World Class Standards

STF 294

Phase 2:

Step 1: Data Selection

Step 2: Development of Objective Model

STQ Workshop (Sophia Antipolis, May, 23th 2007)

S. Poschen, H.W. Gierlich
HEAD acoustics GmbH

Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Introduction

- Initial problem: known, standardized methods (PESQ, TOSQA2001, …) do not work for noisy environments

- Proceeding within STF:
  1. Experts: data selection acc. ToR
  2. Model Development:
     - Expert analysis of listening test results
     - Model development

- Afterwards: validation of model by Telefonica / Universidad de Valladolid

Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Expert Analysis of Listening Test Results

- **Two Purposes:**
  - Select conditions within the scope of the model
  - Extract parameters influencing human's assessment decision

- **Criteria for the condition selection:**
  - Artifacts, others than intended by the data generation process
  - Inconsistencies within one condition due to the selection of the individual speech samples for the listening test
  - Inconsistencies within one condition due to statistical variation in signal generation process (packet loss, …)
  - Inconsistencies due to P.56 level adjustment process chosen for the complete files including the background noise (French data)
  - Influences of different listening levels used in French and Czech database

Results of Experts Selection

- **General:**
  - proposed sample length of 4s (P.835) may be too short even for expert listeners
  - mostly all samples (French) of network condition 1&3 (0% PL, 3% PL) were retained;
  - network condition 2 (1% PL, 20ms jitter) samples were rejected due to inconsistent distribution of PL in the 8 sentences of 1 condition

- **French database:**
  - 6 sentences of one condition were rated by 4 listeners each
  - level of each sentence + background noise adjusted to 79dB$_{SPL}$ (P.56)
  - most samples retained, but 28 conditions (of NI and NIII) rejected of due to
    - not consistent (high) signal levels caused by amplification to 79dB$_{SPL}$
    - insufficient S/N → speech almost inaudible

→ 260 of 432 conditions were retained (60%), 179 for training and 81 for validation

Results of Expert Selection

- **Czech database:**
  - 1 sentence (randomly) per condition was rated by 24 listeners
  - level of each sentence + background noise NOT adjusted after processing
    - → lower level than French samples
    - → level variation within Czech data of up to 16 dB
  - conditions (of NI and NIII) retained, if
    - at least one paket loss occured during speech and one during background noise
    - the overall active speech level is at least 69dB$_{SPL}$
    - the background noise level is consistent compared to the speech level

→ 88 of 432 conditions were retained (20%), 60 for training and 28 for validation

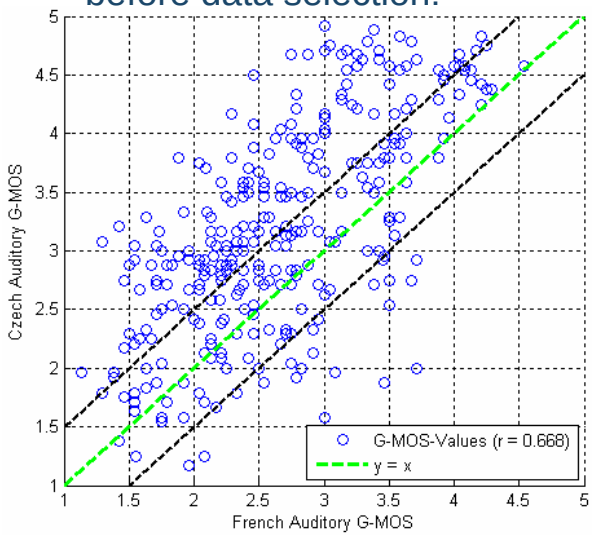
Results of Expert Selection

➔ correlation between the French and Czech S-/N-/G-MOS was increased by the data selection process:

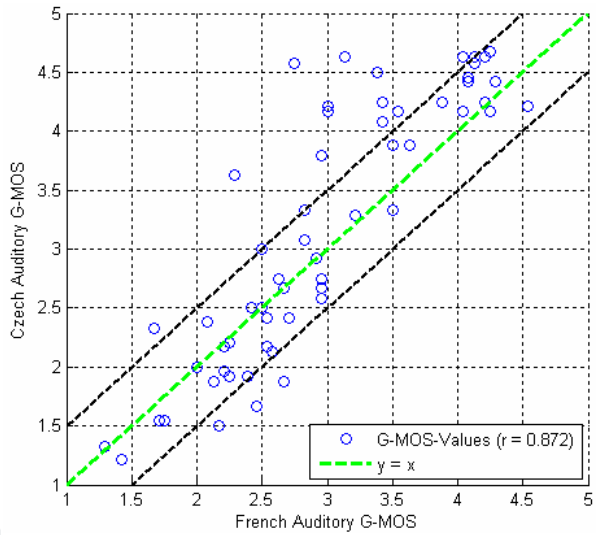
Over all available ratings (French and Czech, 302 condition each)	**Only Czech and French selected MOS Data (NI and NIII conditions, ratings reviewed by experts)** (59 conditions selected for French *and* Czech)
S-MOS: 0.703 N-MOS: 0.816 G-MOS: 0.668	S-MOS: 0.830 N-MOS: 0.897 G-MOS: 0.871

e.g. G-MOS

before data selection:



after data selection:



Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Purpose of Model

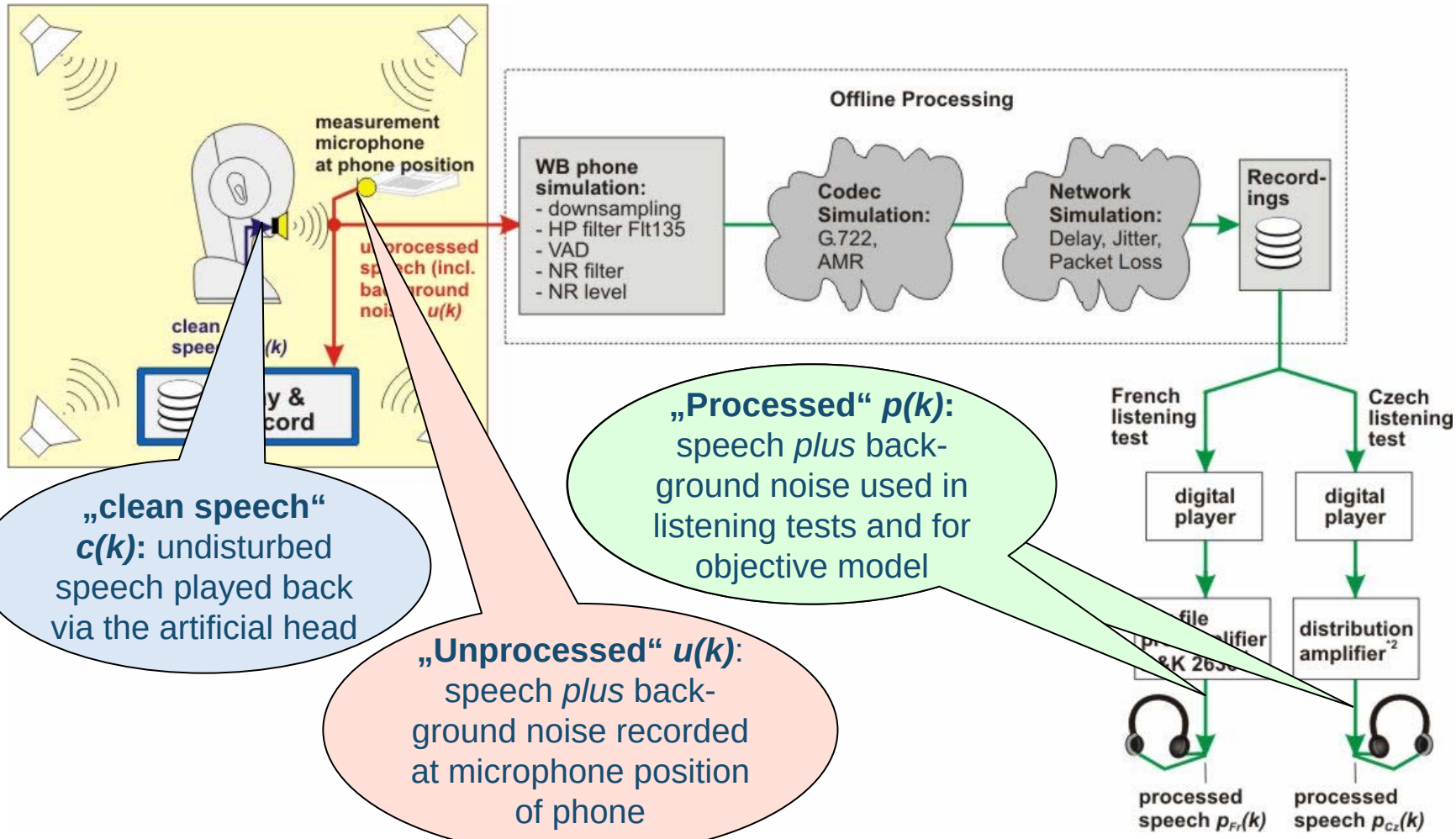
Principles of new model:

