International Telecommunication Union

Evaluation of In-Car Voice Services: Tasks for the New Q.12/12

Sebastian Möller

IKA, Ruhr-University Bochum, Germany

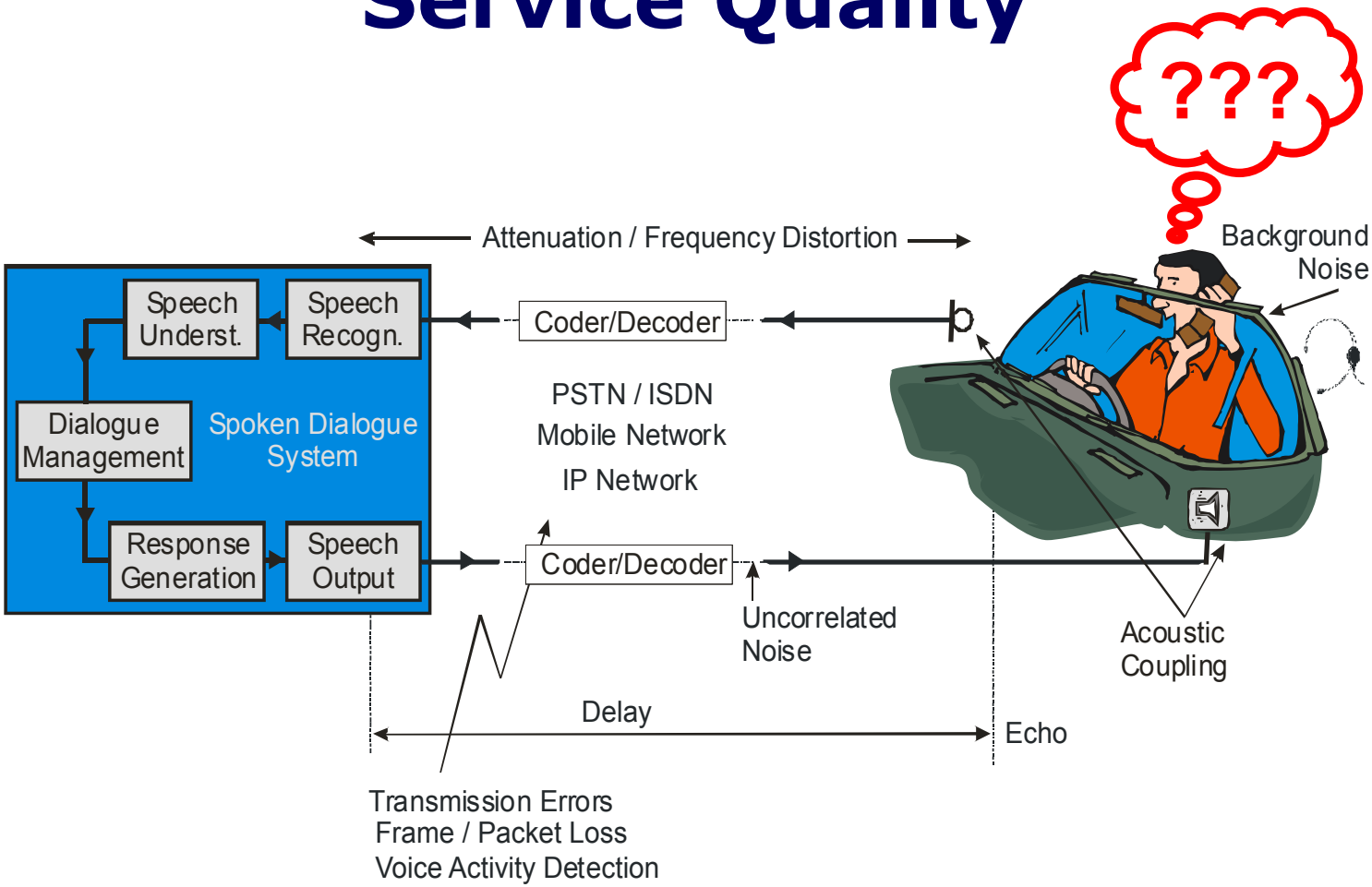
Co-Rapporteur Q.12/12

The Fully Networked Car, A Workshop on ICT in Vehicles
PALEXPO, Geneva, 2-4 March 2005

