Session III: New ETSI Model on Wideband Speech and Noise Transmission Quality – Phase II

STF 294 – Validation Results

TELEFÓNICA I+D / UNIVERSIDAD DE VALLADOLID
Date:23/05/2007

Index

- Objective: Validate the Objective Test Method
- 130 out of the 432 initial conditions per language
  - Due to the consistent problems -> 81 French and 28 Czech
- The process carried out to validate the Objective Test Method had the following steps:
  - Objective results obtaining using the developed calculation algorithms -> N/S/G-MOS
  - Comparison between previously obtained objective results and the subjective results
  - Results comparison -> global and per language

Metrics (I)

- **Obtain -> accuracy, monotonicity and consistency of the Objective Test Method**
- **Characterization through Statistical Metrics:**
  - Root Mean Square Error
  - Pearson Correlation
  - Spearmans' Rank Correlation Coefficient
  - Kendall Tau Rank Correlation Coefficient
  - Residual Error Distribution

Root Mean Square Error (RMSE)

- RMSE measures the difference between values predicted by the algorithm and the auditory values to evaluate its accuracy
- Optimum value RMSE = 0

$$RMSE = \sqrt{\frac{1}{N}\sum_{N} Perror[i]^2}$$

$$Perror(i) = MOS(i) - MOS_p(i)$$

N = number of samples
MOS = subjective MOS
MOSp = predicted MOS

Pearson Correlation Coefficient (R)

- R measures the linear relationship between the algorithm performance and the subjective data
- R varies from -1 to 1 (R=1 -> perfect linear relationship)

$$R = \frac{\sum_{i=1}^{N} (Xi - \overline{X}) * (Yi - Y)}{\sqrt{\sum (Xi - \overline{X}^2) * \sqrt{\sum (Yi - \overline{Y}^2)}}}$$

Xi = subjective MOS
Yi = predicted MOS

- The 95% confidence interval – values of R for which the difference between the parameter and the observed estimate is not statistically significant at the 5% level

$$z \pm 2 \cdot \sigma_z$$

$$z = 0.5 \cdot \ln\left(\frac{1+R}{1-R}\right) \qquad \sigma_z = \sqrt{\frac{1}{N-3}}$$

N = number of samples

Spearman's Rank Correlation Coefficient (ρ)

- ρ assesses how well an arbitrary monotonic function could describe the relationship between two variables
- ρ varies from -1 to 1 (optimum value ρ =1)

$$\rho = 1 - \frac{6 \cdot \sum d_i^2}{N(N^2 - 1)}$$

di = difference between each rank of corresponding values of x and y
N = number of samples

Kendall Tau Rank Correlation Coefficient (T)

- T measures the degree of correspondence between two rankings.
- T varies from -1 to 1 (optimum value T =1)

$$\tau = \frac{4\sum q_i}{N(N-1)} - 1$$

qi = sum of samples ranked after the given sample
N = number of samples

Residual Error Distribution (e)

$$e = |MOS_{auditory} - MOS_{objective}|$$

- Perfect result: e =0
- To evaluate the consistency we used the Cumulative Density Function (CDF) applied to the error e
  - The graphical representation of the CDF will show the number of conditions which yields a maximum residual error

