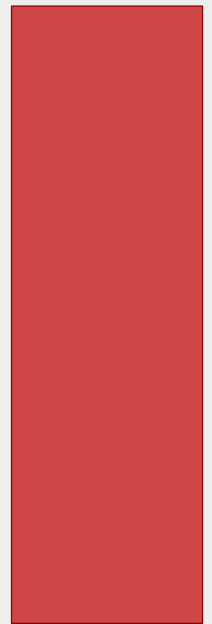


Variable Voice Likability Affecting Subjective Speech Quality Assessments

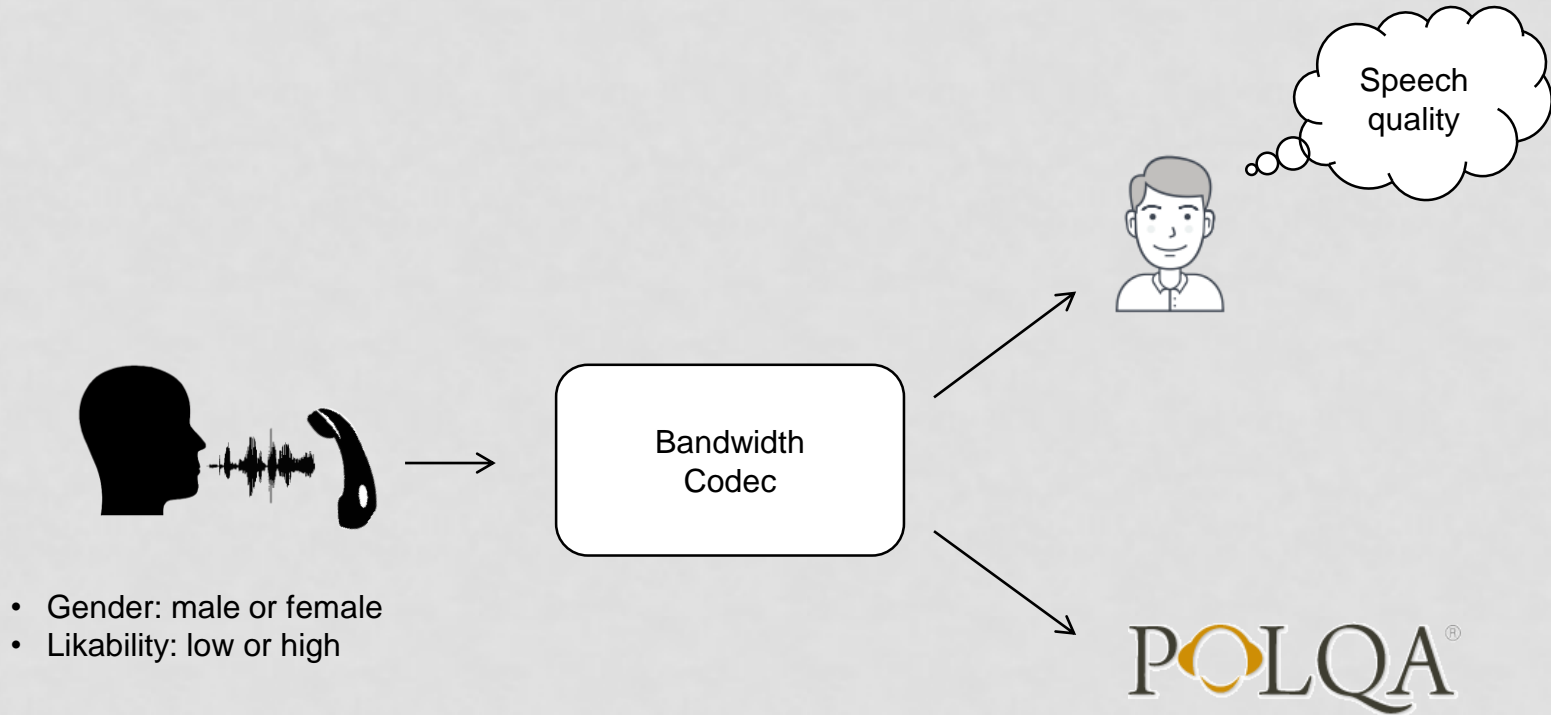
*Laura Fernández Gallardo, Ph.D.¹, Gabriel Mittag¹,
Prof. Dr.-Ing. Sebastian Möller¹, Dr. John Beerends²*