In-Car Voice Services

- **Rely on Speech Technology:**
  - Speech recognition (ASR), speaker recognition
  - Natural language understanding, dialogue management
  - Speech synthesis (TTS, CTS)
- **Application Examples:**
  - Voice dialling
  - Navigation
  - Office applications (dictation, email reading)
  - Information services
  - Control of remote devices, ...

Factors Influencing Service Quality



Technology Performance ↔ Perceived Quality

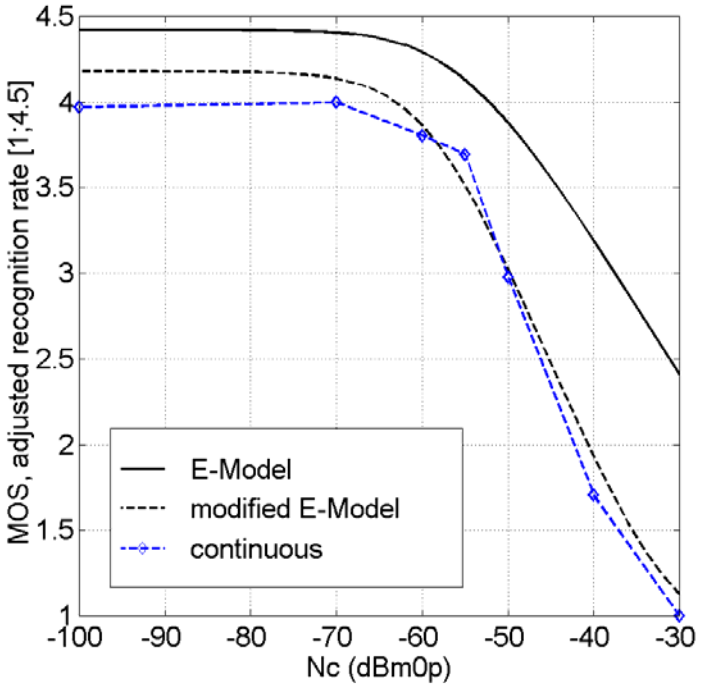
ITU-T

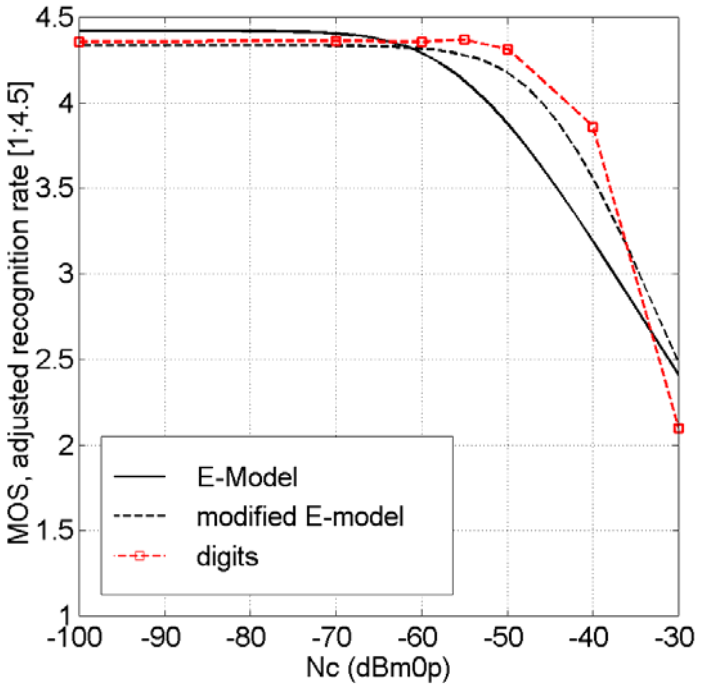
Q.12/12:
Performance Evaluation of Services Based on Speech Technology