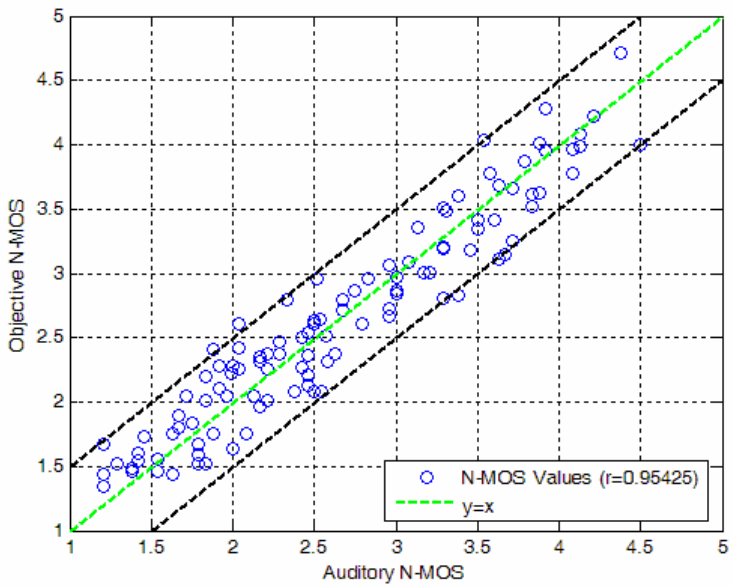
02 All Conditions Results Analysis

Comparing Subjective and Objective N-MOS Results

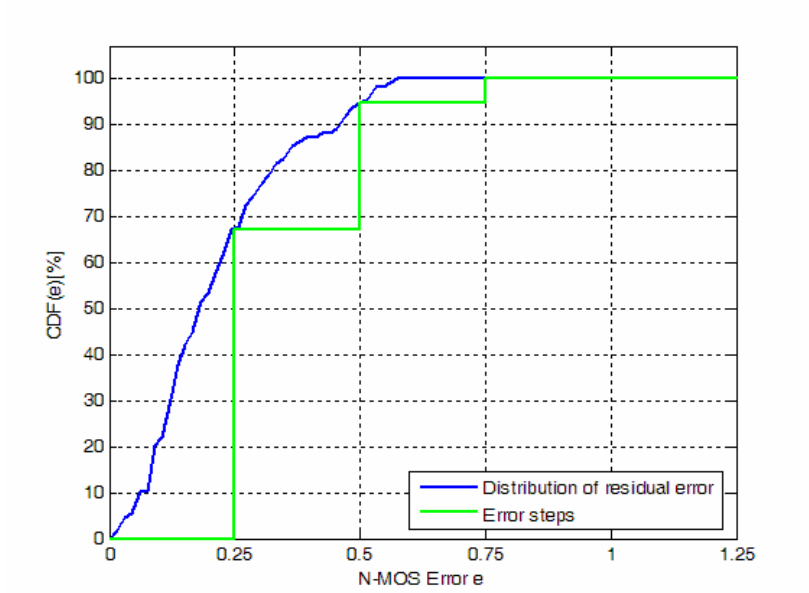
Pearson correlation = **0.954**; confidence interval **[0.933, 0.969]**