¹Quality and Usability Lab, Technische Universität Berlin, Germany

²TNO, The Hague, The Netherlands




Motivation



Motivation

- Recommended speech material:
 - For auditory tests (ITU-T Rec. P.800):
 - And also adopted by the POLQA model (ITU-T Rec. P.863)
“Two sentences from both, male and female speakers”
- Gender must be taken into account
- But what about other speaker peculiarities?
 - Age, language, emotions, personality, speaker/voice likability..
 - Under examination (ITU-T Study Group 12)

Outline

- Speech material 
 - Speakers
 - Channel degradations
- Speech quality assessments
 - Subjective
 - POLQA
- Effects of speakers' "warmth-attractiveness"
 - On subjective MOS
 - On POLQA MOS
- Conclusions

Speech stimuli selection

Corpus	Nautilus Speaker Characterization (NSC)
Number of speakers	300 speakers; 126 males, 174 females
Labels	34 interpersonal speaker attributions 34 voice descriptions for selected speakers
ISLRN	157-037-166-491-1
Author / Data owner	Laura Fernández Gallardo



Speech from 6 males and 6 females with extreme **warmth – attractiveness (WAAT)**

Speech stimuli selection

- 12 speakers (male, female, high WAAT, low WAAT)
- 8 excerpts, scripted sentences
- 8 channel conditions: reference and 7 degradations

↓ balanced
design


The same speaker, the same content, and the same distortion appears the same number of times.

Total: $12 \times 8 = 96$ stimulus files (mean 8.7 s)

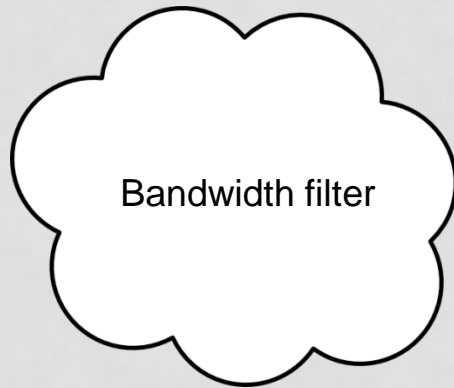
Channel conditions

ref	Reference (48 kHz)
EVS	(SWB) Enhanced Voice Services (EVS) at 16.4 kbit/s
G722	(WB) G.722 at 64 kbit/s
AMRWB	(WB) AMR-WB at 6.6 kbit/s
G711	(NB) G.711 at 64 kbit/s
AMRNB	(NB) AMR-NB at 4.65 kbit/s
BP900_2500	Bandpass filtering 900–2500 Hz
MNRU_10DB	Modulated Noise Reference Unit (MNRU) of 10 dB

Outline

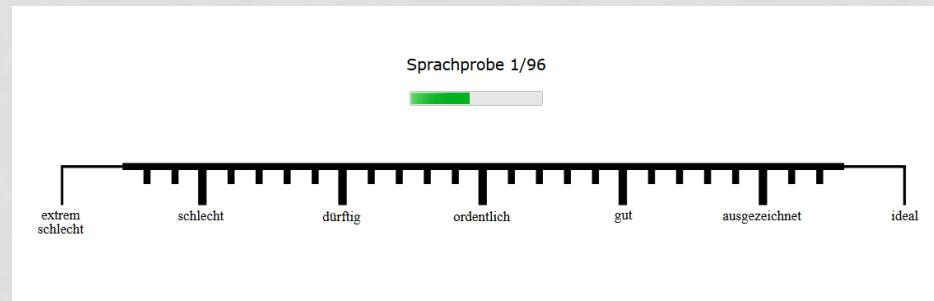
- Speech material
 - Speakers
 - Channel degradations
- Speech quality assessments 
 - Subjective
 - POLQA
- Effects of speakers' "warmth-attractiveness"
 - On subjective MOS
 - On POLQA MOS
- Conclusions