- Technology Performance
  - Impact of acoustical conditions and transmission impairments on ASR performance
  - Measurement of speech synthesis performance
- Perceived Service Quality and Usability
  - Quality evaluation methods
  - Usability inspection methods
  - Quality prediction models

Transmission Channel Impact

ASR Performance under Circuit Noise

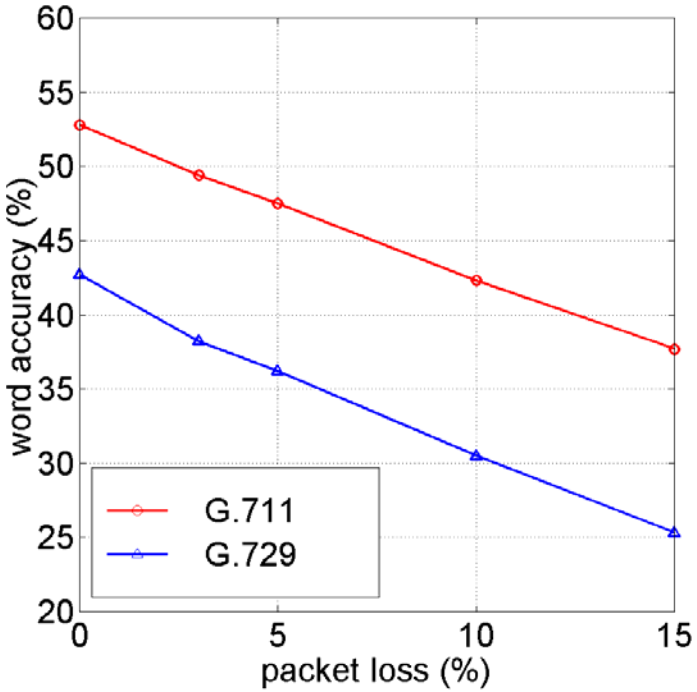


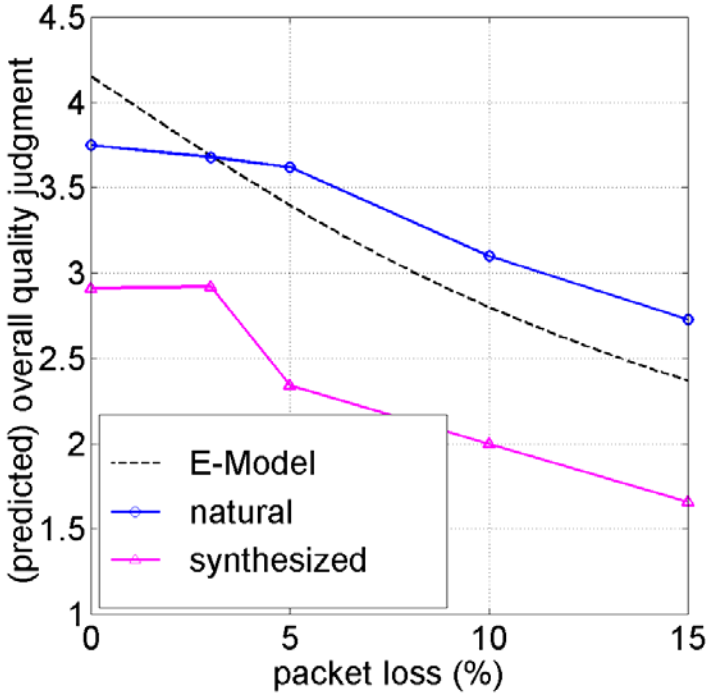


see ITU-T COM 12-14 (2001) and Möller (2005)

