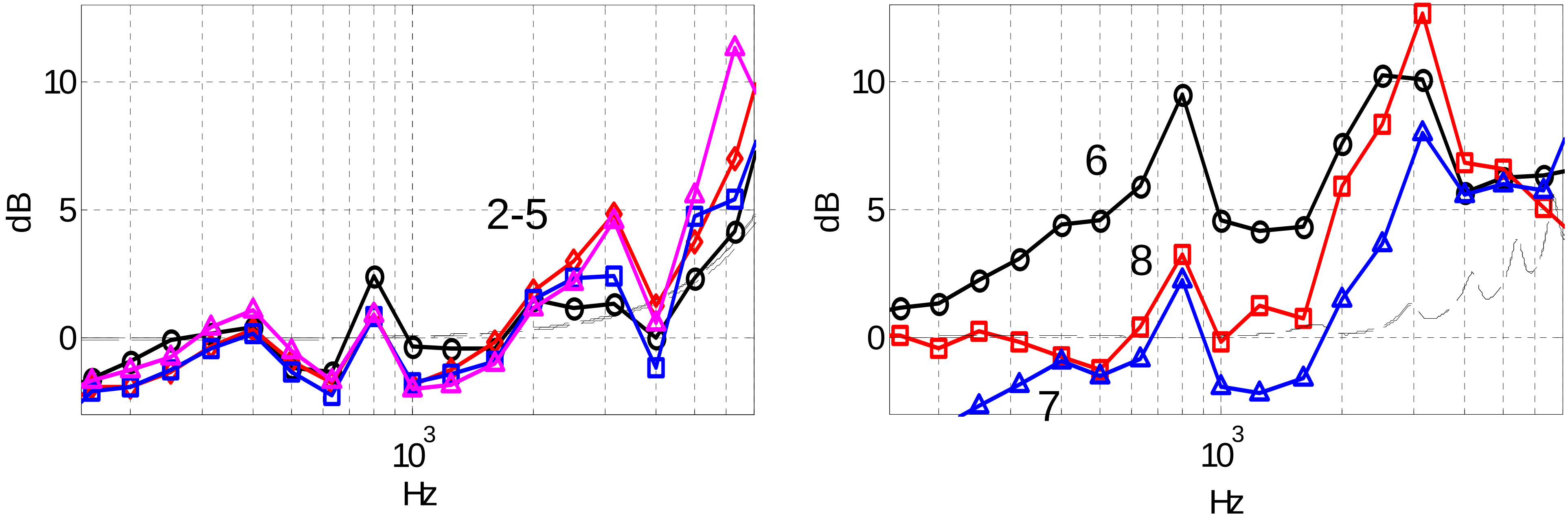
- Transfer function from mic. pos. 1 (0.5m) to MRP.
  - Normal situation (red).
  - HATS without vest (blue).
  - Only the head (magenta).
- Head-piston model (dashed black) and model with baffle (dotted black).