Subjective Speech Quality Assessments



„Bitte beurteilen Sie die nachfolgenden Sprachproben nach ihrer Gesamtqualität“

Listeners:
10 males
10 females




Instrumental Speech Quality Assessments

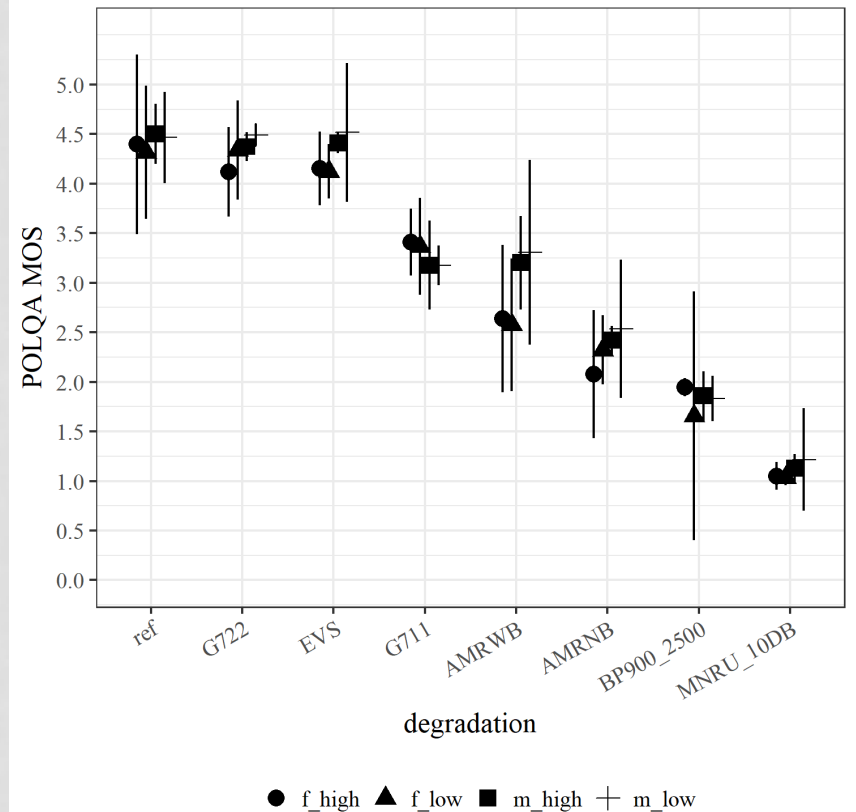
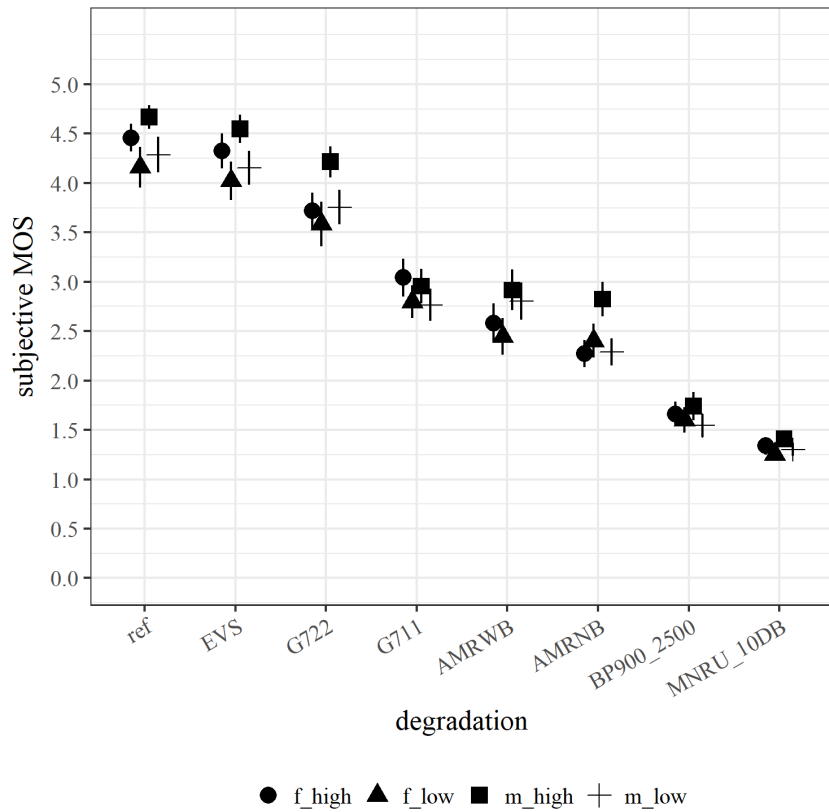