Spearman Correlation Coefficient = **0.952**; Kendall Tau = **0.821**

RMSE = **0.255** e<**0.25** for **67%**; e<**0.6** for **99%**



Objectively calculated N-MOS versus auditory N-MOS for validation conditions



Objectively CDF of residual error versus N-MOS error e for validation conditions

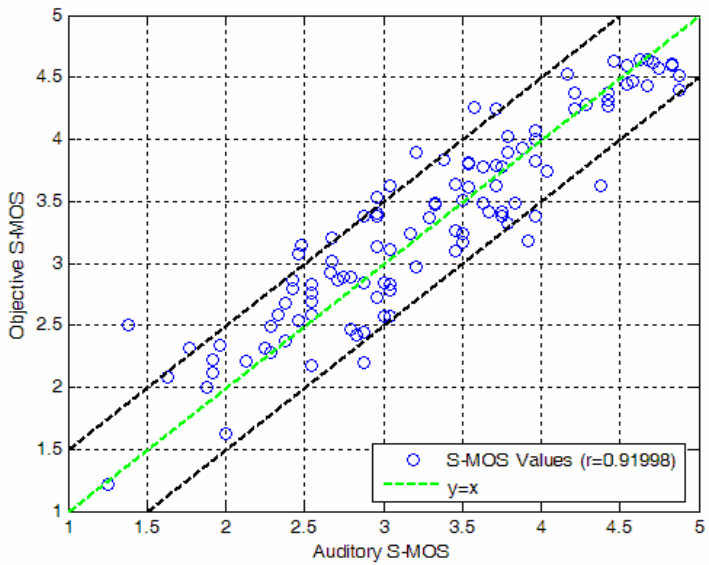
02 All Conditions Results Analysis

Comparing Subjective and Objective S-MOS Results

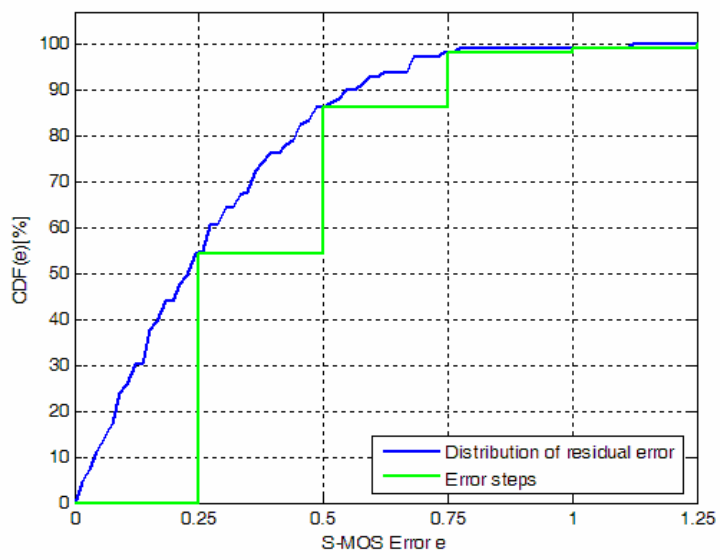
Pearson correlation = **0.920**; confidence interval **[0.884,0.945]**