- **Reproduction of human perception by choosing a hearing-adequate analysis**
  - → **Use parameters from expert analysis results**
- **High correlation to given STF database**
- **Assurance of robustness for other databases**

Relevant Parameters (based on experts analysis)

- **for N-MOS**
  - absolute background noise level
  - modulation of background noise
  - “naturalness” of background noise
  - lost packets
- **for S-MOS**
  - Level and quality of processed background noise
  - Signal to noise ratio (SNR) between speech and noise in the processed signal
  - Change in SNR before and after processing
  - modulation of speech, speech sound, “naturalness”
  - lost packets

Sample Generation Procedure



Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Base Analysis: Relative Approach

Relative Approach:

- Hearing-adequate time and frequency resolution → 3D „spectrograph“
- forward estimation analogue human expectation based on signal history
- unexpected patterns shown as „estimation errors“
- no reference needed
- applicable for packet loss, VAD, background noises

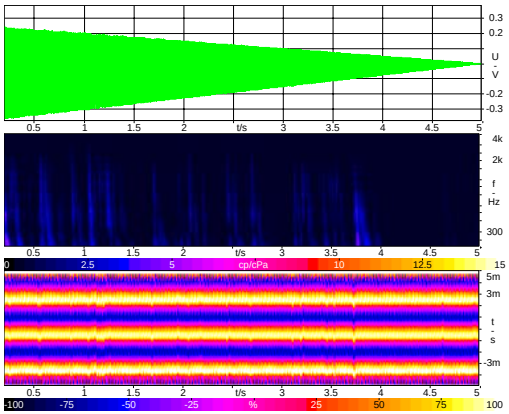
Variant: „Δ Relative Approach“

- Determination of „similarity“ between two signals by subtracting two Relative Approach 3D spectrographs

→ 3D „delta-spectrograph“

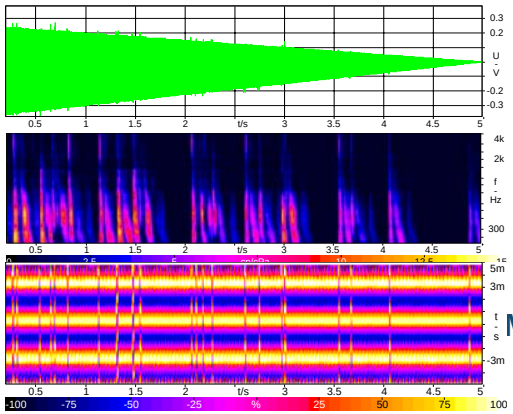
Relative Approach: Analysis Examples

RA: PLC



0% 5%

MOS-LQO 3.67

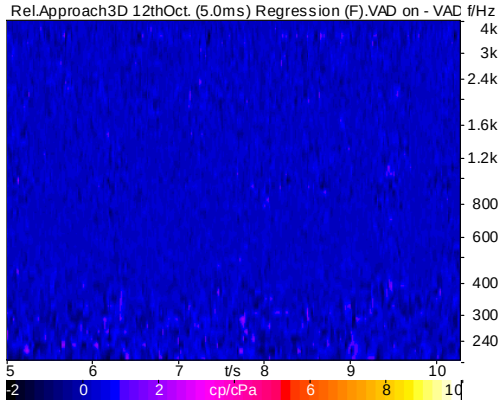


0% 5%

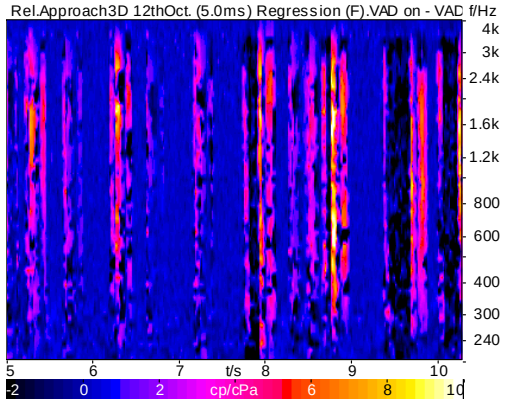
MOS-LQO 3.31

ΔRA: BGNT + VAD

Original Background Noise:



VAD implementation 1:

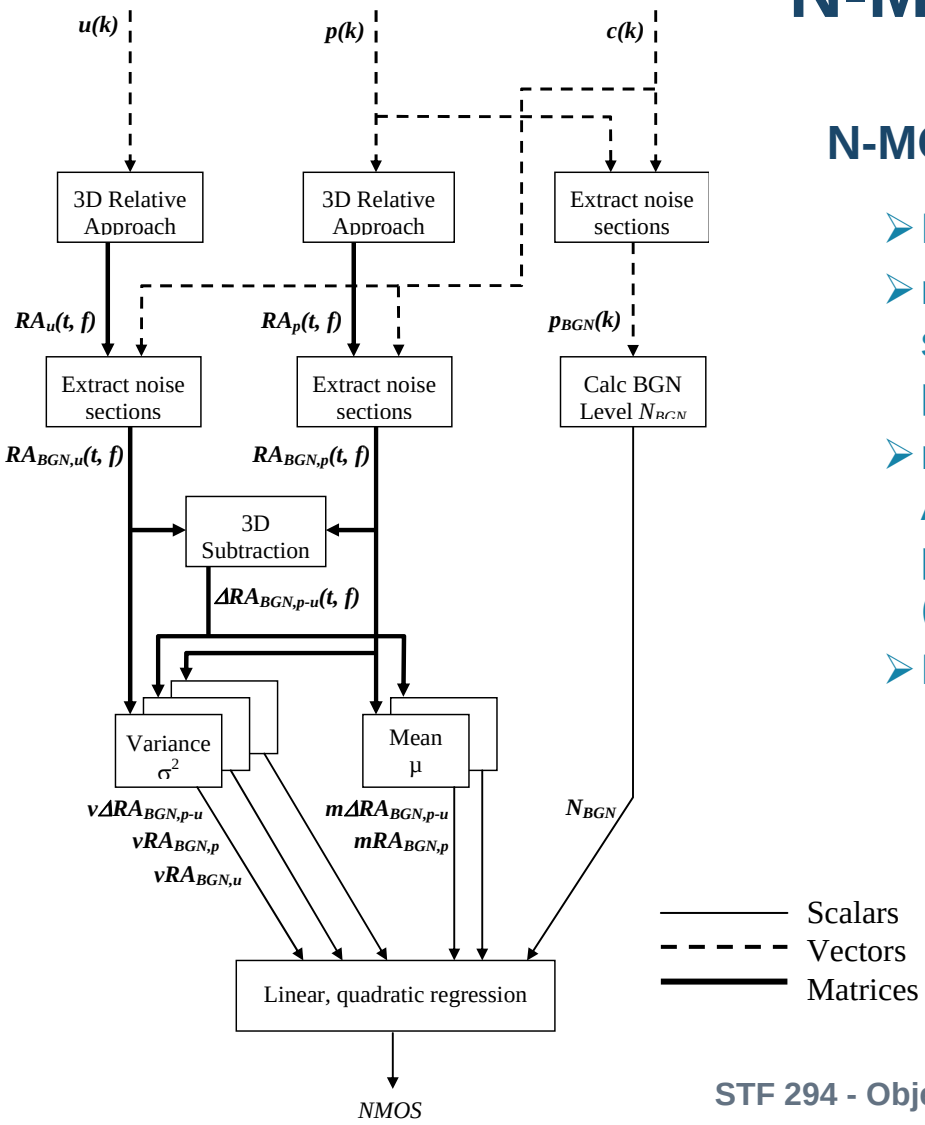


VAD implementation 2:

Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

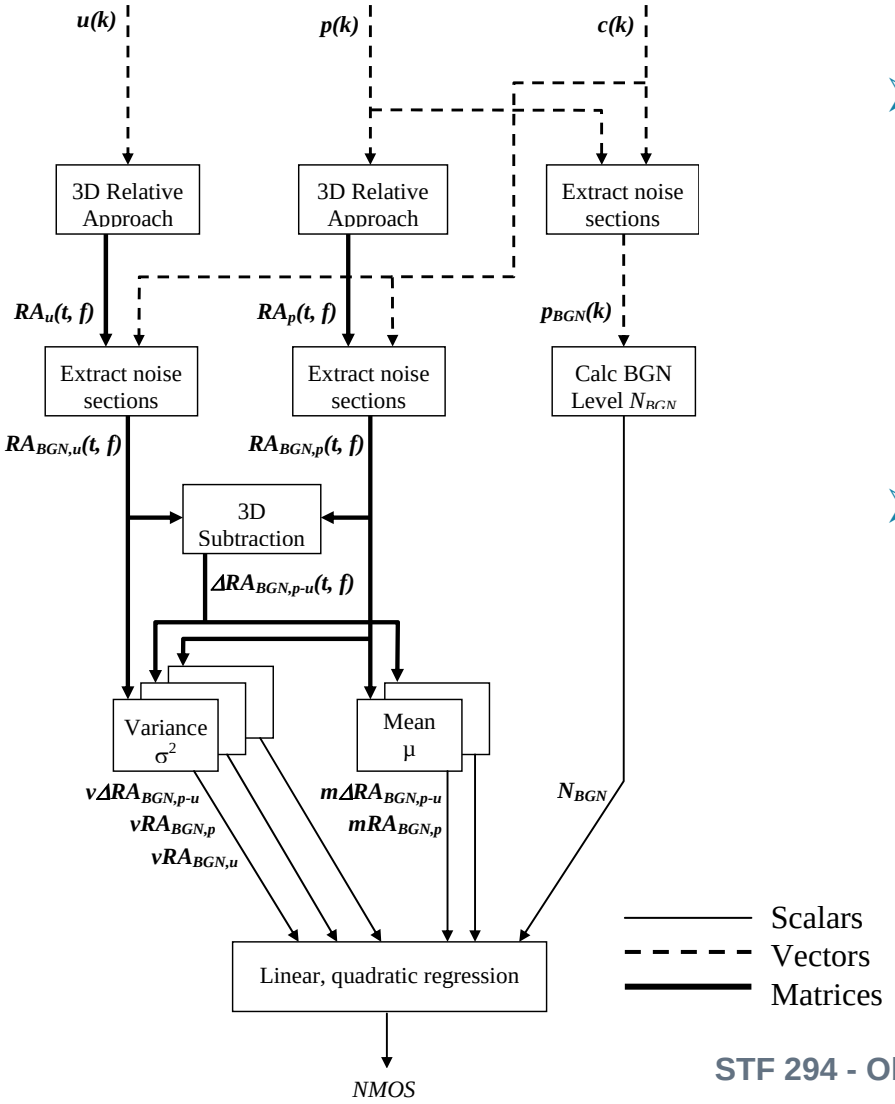
N-MOS



N-MOS calculation is based on:

- Level of processed background noise
- mean and variance of 3D Relative Approach spectrographs of unprocessed and processed signal (during only BGN parts)
- mean and variance of 3D Δ Relative Approach spectrograph: processed – unprocessed signal (during only BGN parts)
- Mapping by linear quadratic regression

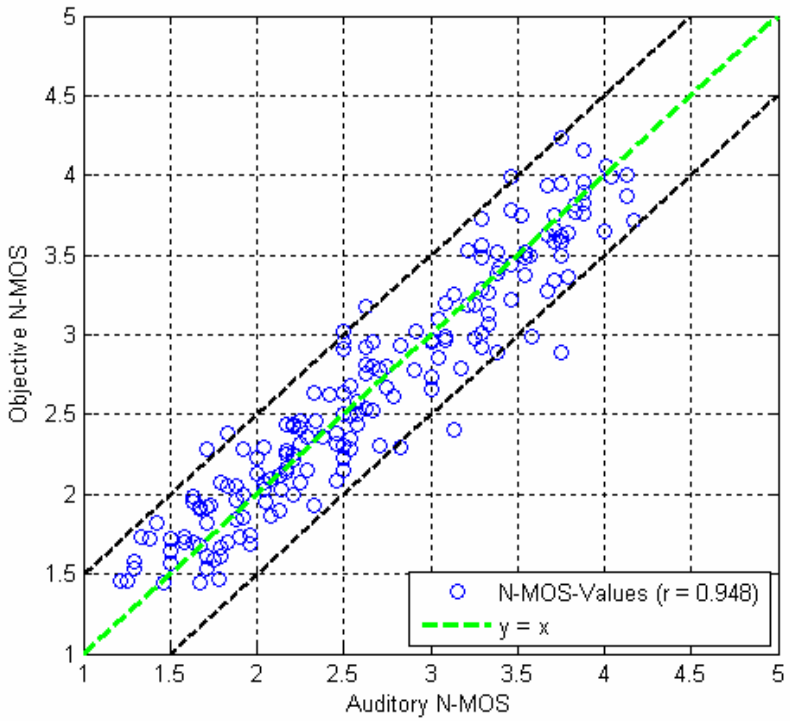
N-MOS – more descriptive…

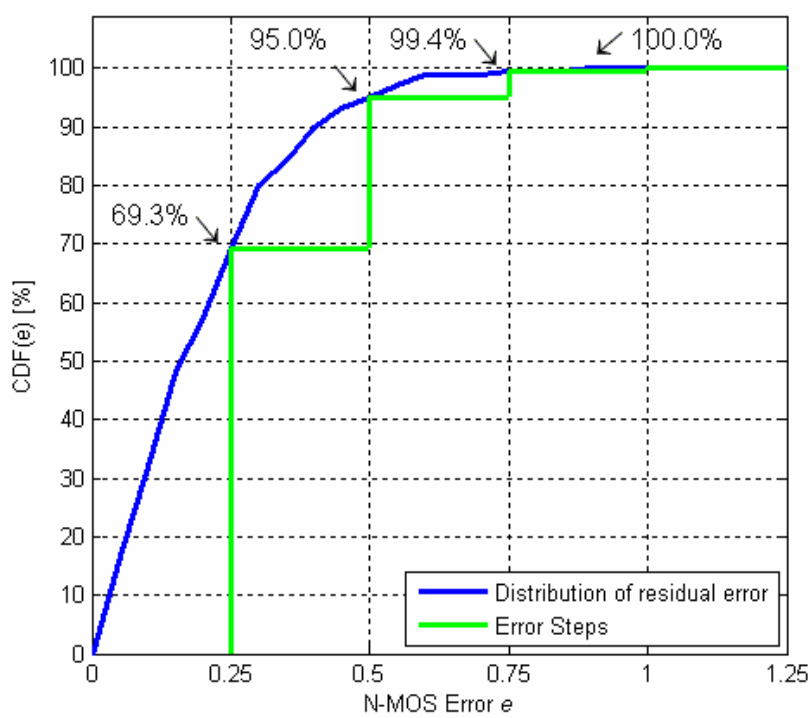


- characterization of *changes* in background noise (Δ Relative Approach)
  - mean: amount of similarity, "truth-to-original"
  - variance: covers musical tones, modulations …
- "anchors":
  - characterization of background noise before and after processing (single Relative Approach)
  - level of processed background noise

Calculation works aurally adequate !