Transmission Channel Impact

ASR and TTS with IP Packet Loss



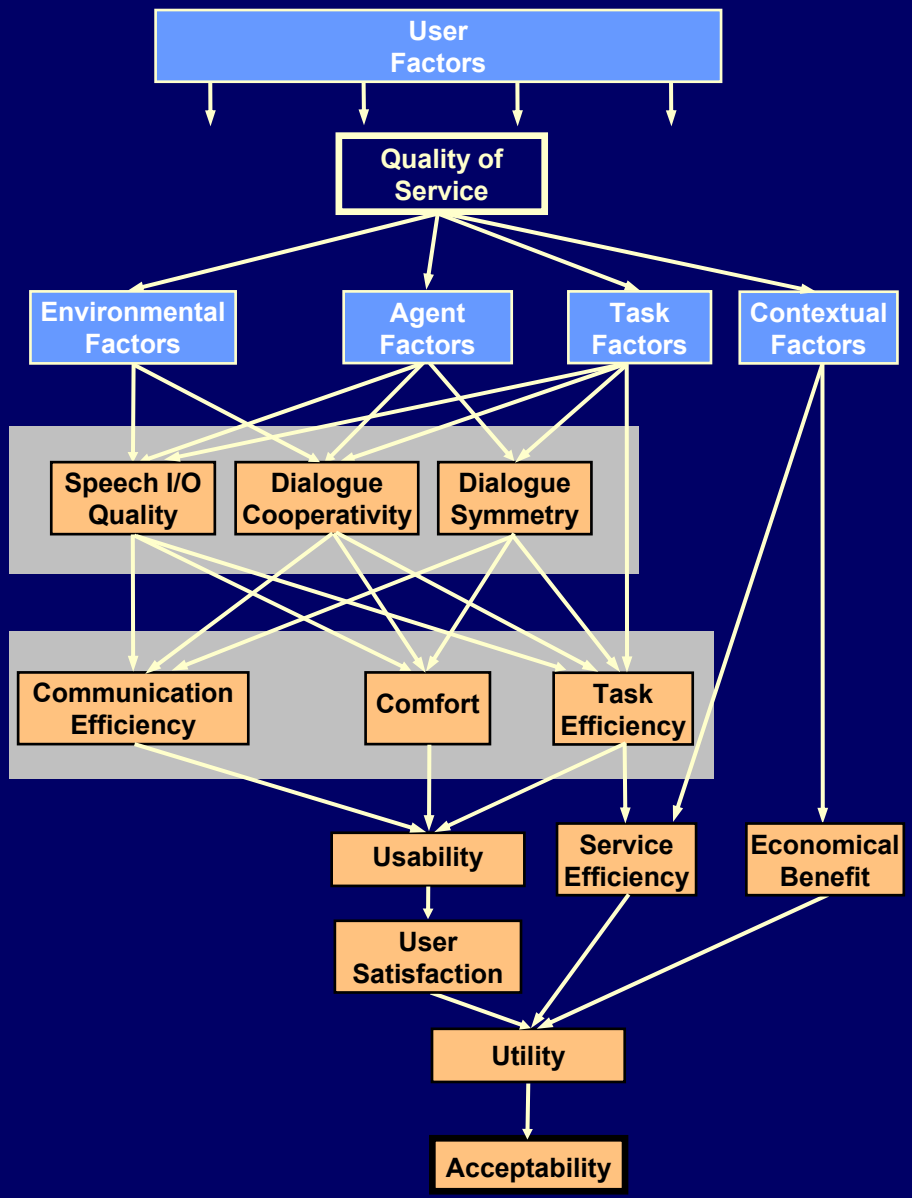


see Möller, Krebber & Raake (2004)

Predicting the Performance of Speech Technology

- ASR Performance
  - Adequate estimations may be obtained with (modified) network planning models, e.g. the E-model (Rec. G.107, 2003)
  - Other approaches: Signal-based models, e.g. Rec. P.862 (see Möller, 2005)
- TTS Quality
  - Can only be measured in auditory experiments (see Rec. P.85, 1994)
  - Prediction perhaps possible with single-ended models (e.g. Rec. P.563, 2004)?

