

Predicting Narrow-band and Wideband Speech Quality with WB-PESQ and TOSQA

Nicolas Coté^{1,2}, S. Möller², V. Gautier-Turbin¹

¹ France Télécom R&D, TECH/SSTP/MOV, France

² Deutsche Telekom Laboratories, Germany

(nicolas.cote@orange-ftgroup.com)

ETSI Workshop on Speech and Noise in Wideband Communication,
22nd and 23rd May 2007 - Sophia Antipolis, France



Summary

- 1 ■ Introduction: "Auditory tests" vs. "Prediction models"
- 2 ■ Presentation of the WB models
 - WB-PESQ
 - TOSQA-2001
- 3 ■ Analysis of WB models performance
 - Evaluation procedure
 - Results
 - Analysis of modified WB models
- 4 ■ Conclusion

1

Introduction: "Auditory tests" vs. "Prediction models"

Auditory tests

- WB in Auditory test (User's quality judgments)
 - WB speech assessed since earlier 80's (development of G.722 speech codec)
 - Few studies with 5-point MOS scale in a mixed-band context
 - Introduction of the "full-band"
- Improvement of quality for WB transmissions compared to NB transmissions: $\approx 29\%$ (ITU-T Rec. G.107 Appendix 2)

Prediction models

- Prediction models for WB speech quality: Several models developed during the ITU-T competition for PESQ.
 - WB-PSQM
 - WB-PAMS
 - TOSQA-2001
- WB-PESQ: normalized in 2005 (ITU-T Rec. P.862.2)
- P.OLQA (Objective Listening Quality Assessment): New ITU-T competition for a full-band speech quality model

"Auditory tests" vs. "Prediction models"

- AQUAVIT project (EURESCOM P.905): the first study of several WB speech models
 - The WB-PESQ is better than TOSQA-2001 for a NB context. However, TOSQA-2001 gives better predictions for Mixed-band and WB contexts.

Test:	Bandwidth:	WB-PESQ	TOSQA-2001	WB-PAMS
1	MB	0.952	0.966	0.946
2a	NB	0.981	0.954	0.981
2b	WB	0.977	0.982	0.992

2

Presentation of WB models

PESQ (1/2)

- PESQ: Perceptual Estimation of Speech Quality (ITU-T Rec. P.862, 2001).
- An intrusive model which needs two audio files
- Consists of four stages:
 1. Input filter (*IRS*: Intermediate Reference System)
 2. Time alignment
 3. Perceptual model
 4. Mapping function

PESQ (2/2)

- Compensates the system's characteristics which are not relevant in auditory judgments
 - Frequency response
 - Variable gain
 - Variable delay (as seen above)
- Includes an asymmetrical measure (positive or negative disturbance)
- Calculates the speech quality prediction by the difference of the two signals in the perceptual domain and mapped to the MOS scale

WB-PESQ

- WB-PESQ, the WB version of PESQ (ITU-T P.862.2)
- Predicts auditory speech quality judgments in a Mixed-band context
- Two differences to PESQ:
 1. The input filter: FLAT
 2. The output mapping function
- Some changes were proposed in [Côté et al. 2006] in order to improve the correlation of WB-PESQ predictions with auditory judgments

TOSQA

- Predicts the speech quality of end-to-end systems, including both terminals. Different filters are available:
 1. IRS (modified send)
 2. Handset
 3. Telephone-band (300-3400 Hz)
 4. HATS 3.4 ear in free-field (in NB or WB)
 5. Wideband (200-7000 Hz)

- Calibration option: Digital level corresponding to 79 dB_{SPL}

- Speech quality predictions based on the correlation between both perceptual-transformed speech signals

TOSQA

■ Characteristics of TOSQA

1. The calculation of the frequency response is different

■ Algorithm:

1. Calibration
2. Overall noise calculation
3. Fix delay
4. VAD -> Delay estimation, Gain estimation, Perceptual transformation, Correlation
5. Mapping function