N-MOS: objective vs. subjective (French Data)

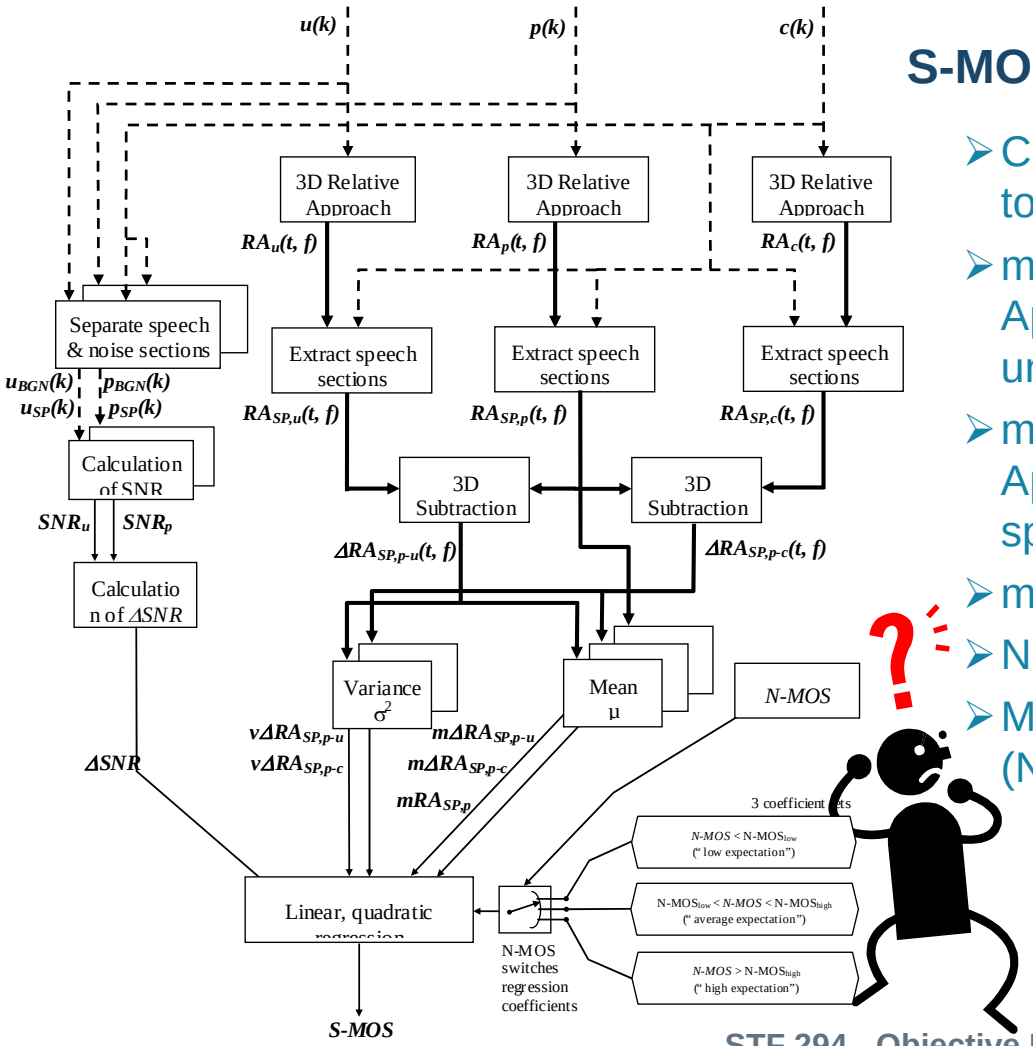




Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

S-MOS



S-MOS calculation is based on:

- Change of signal to noise ratio (SNR) due to processing (influence of NR, VAD ...)
- mean and variance of 3D Δ Relative Approach spectrograph between unprocessed and processed signal
- mean and variance of 3D Δ Relative Approach spectrograph between clean speech and processed signal
- mean of processed signal
- N-MOS
- Mapping by linear quadratic regression (N-MOS defines coefficient set)

S-MOS – more descriptive …

Analysis of listening test results:

- → high influence of SNR on S-MOS for speech transmission out of noisy environments
- → high influence of transmitted background noise on (subjectively) perceived speech quality

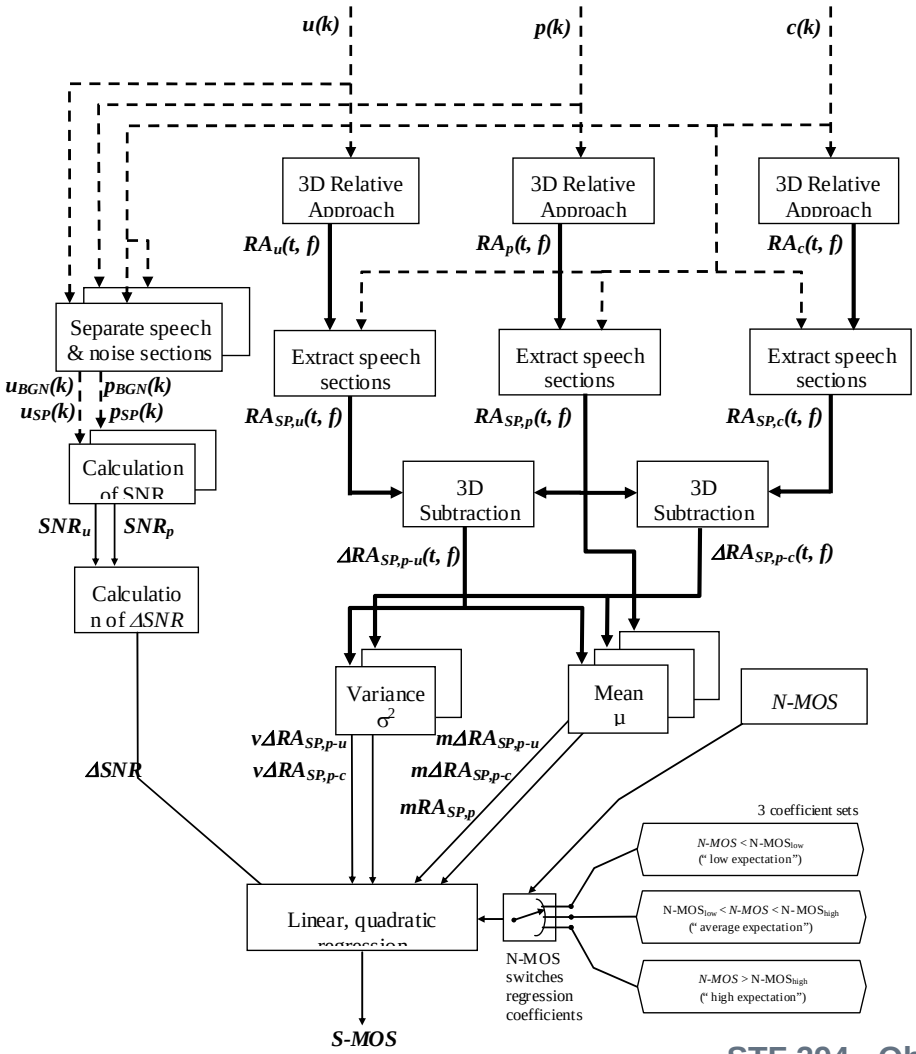
high quality of background noise:

- → high expectation to speech quality
- → compare speech quality between clean speech signal (only speech!) and processed signal

low quality of background noise:

- → low expectation to speech quality
- → compare speech quality between unprocessed speech signal (speech plus background noise!) and processed signal

S-MOS – more descriptive …



Δ Relative Approach:

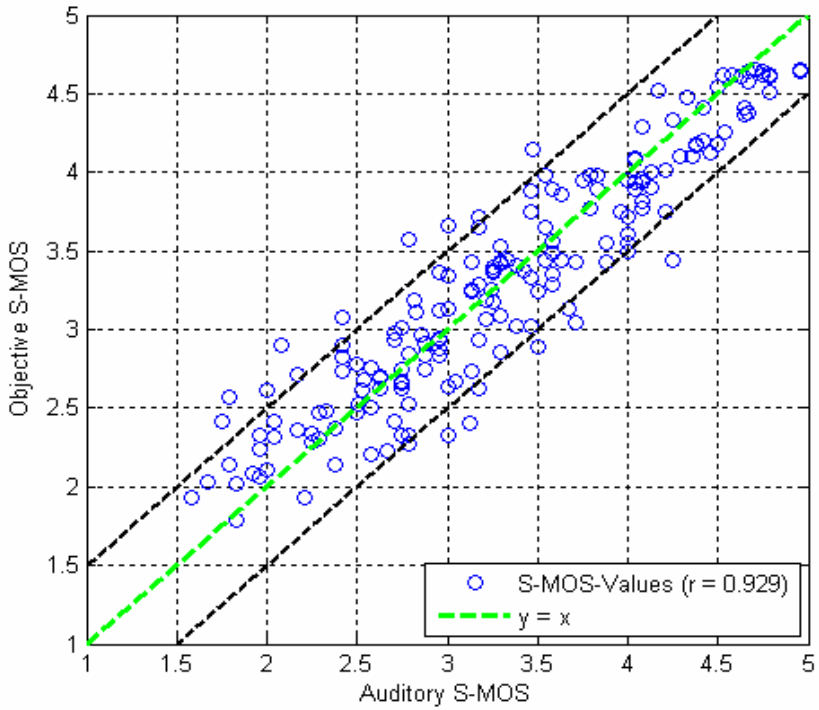
- comparison of two speech qualities
- mean: amount of similarity, "truth-to-original"
- variance: covers musical tones, modulations …

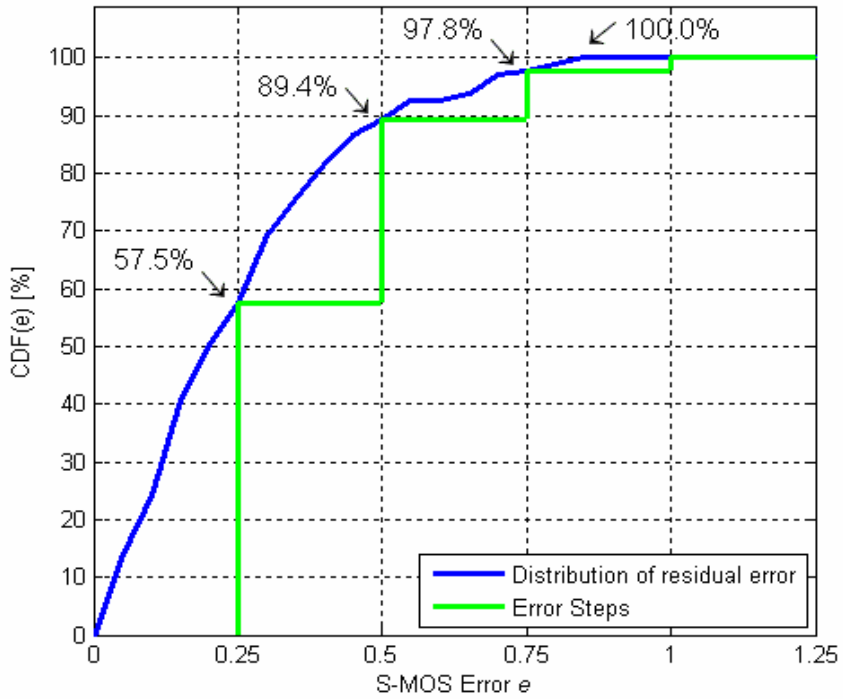
N-MOS:

- determines speech quality expectation
- weight of comparison:
  clean speech ⇆ unprocessed

Calculation works aurally adequate !

S-MOS: objective vs. subjective (French Data)

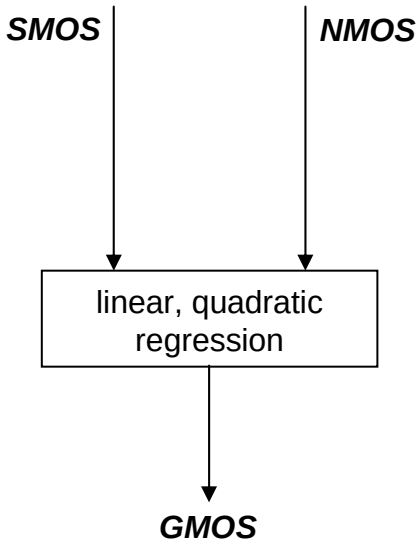




Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

G-MOS



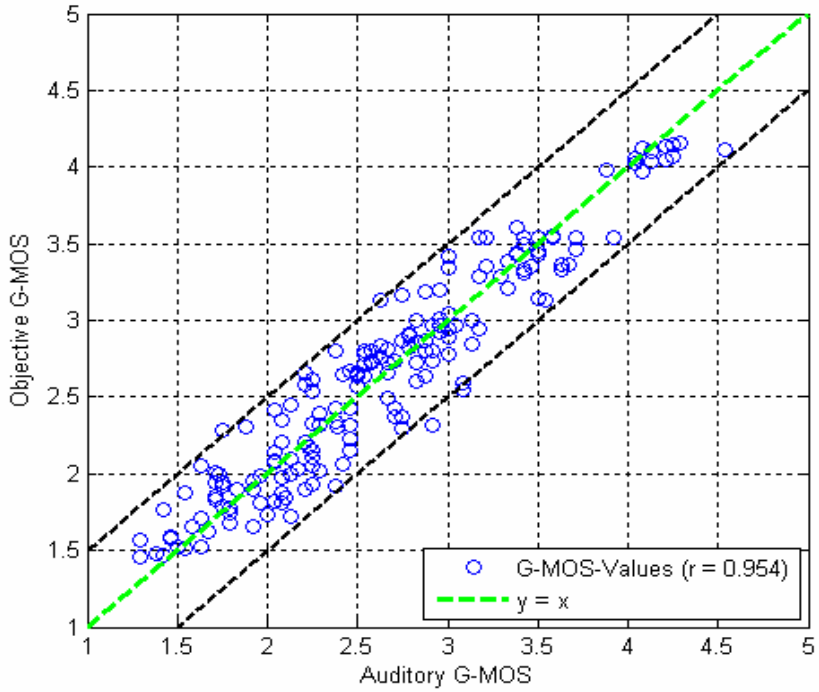
G-MOS calculation is based on:

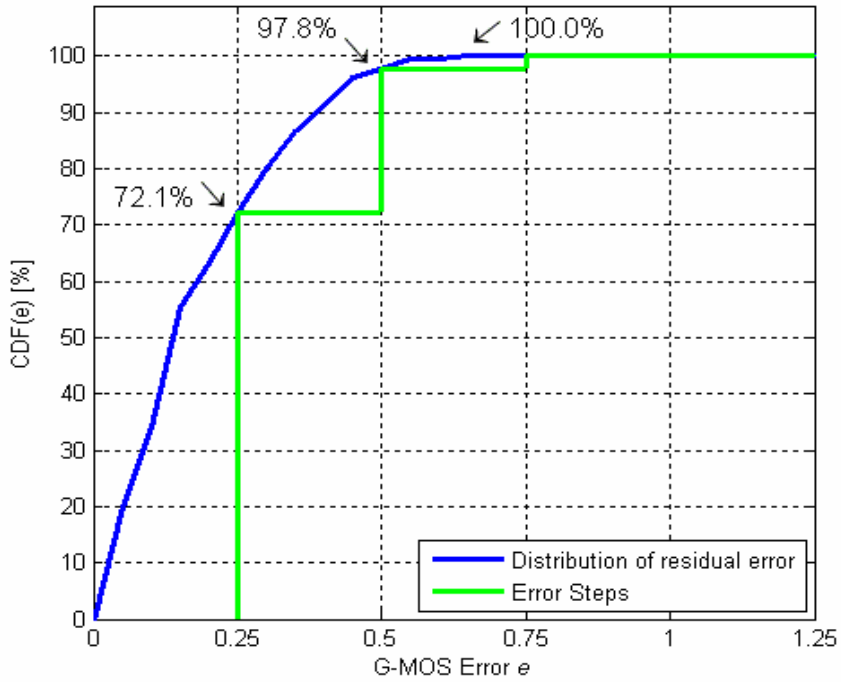
- S-MOS
- N-MOS
- Mapping by linear quadratic regression

→ **S-MOS und N-MOS cover all subjective, perceptual effects during speech and noise assessment**