NOKIA

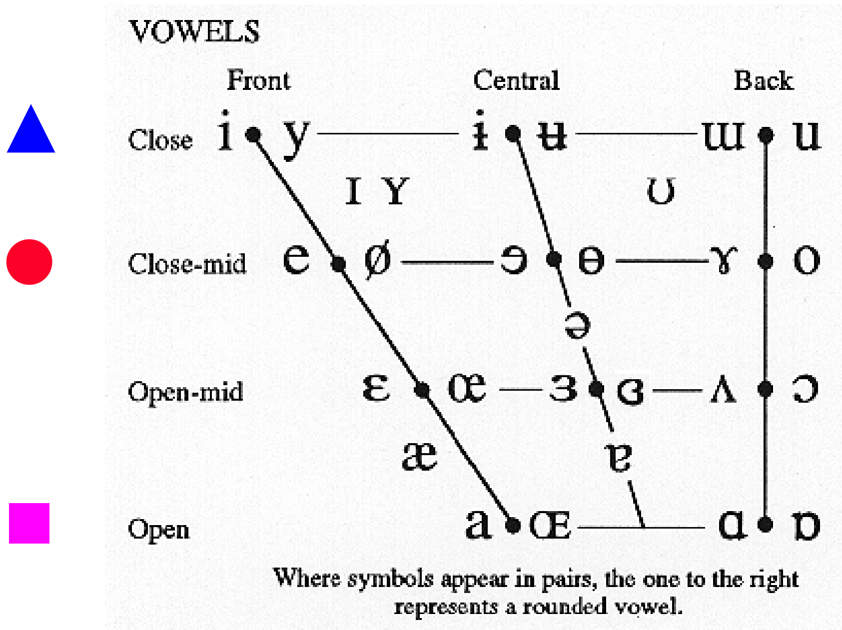
6. Results: Human - HATS

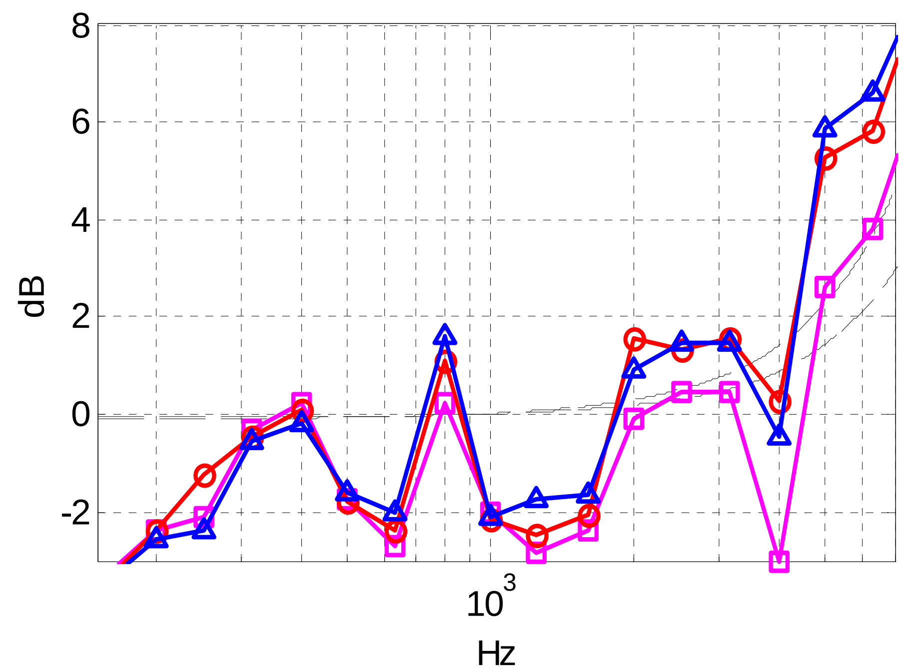
- Comparison between transfer functions, referred to 0.5m position
  - TF_{Human} / TF_{HATS}
- 1/3 mouth cross-section area ratio in models.



6. Results: Speech content

- Speech content changes the directivity pattern
  - Speech style: loud – silent etc..
  - Articulation: continuous phonation – natural speech
  - Phonemes (Human/HATS difference for mic. pos. 4 below)
- Mouth aperture size = speepness of the response on high frequencies
- Statistical approach: ANOVA model did not apply = speech data scattered