■ WB version of TOSQA: TOSQA-2001

3

Experiments

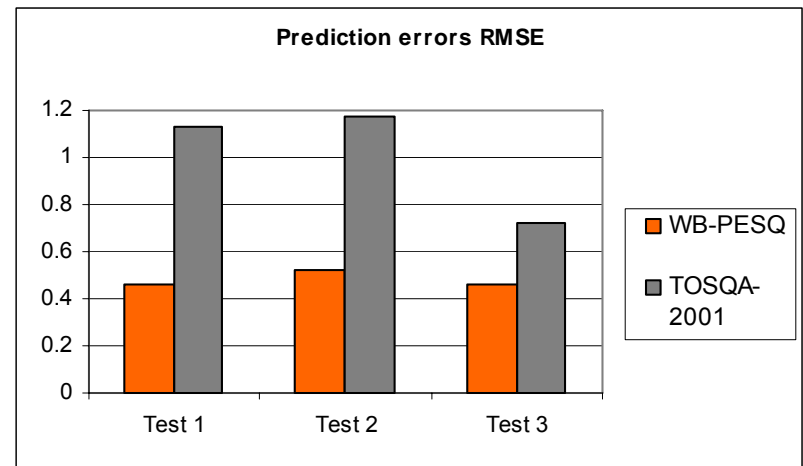
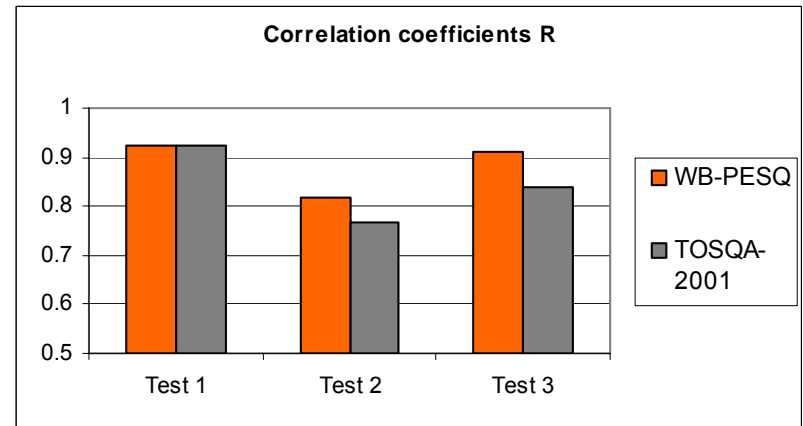
Evaluation Procedure

- 3 databases
- ACR listening quality in a Mixed-band context (MOS scale)
- Comparison:
 - MOS from **Auditory Tests** (LQSM)
 - MOS predictions from **WB models** (LQOM)

N°	Test description
Test 1	France Télécom R&D (2004) 25 NB and 11 WB conditions
Test 2	IKA – Ruhr-Universität (2005) 9 NB and 9 WB conditions (with band-pass conditions)
Test 3	France Télécom R&D (2006) 30 NB and 30 WB conditions, 36 conditions with packet-loss

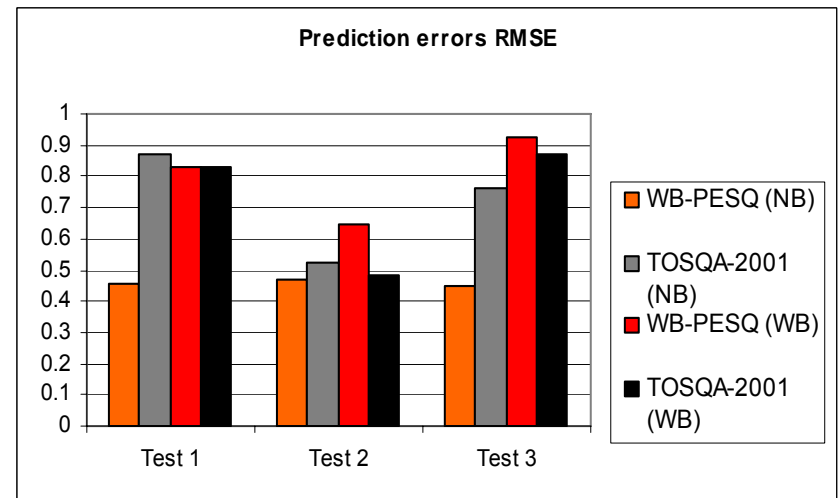
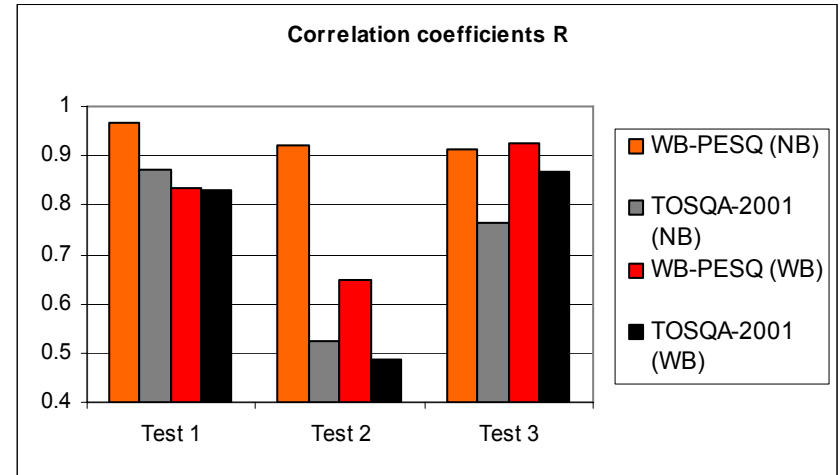
WB-PESQ vs. TOSQA-2001 (1/2)

