A Taxonomy of Quality Prediction Models Recommended by the ITU-T
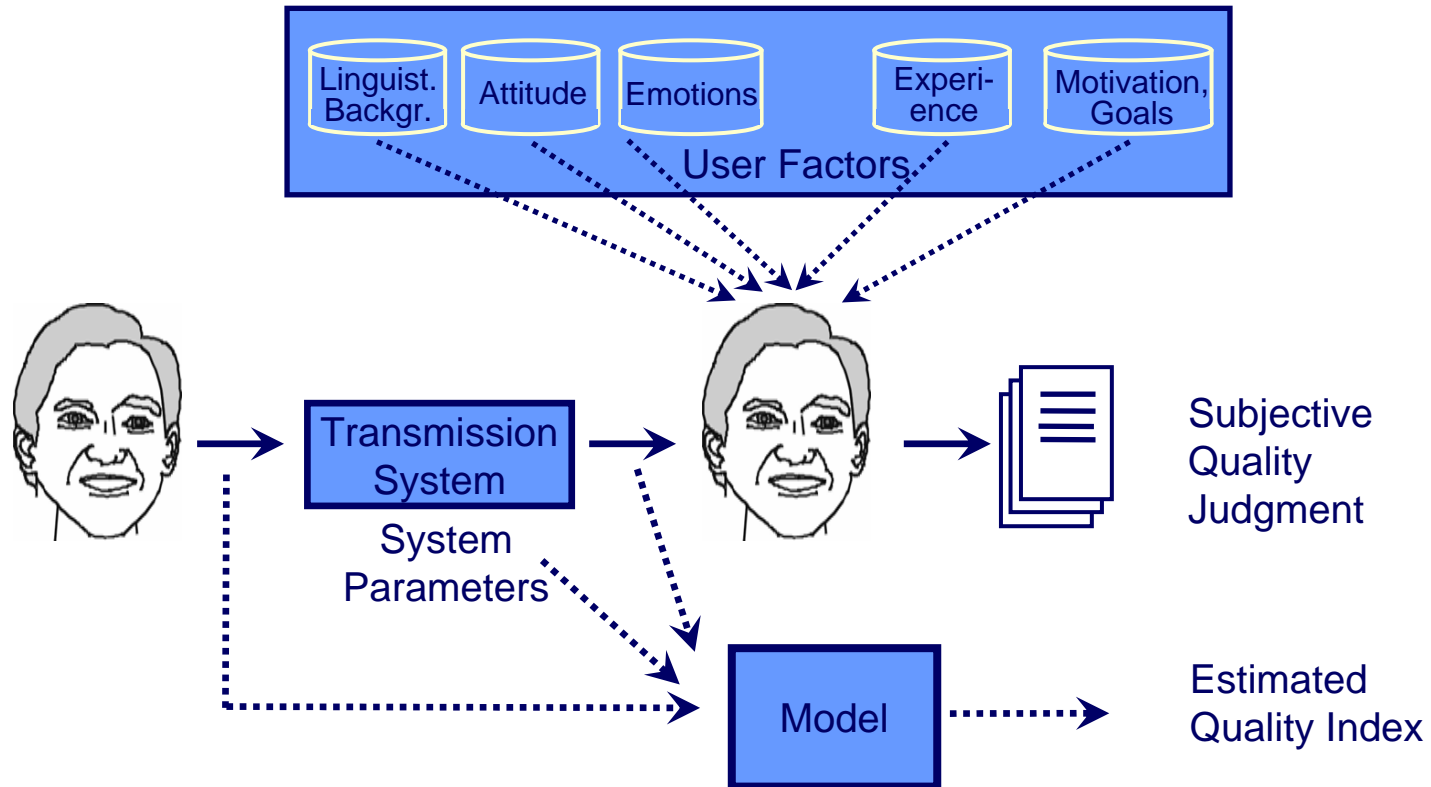
Sebastian Möller, Alexander Raake

Deutsche Telekom Labs, TU Berlin

Disclaimer

Information contained in this presentation
does not necessarily reflect the opinion of
and the recommendations given by the ITU-T!