6. Results: Reliability

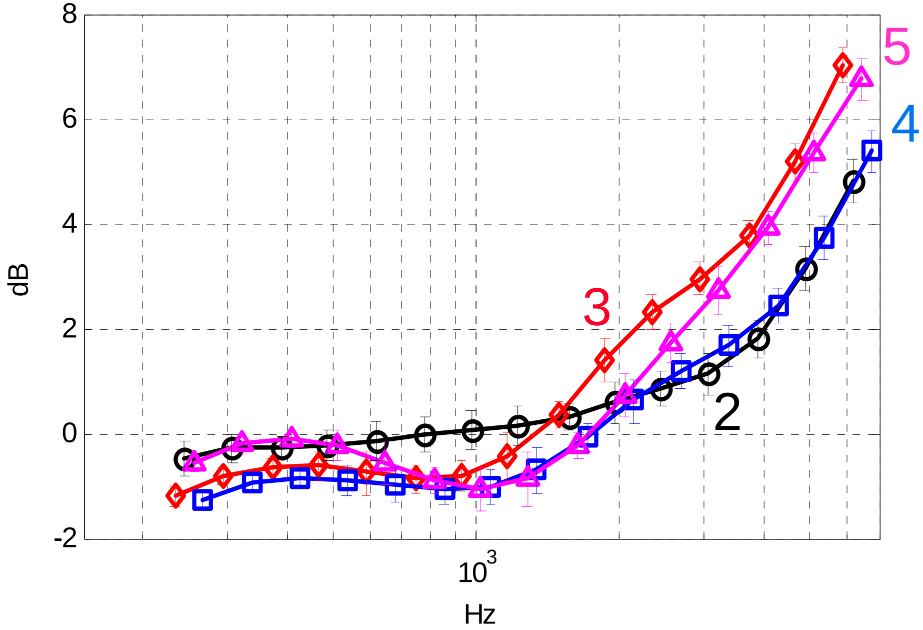
- Beforehand by careful desing of the measurement processes
- SNR and Coherence (estimates)
  - SNR on wide-band: >25dB
  - Coherence on narrow-band: >0.8 wide-band: >0.65
- Normal distribution hypothesis
  - 95% confidential intervals on wide-band: <±0.5dB
- HATS measurements are omitted

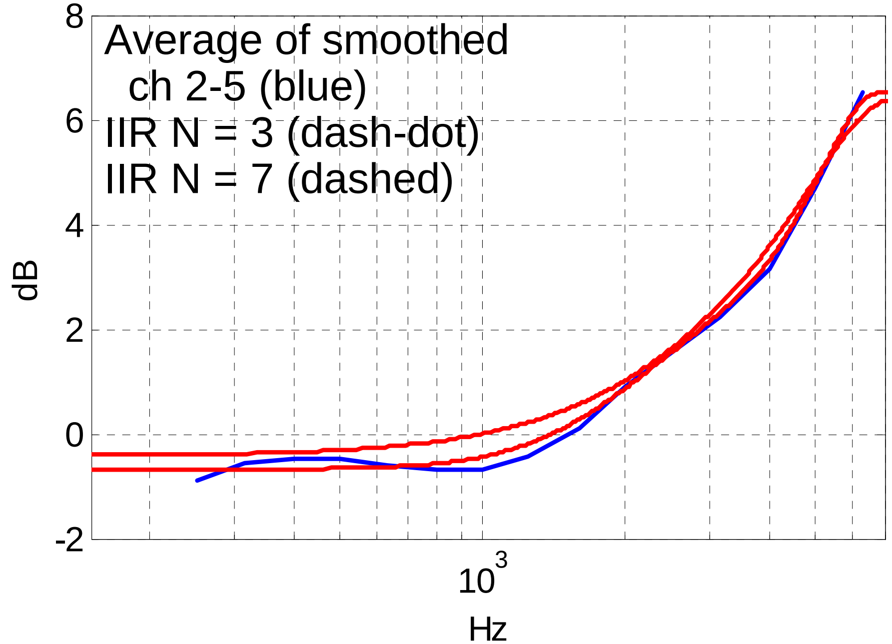
Part III

The OUTCOME

7. IP: Equalization

- HATS does not correspond averaged test subject
- Difference curves are ”well-behaving” -> A low order filter could equalize the difference near cheek.
- Example: Yulewalk recursive IIR filter design (least-squares method) was applied in Matlab for smoothed and averaged differences.