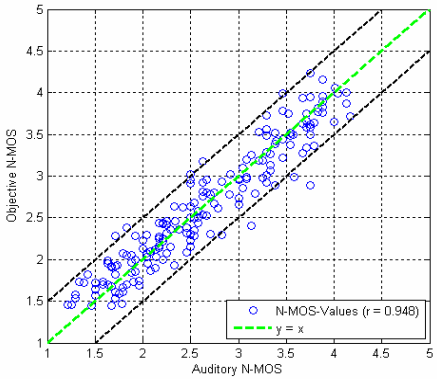
→ **simple combination leads to „global“ MOS**

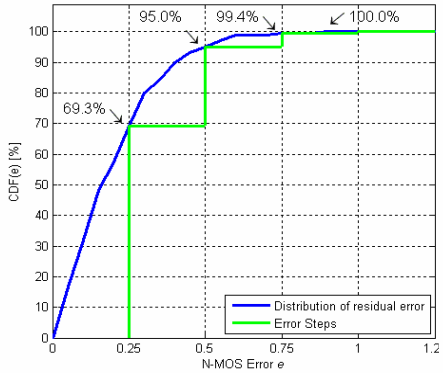
G-MOS: objective vs. subjective (French Data)



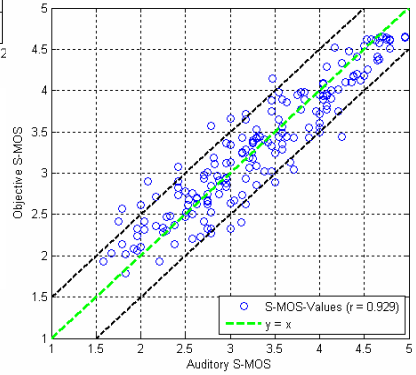


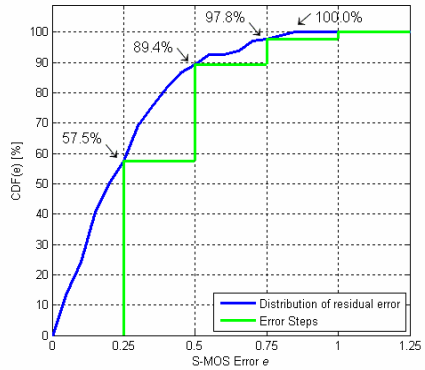
Objective MOS - Summary (French Data)



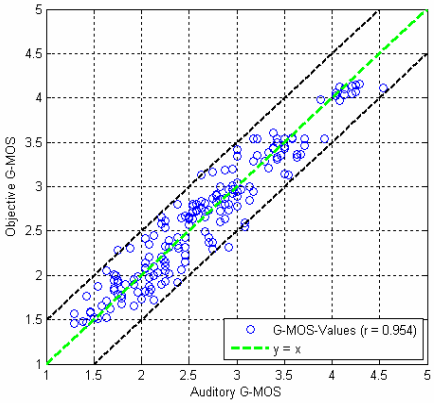


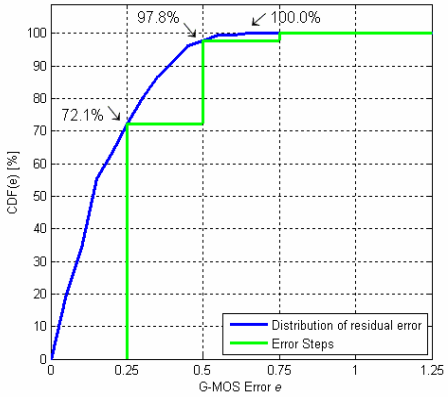
N-MOS (94.8%)





S-MOS (92.9%)





G-MOS (95.4%)

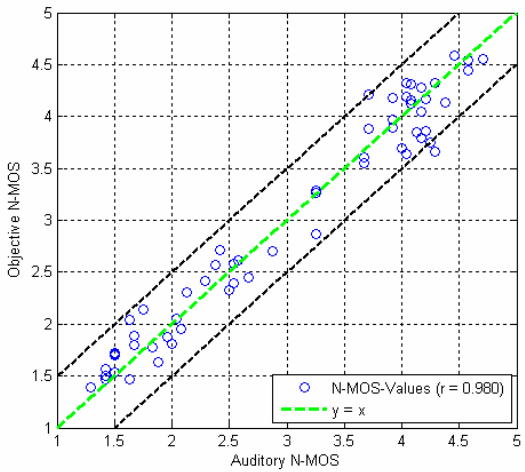
Overview

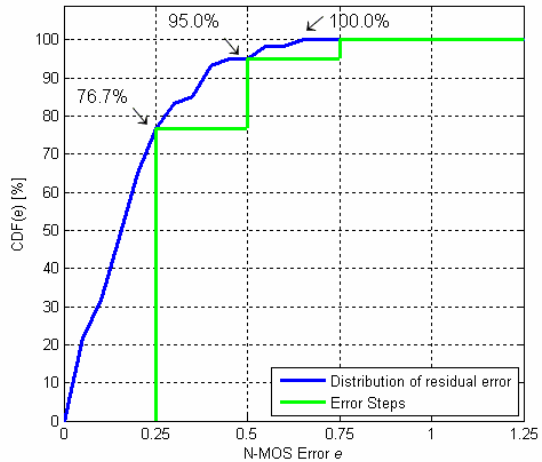
- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

French vs. Czech Data

- **Main difference: level strategy**
- **Assumption: same perceptual processes active during French and Czech listening test**
  - Use same algorithms for S-MOS, N-MOS and G-MOS
  - but: separate training for Czech data → Czech regression coefficient set
  - Training uses ALL NI conditions + retained NIII conditions → provide higher numerical stability

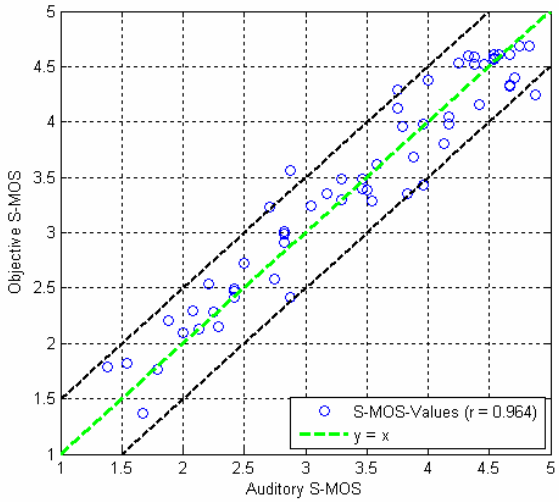
N-MOS (Czech data)

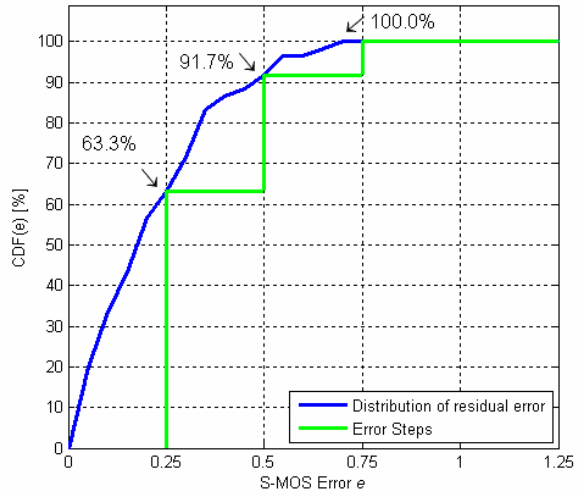




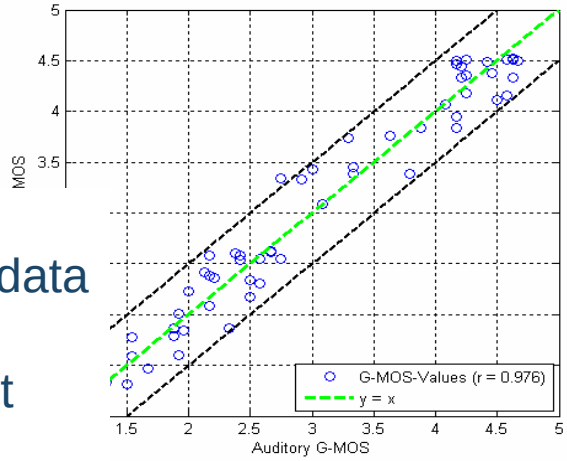
S-MOS / G-MOS: objective vs. subjective (Czech Data)