- POLQAv2
- SQquadAnalyzer v.2.4.2.7 in SWB mode
- Speech pre-processing as indicated in ITU-T Rec. P.863

Outline

- Speech material
 - Speakers
 - Channel degradations
- Speech quality assessments
 - Subjective
 - POLQA
- Effects of speakers' "warmth-attractiveness" 
 - On subjective MOS
 - On POLQA MOS
- Conclusions

Subjective and POLQA MOS

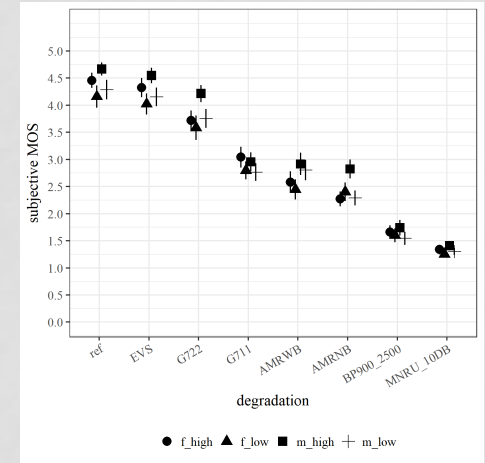


Subjective MOS

- For each channel condition separately:
- 2-way repeated measures ANOVA to test the effects of:
 - gender
 - high or low WAAT
 - their interactions (gender:WAAT)
- Significance level of $p < 0.01$

Subjective MOS

- 1) higher (lower) perceived quality when speakers' WAAT was also high (low)
- 2) male speech always rated with higher quality than female speech, except for G.711
- 3) Stronger effect of WAAT compared to the effect of gender

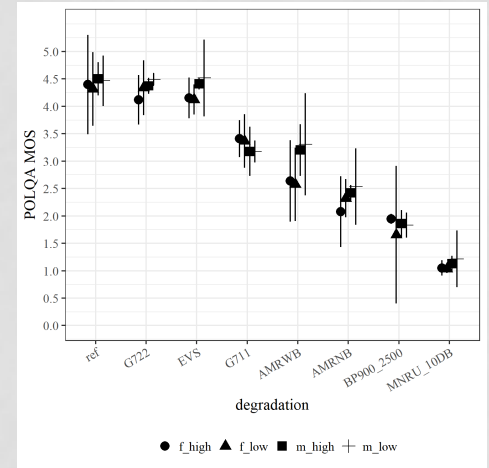


POLQA MOS


- For each channel condition separately:
- Two-sample T-Tests evaluating the effects of:
 - gender
 - high or low WAAT
- Significance level of $p < 0.01$

POLQA MOS

- 1) Unlike for subjective MOS: POLQA does not account for perceptual differences in WAAT
- 2) Like for subjective MOS: POLQA accounts for perceptual differences in gender



Outline

- Speech material
 - Speakers
 - Channel degradations
- Speech quality assessments
 - Subjective
 - POLQA
- Effects of speakers' “warmth-attractiveness”
 - On subjective MOS
 - On POLQA MOS
- Conclusions 

Conclusions

- Speech quality assessments:
 - Human listeners take WAAT and gender into account
 - POLQA takes only gender into account (not WAAT)
- Discussion: is the purpose of POLQA...
 - ...to accurately predict subjective scores?
 - ...or to predict the quality disregarding speaker peculiarities?
- For subjective tests, we recommend:
 - Selecting high and low WAAT speakers
 - Or selecting only moderate WAAT speakers
- A WAAT listening test protocol as Annex in ITU-T P.800

Thank you for your attention!

Laura Fernández Gallardo

laura.fernandezgallardo@tu-berlin.de

<http://www.qu.tu-berlin.de/?id=lfernandez>