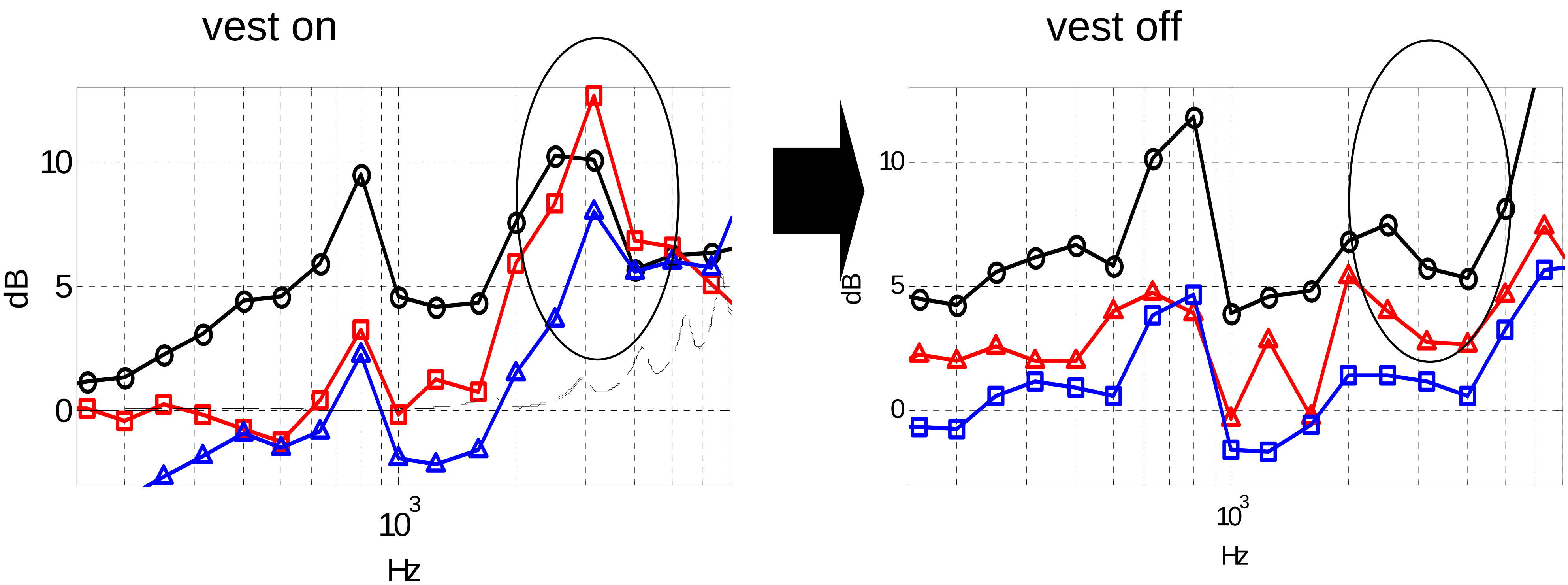




NOKIA

7. IP: Vest

- About 2cm thick vest should be used in HATS measurements. How does it change the directivity in position on the chest?
- Let’s take off the vest and compare to average test subject (figure on right)
- The near-chest reflection shifts to higher frequency



NOKIA

8. Conclusions

- Key results:
  - On high frequencies the HATS is too directional.
  - Near cheek the difference (human/HATS) is substantially índependent of the position.
  - Near chest differences in chest reflections are significant.
  - Speech content affects on the directivity.
- Reasons:
  - Mouth cross-section area during speech is smaller than in HATS. Speech content and mouth cross-section area are linked.
  - The design of the HATS does not correspond the acoustical characteristics of real body.
- What could be improved?
  - Near cheek: simple EQ was proposed.
  - Vest usage should be considered always carefully.
  - Improvement of the structure of HATS (mouth size and torso design). This path requires more measurements and analysis.

Agenda

Part I:	WHY?	1. Motivation and Background
		2. Goals
Part II:	The STUDY	3. Measurements
		4. Analyzes
		5. Modelling
		6. Results
Part III:	The OUTCOME	7. Improvements to telephonometry
		8. Conclusions