Spearman Correlation Coefficient = **0.914**; Kendall Tau = **0.749**

RMSE = **0.338** e<**0.25** for **55%**; e<**0.75** for **99%**



Objectively calculated S-MOS versus auditory S-MOS for validation conditions



Objectively CDF of residual error versus S-MOS error e for validation conditions

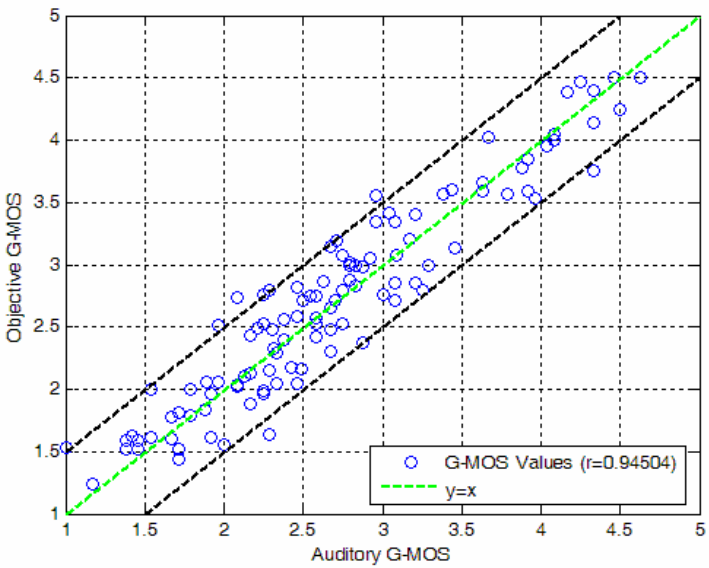
02 All Conditions Results Analysis

Comparing Subjective and Objective G-MOS Results

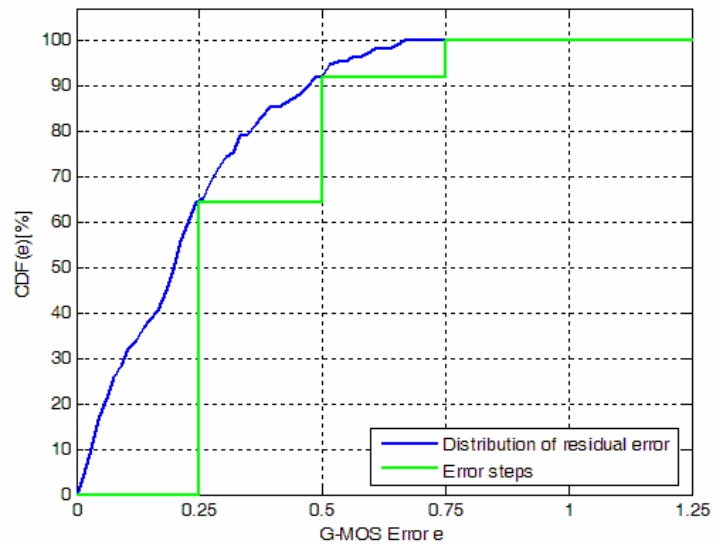
Pearson correlation = **0.945**; confidence interval **[0.920,0.962]**

Spearman Correlation Coefficient = **0.935**; Kendall Tau = **0.793**

RMSE = **0.272** e<**0.25** for **65%**; e<**0.7** for **99%**



Objectively calculated G-MOS versus auditory G-MOS for validation conditions



Objectively CDF of residual error versus G-MOS error e for validation conditions

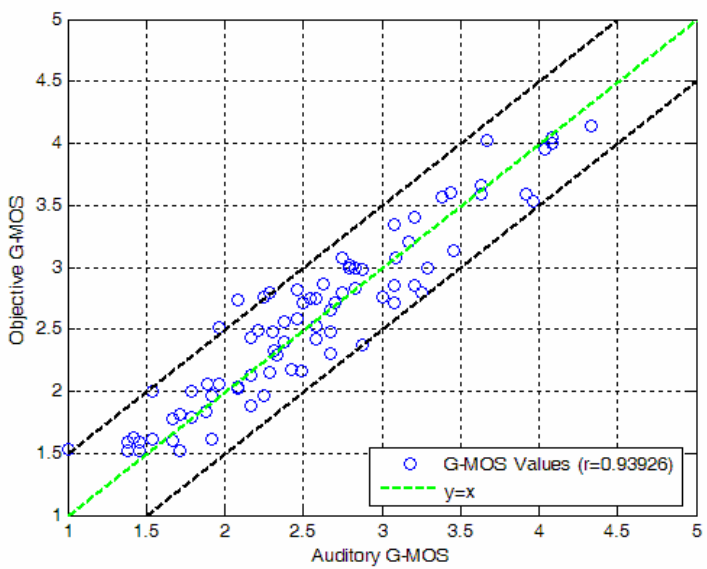
03 French Conditions Results Analysis

Comparing Subjective and Objective G-MOS Results

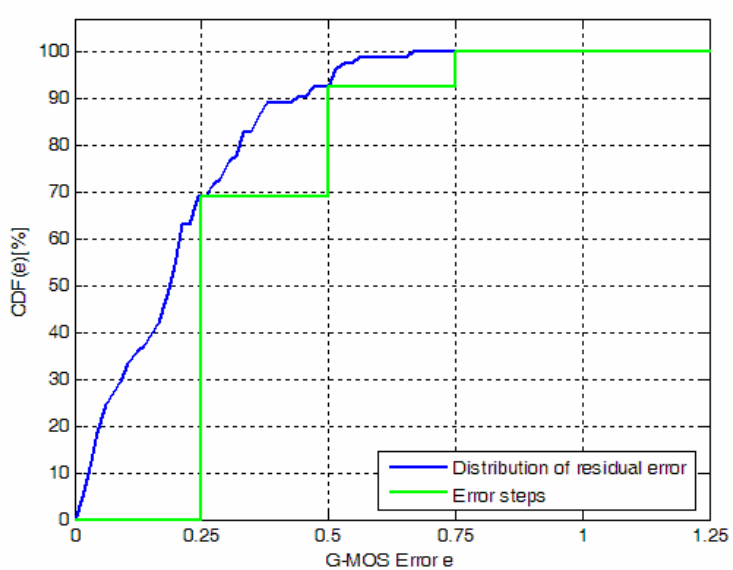
Pearson correlation = **0.939**; confidence interval **[0.906,0.961]**