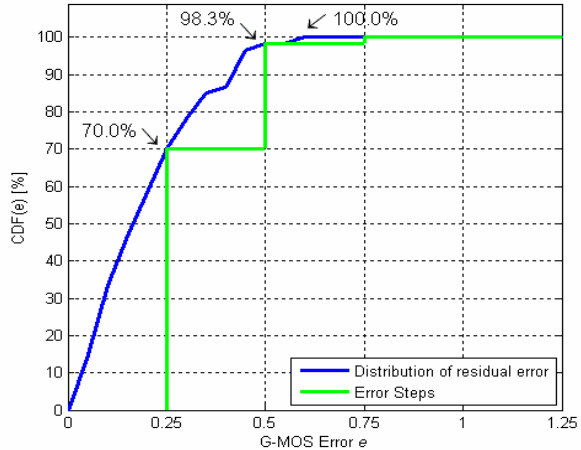
S-MOS (Czech Data)





G-MOS (Czech Data)

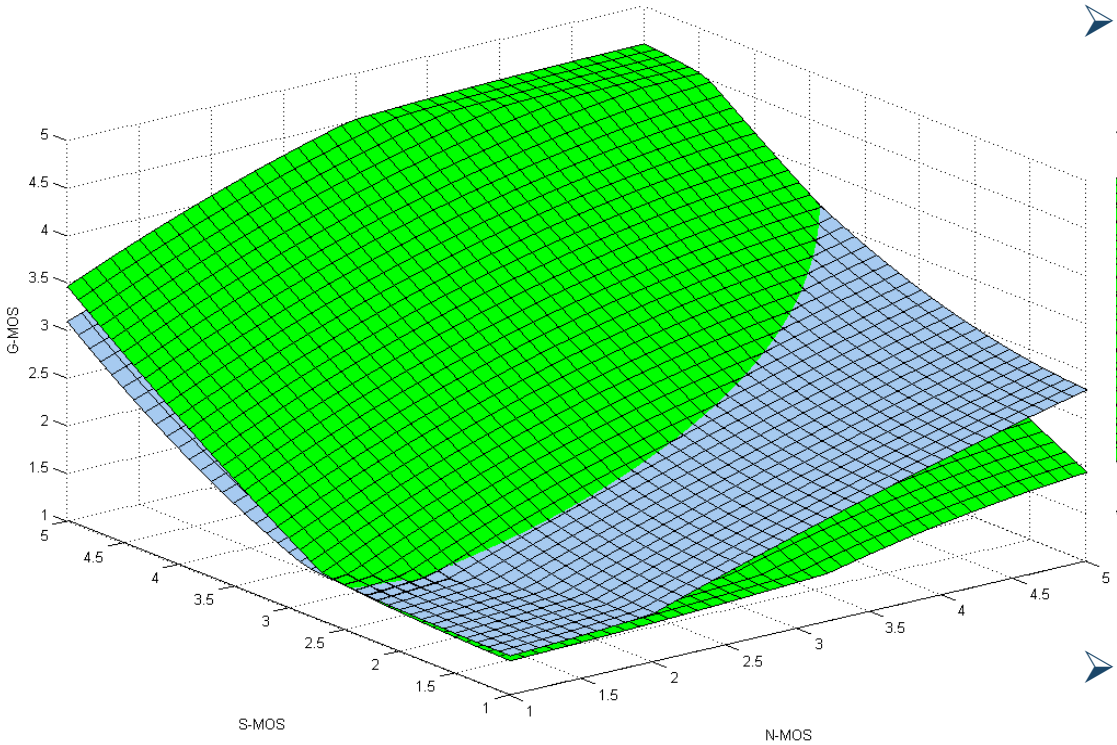


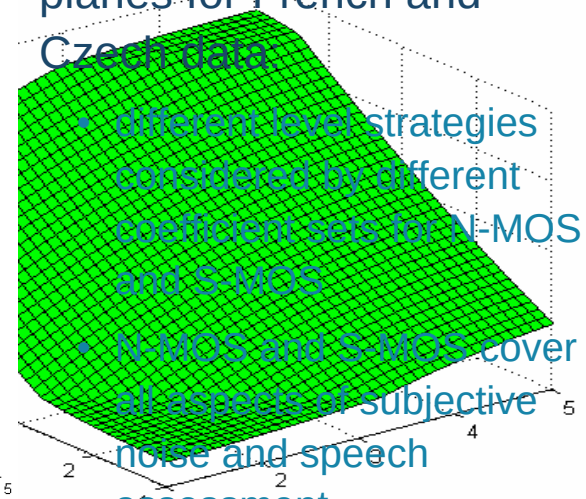


- New model also applicable on Czech data
- But: new training of model due to different level strategy

G-MOS: French vs. Czech

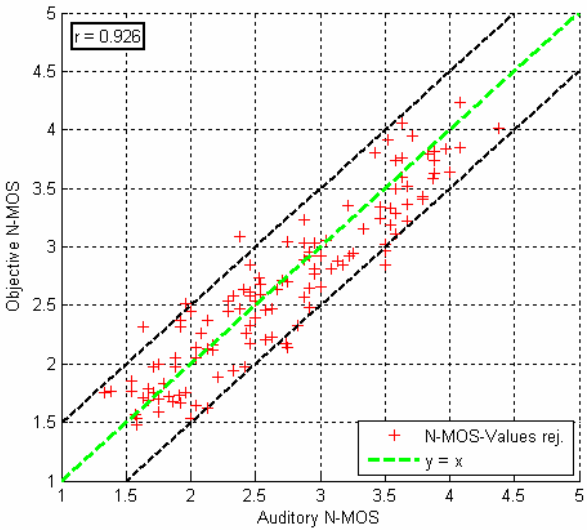
- **Check assumption: “same perceptual processes active during French and Czech listening test”**
- **Compare G-MOS regression planes: G-MOS = f(S-MOS, N-MOS)**

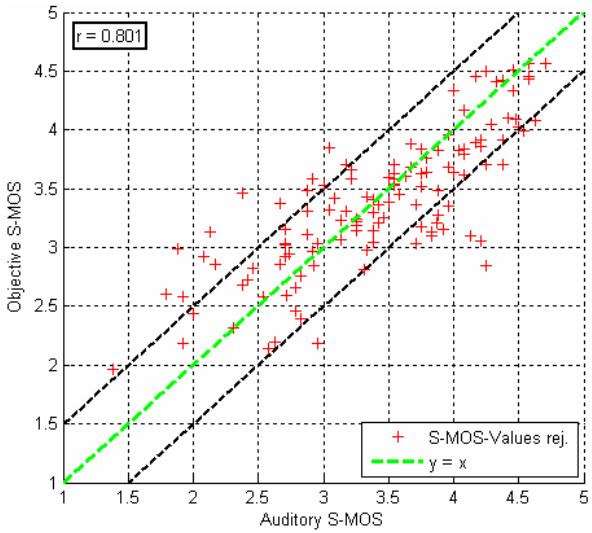


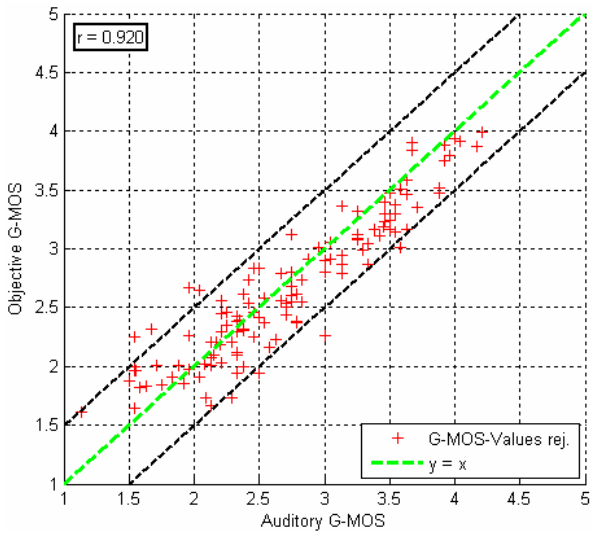


- Similar G-MOS regression planes for French and Czech data:
  - different level strategies considered by different coefficient sets for N-MOS and S-MOS
  - N-MOS and S-MOS cover all aspects of subjective noise and speech assessment
- Similar “overall” perception in both listening tests