Motivation

Quality evaluation and prediction:

User Factors: Linguist. Backgr., Attitude, Emotions, Experi-ence, Motivation, Goals

Transmission System

System Parameters

Model

Subjective Quality Judgment

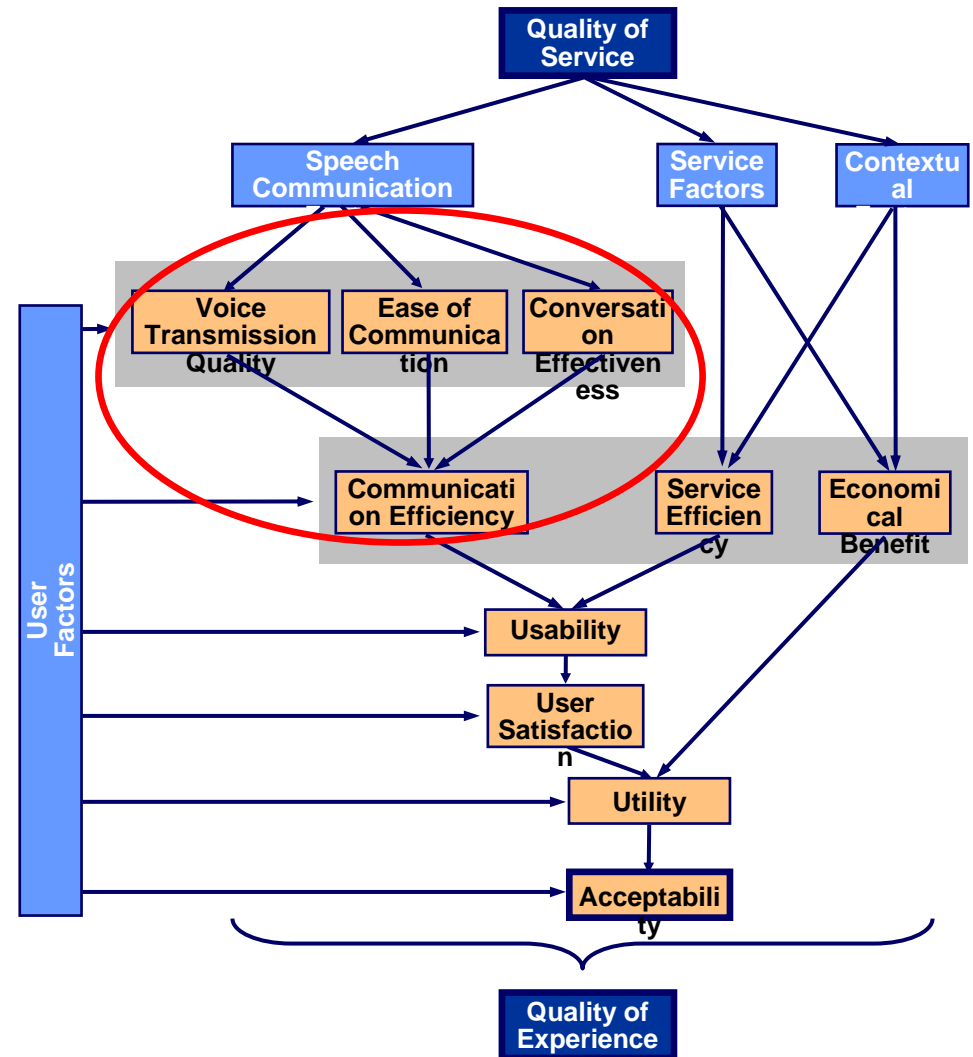Estimated Quality Index

Some Definitions (1)

Quality:

Result of judgment of the perceived composition of an entity with respect to its desired composition. (Jekosch, 2005)

- Quality of Service (Rec. E.800)
- Quality of Experience

Measurement:

o **Subjective** measurement:
Measurement involving human test subjects as measurement organs

o Objectivity:
Degree of interpersonal agreement
Here: Independence of the evaluator
Not: Independence of the test subject!