- Model's options:
 - WB-PESQ: +16000 +wb
 - TOSQA-2001: FLAT/WIDEBAND F250 NORM
- WB-PESQ better than TOSQA-2001
 1. The Correlation coefficients are higher
 2. Prediction errors are lower



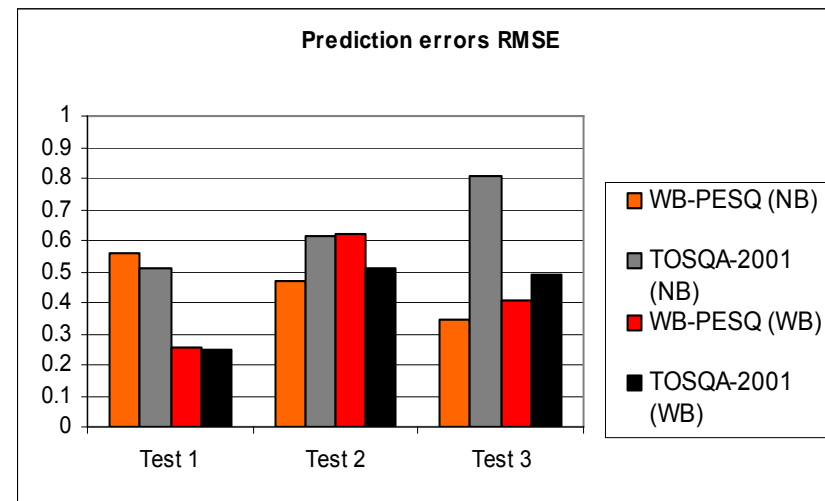
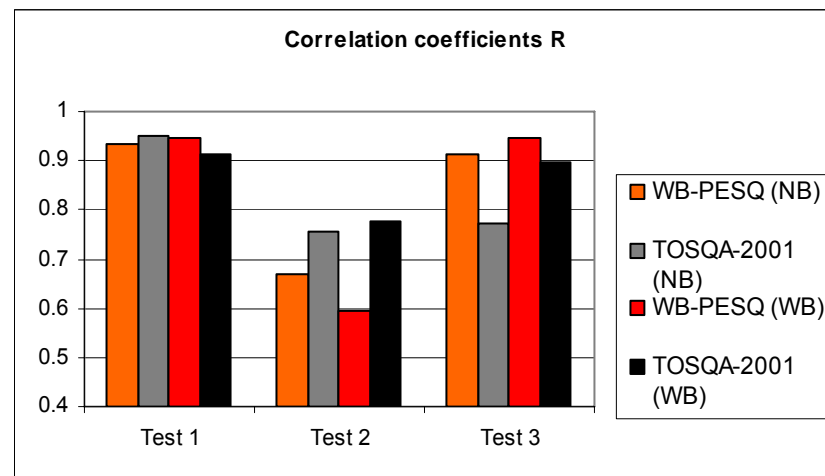
WB-PESQ vs. TOSQA-2001 (1/2)

- For NB conditions:
 - WB-PESQ is clearly better than TOSQA-2001
- For WB conditions:
 - The correlation coefficients are higher for WB-PESQ than TOSQA-2001
 - The RMSE is lower for TOSQA-2001
- WB-PESQ underestimates several WB conditions
- TOSQA-2001 underestimates the NB conditions (due to the IRS send filter)



Modified models results

- **Modified WB-PESQ:**
 - WB conditions: improve the reliability of WB-PESQ predictions with auditory judgments for
 - NB conditions: under-estimated
- **Modified TOSQA-2001:**
 - improve the TOSQA-2001 predictions of auditory judgments, especially for WB conditions.



4

Conclusions

Conclusions

- WB-PESQ speech quality model provides better estimations of user's judgments than TOSQA-2001

- Problems in both WB models:
 - WB-PESQ on several WB conditions
 - TOSQA-2001 on NB conditions

- Slight changes result in a better prediction for both models
 - improvements of the estimations
 - some problems still remain

Thank you