Spearman Correlation Coefficient = **0.925**; Kendall Tau = **0.781**

RMSE = **0.253** e<**0.25** for **70%**; e<**0.65** for **99%**



Objectively calculated G-MOS versus auditory G-MOS for French validation conditions



Objectively CDF of residual error versus G-MOS error e for French validation conditions

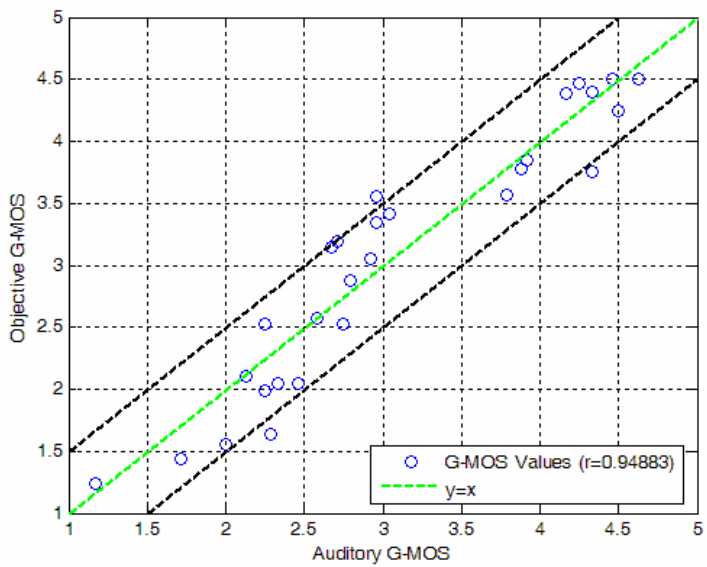
04 Czech Conditions Results Analysis

Comparing Subjective and Objective G-MOS Results

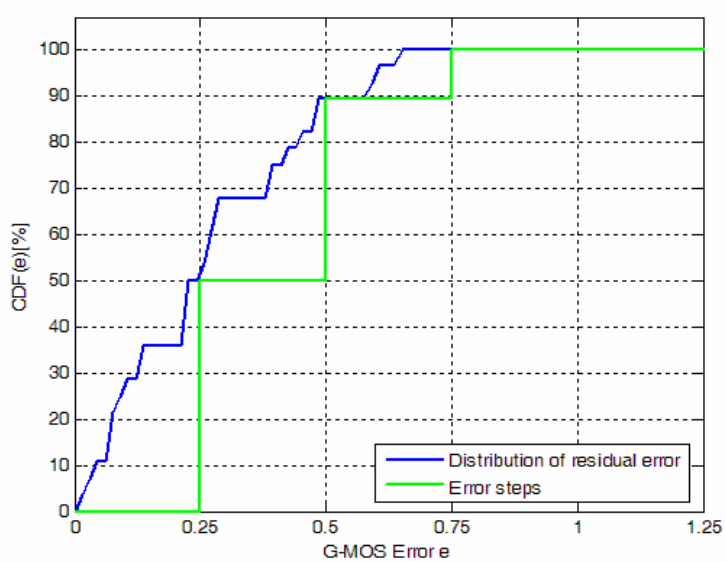
Pearson correlation = **0.949**; confidence interval **[0.892,0.976]**

Spearman Correlation Coefficient = **0.935**; Kendall Tau = **0.793**

RMSE = **0.321** e<**0.25** for **50%**; e<**0.65** for **99%**



Objectively calculated G-MOS versus auditory G-MOS for Czech validation conditions



Objectively CDF of residual error versus G-MOS error e for Czech validation conditions


Telefonica


Universidad de Valladolid
UVa