o **Instrumental** measurement:
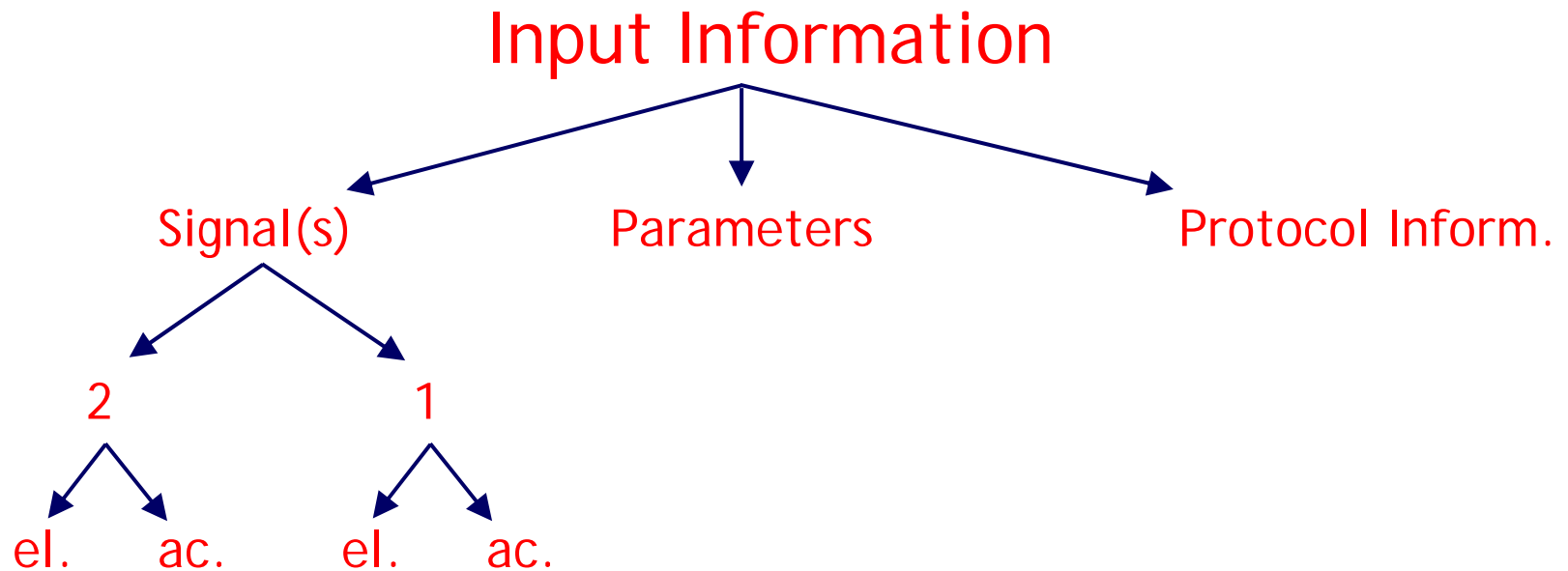Measurement using physical instruments and/or algorithms

ITU-T

Quality prediction:

o **Estimation** of quality, i.e. of subjective judgments which could have been given by an "average" test subject, e.g. on the basis of

- signals (acoustical or electrical)

- protocol information

- parameters

o All available "objective" models provide quality estimations (or predictions) only, **not** real measurements of quality!

Taxonomy (1)

Quality prediction models can be categorized according to:
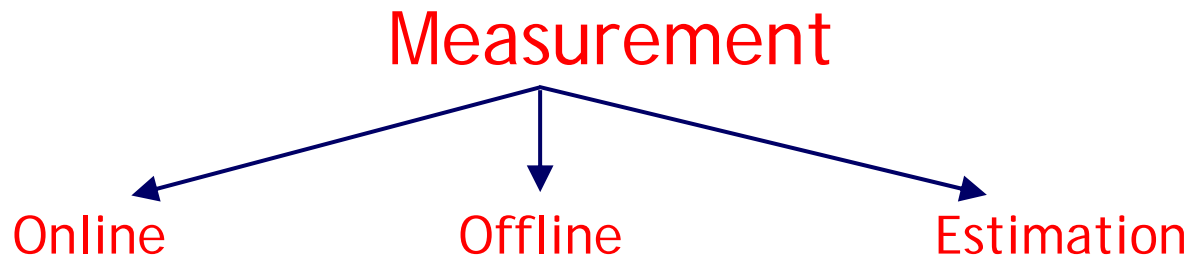
Input Information

Signal(s)   Parameters   Protocol Inform.

2   1

el.   ac.   el.   ac.

ITU-T

Quality prediction models can be categorized according to:

Measurement

Online   Offline   Estimation

Taxonomy (3)

Quality prediction models can be categorized according to:

Output Information

- **Listening**
  - Overall Quality
  - Quality Features

  MOS-LQ

- **Conversation**
  - Overall Quality
  - Quality Features

  MOS-CQ

- **Talking**
  - Overall Quality

  MOS-TQ
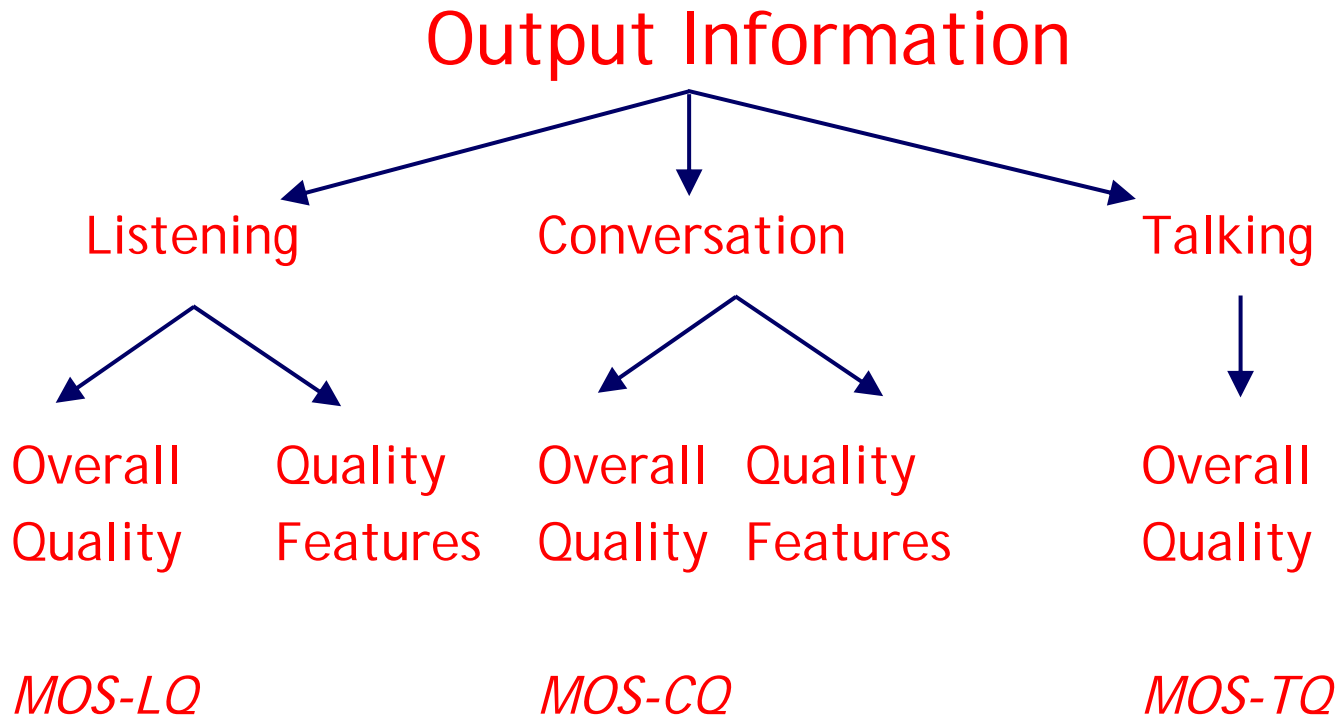
Taxonomy (4)

Quality prediction models can be categorized according to:

Application Area

Planning   Set-up   Monitoring
                    Optimization

NB   WB   NB   WB   NB   WB
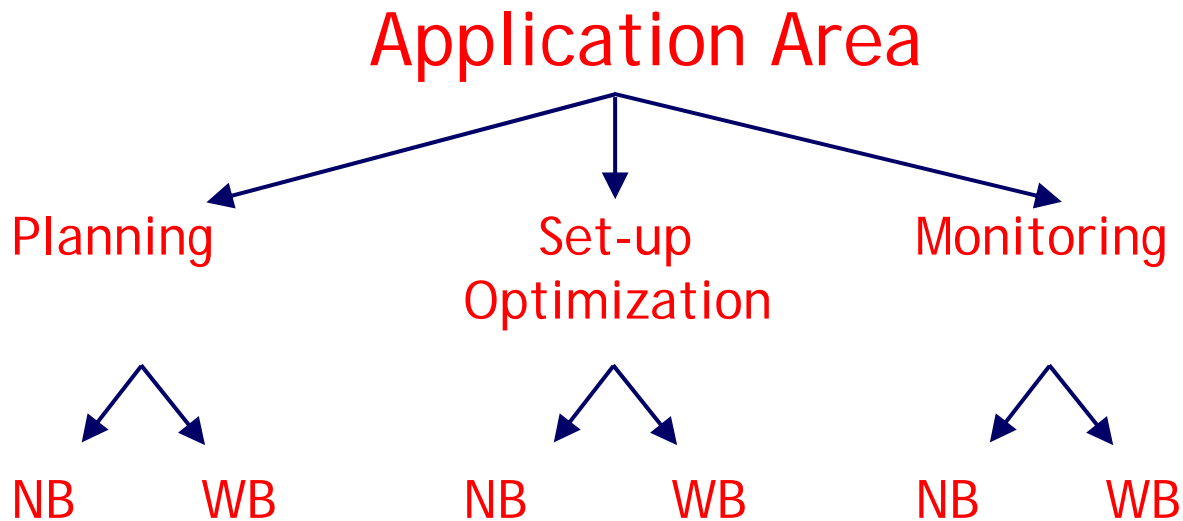
Available Models: NB