Check: Rejected Data (French)

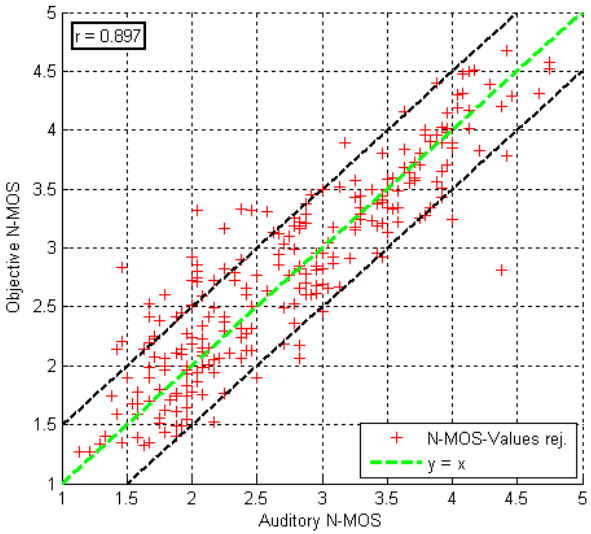


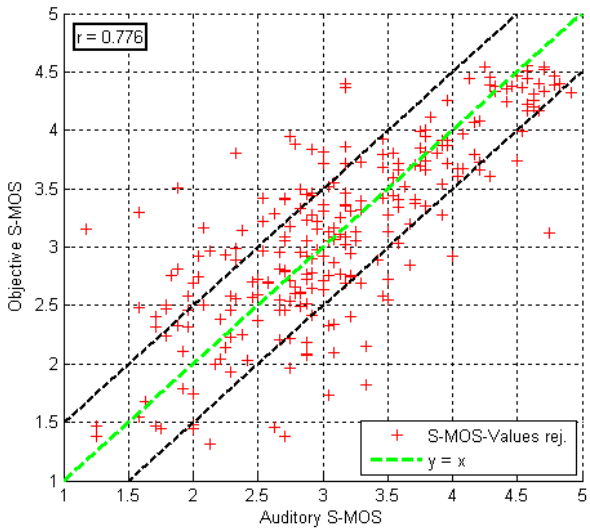


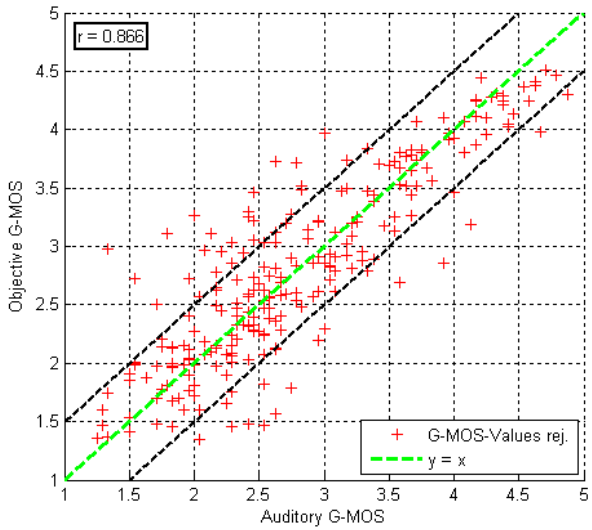


- Model works in principle also for rejected French data
- correlation for N-MOS and G-MOS > 92%
- S-MOS calculation more critical, still correlation of 80%

Check: Rejected Data (Czech)

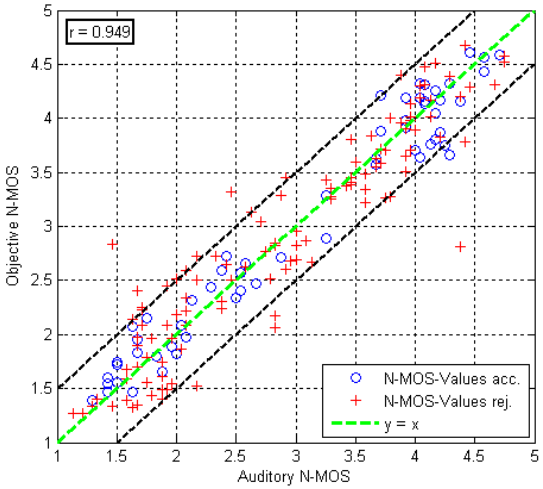


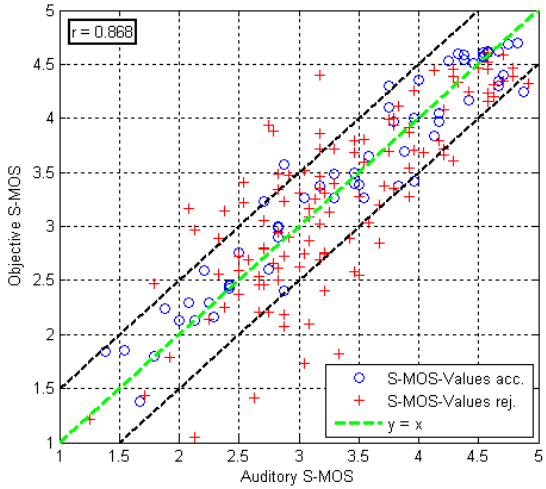


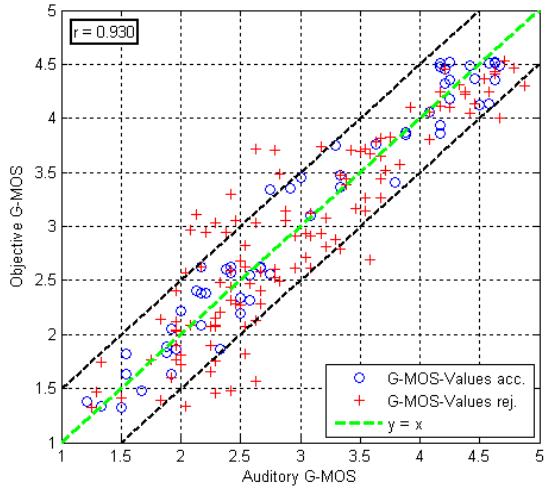


- most of the Czech conditions can be assessed by new model
- lower overall correlation due to higher level variations within Czech data

Czech Results: level > 69dB







- rejected with level higher than 69dB and retained Czech data
- correlation for N-MOS and G-MOS > 93%
- S-MOS calculation most critical, still correlation of 86%
- new model requires “typical” signal levels

Overview

- Introduction
- Step 1: Expert review of listening test results and data selection
- Step 2: Development of new objective model
  - Purposes of model
  - Principles of Relative Approach
  - N-MOS:
    - calculation
    - N-MOS: objective vs. subjective
  - S-MOS:
    - calculation
    - S-MOS: objective vs. subjective
  - G-MOS:
    - calculation
    - G-MOS: objective vs. subjective
  - Further Analyses
- Summary

Ideas for Future Upgrades

- Additional analyses based on Relative Approach
- Extension to narrowband transmission systems
- Improvement concerning assessment of jitter and packet loss concealment
  - only few data with jitter in STF database
  - no packet loss concealment algorithms considered
- Enhancement to other transmission codecs (G.729, G.729.1, GSM, CDMA, …)
- …

Conclusion

- **New model able to predict P.835 listening tests – (for moderate level differences within databases)**
- **No language dependencies observed**
  - due to different designs of the listening tests
  - due to complex assessment task
- **N-MOS, S-MOS calculation**
  - uses psycho-acoustic model → Relative Approach
  - are based on human perception
  - are INDEPENDENT of the specific signal processing (NR, VAD …) implemented
- **Complex, hearing-adequate calculation of N-MOS and S-MOS allows simple combination to / calculation of G-MOS**