Output Information

		Listening		Conversation	
		Overall Quality	Quality Feat.	Overall Quality	Quality Feat.
Signals	1, el.	P.563	Psychoac. Measures	CCI (P.562)	
	1, ac.		Psychoac. Measures		
	2, el.	P.862 P.OLQA	[Heute05]	P.CQO?	P.CQO?
	2, ac.	P.OLQA	[STI] P.OLQA?		
Protocol	Meas.			P.VTQ? P.CQO?	P.CQO?
Parameters	Estim.			G.107	
	Meas.			NIEM (P.562)	

Input Information

Available Models: WB

Output Information

		Listening		Conversation	
		Overall Quality	Quality Feat.	Overall Quality	Quality Feat.
Signals	1, el.		Psychoac. Measures		
	1, ac.		Psychoac. Measures		
	2, el.	P.862.2 P.OLQA	[Heute05]		
	2, ac.	P.OLQA	[STI] P.OLQA?		
Protocol	Meas.				
Parameters	Estim.			WB-E-Model	
	Meas.				

Input Information

Application Areas

Planning:

o   Only estimated or offline-measured **parameters** are available as an input

- ➢ Parametric models (G.107, WB-E-Model)

Set-up / optimization:

o   (Simulated) **signals** are available

- ➢ P.862/P.862.2 for el2el MOS-LQ estimation
- ➢ P.OLQA for ac2ac MOS-LQ estimation

Monitoring:

o   Measured **signals**, **protocols** and **parameters** are available

- ➢ P.563, P.562, P.CQO and P.VTQ for NB non-intrusive monitoring
- ➢ P.862/P.862.2, P.CQO and P.OLQA for intrusive monitoring

Prediction Accuracy

General requirements:

o **Validity**: Do not try to predict what the model does not predict!

o **Reliability**: How accurate are the predictions?

Error sources:

o **Input information**

  ➢ Type of input information (signals, parameters)

  ➢ Measurement principle (online, offline, estimation)

  ➢ Accuracy of electrical/acoustical measurements

  ➢ Assumptions about unknown parts of the transmission channel

o **Model accuracy**

➢ As long as cross-model comparisons are missing, do not make assumptions about the superiority of a specific modeling approach!

Summary

Conclusions:

o According to the **taxonomy**, quality prediction models can be classified according to

   ➢ Input information

   ➢ Measurement principle

   ➢ Output information

   ➢ Application area

o All models estimate MOS-LQ or MOS-CQ, but they differ in

   ➢ the type and measurement principle of the available input information

   ➢ the application area

o Data on the accuracy of different approaches are still missing

Summary

Future work:

o Extend existing models towards wideband network scenarios

  ➢ P.862 → P.862.2 → P.OLQA

  ➢ G.107 → WB-E-Model

  ➢ P.CQO → ?

o Perform a thorough validity and reliability analysis

o Modeling approaches are still missing for

  ➢ Diagnostic information (except P.OLQA)

  ➢ Other audio signals (except P.OLQA, PEAQ)

  ➢ Other interaction scenarios and services

o Review ITU-T terminology and provide guidance document for quality prediction models

ITU-T

Quality prediction for dialogue systems:
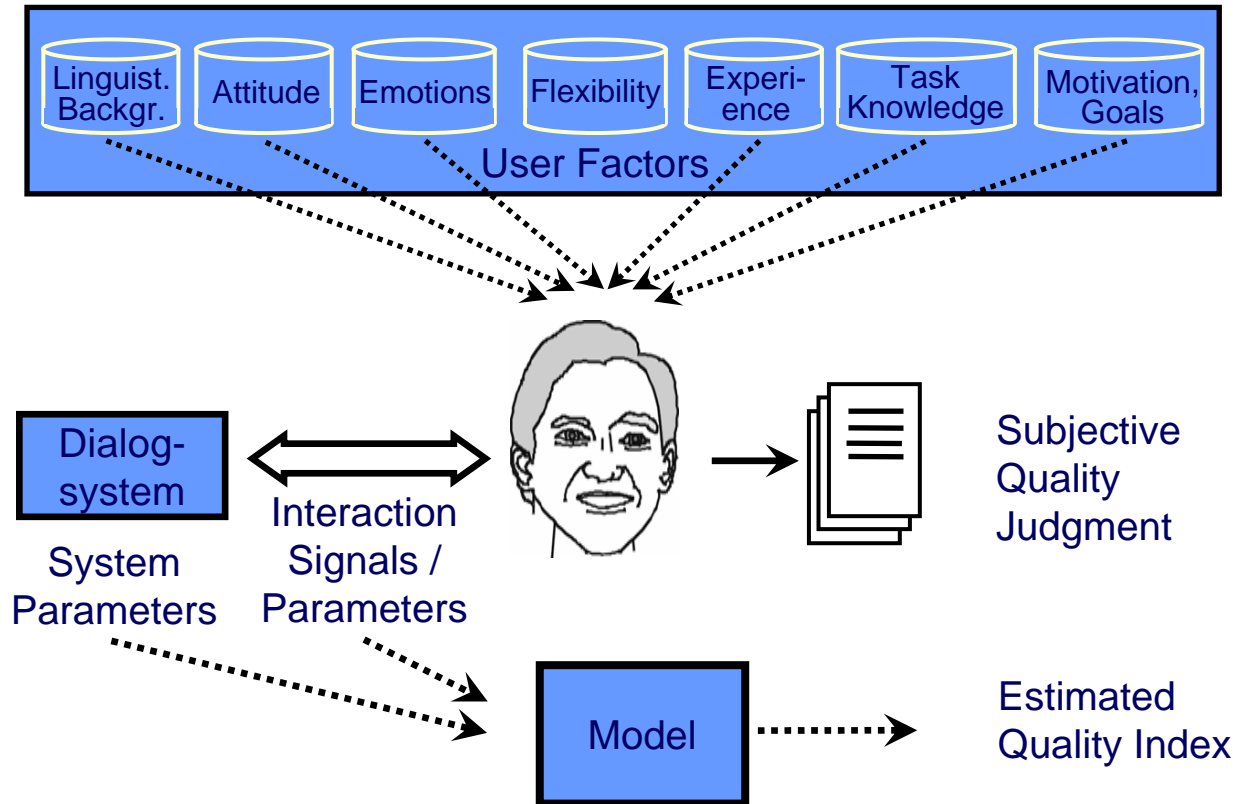
o Models for overall interaction quality?
o Models for text-to-speech quality?

References

ITU-T

Heute, U., Möller, S., Raake, A., Scholz, K., Wältermann, M. (2005). *Integral and Diagnostic Speech-Quality Measurement: State of the Art, Problems, and New Approaches*, in: Proc. 4th Europ. Congr. Acoustics (Forum Acusticum Budapest 2005), Budapest, 1695-1700.

IEC Standard 60268-16 (2003). *Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index*, IEC, Geneva.

ITU-T COM 12-26 (2006). *P.OLQA speech quality degradation decomposition benchmark proposal*, ITU-T, Geneva.

ITU-T D.075 (2005). *Requirements for a new model for objective speech quality assessment P.OLQA*, ITU-T SG12 meeting, 17-21 Oct., Geneva.

ITU-T Rec. E.800 (1994). *Terms and definitions related to quality of service and network performance including dependability*, ITU-T, Geneva.

ITU-T Rec. G.107 (2005). *The E-model, a computational model for use in transmission planning*, ITU-T, Geneva.

ITU-T Rec. P.562 (2004). *Analysis and interpretation of INMD voice-service measurements*, ITU-T, Geneva.

ITU-T Rec. P.563 (2004). *Single-ended method for objective speech quality assessment in narrow-band telephony applications*, ITU-T, Geneva.

ITU-T Rec. P.800.1 (2003). *Mean Opinion Score (MOS) terminology*, ITU-T, Geneva.

ITU-T Rec. P.862 (2001). *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, ITU-T, Geneva.

ITU-T Rec. P.862.2 (2005). *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, ITU-T, Geneva.

ITU-T TD 27(WP2/12) (2005). *Proposed scope for P.CQO*, ITU-T SG12 meeting, 17-21 Oct., Geneva.

Jekosch, U. (2005). *Voice and speech quality perception. Assessment and evaluation*, Springer, Berlin.

Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M. (2006). *Impairment factor framework for wideband speech codecs*, submitted to IEEE Trans. Audio, Speech and Language Processing.