ITU-T



Perceived Quality: Taxonomy of Quality Aspects

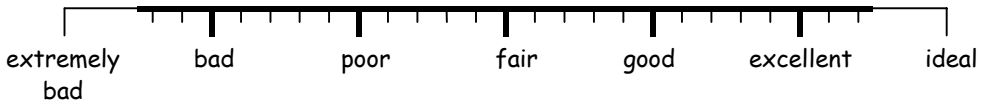
see Möller (2002), Rec. P.851 (2003)

dates

Subjective Evaluation of Service Quality

- Rec. P.851 (2003)
  - Definition of quality aspects and system/service characteristics
  - Description of experimental procedure (test set-up, subjects, scenarios, etc.)
  - Questionnaire example:

How well did you feel understood by the system?



  - First application examples have been reported

Usability and Acceptability Evaluation

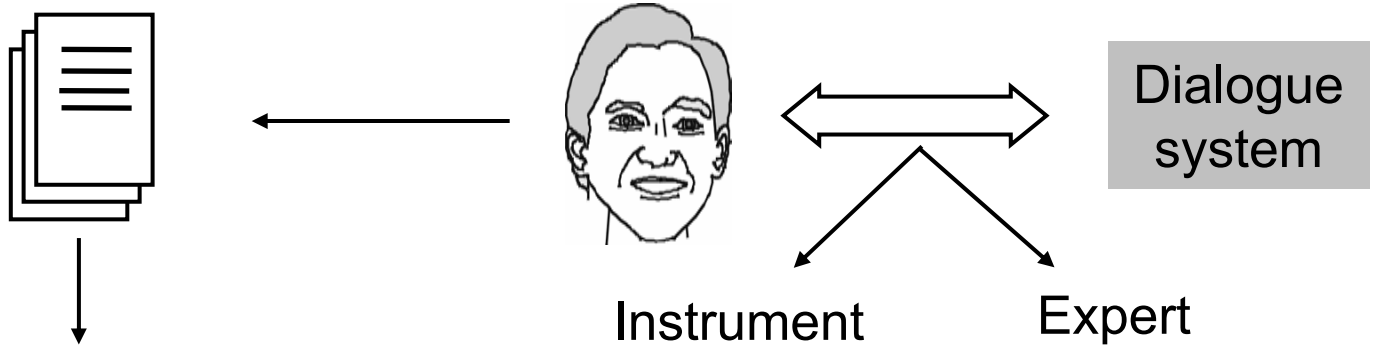
- Methods still need to be defined
  - Cognitive effects of the driving task?

Smeele et al. (2004)

  - New guidelines needed?

Service Quality Prediction

- Parametric Approach: PARADISE Model



$$user\ satisfaction = \alpha \cdot N(\kappa) - \sum_{i=1}^{n} w_i \cdot N(c_i)$$

Walker et al. (1997)

Q.12/12: Current Tasks

- Quantify impact of channel impairments and acoustic conditions on
  - ASR /speaker recognition performance
  - TTS quality
  - Overall service quality

  Prediction possible?
- Define measurement methods for
  - TTS quality:
    - Subjective method: Update Rec. P.85?
    - Instrumental method: Rec. P.563?
  - Service quality: Update of Rec. P.851?
  - Usability and Acceptability: Rec. P.QVS

Q.12/12: Current Tasks (2)

- Predict Perceived Quality
  - Define informative input parameters: Suppl. to P.85x Series (expected 10/2005), Rec. P.PST (expected 2006), see e.g. ITU-T D.030, 2005)
  - Modeling approach: PARADISE or non-linear algorithm?
- Several new Recommendations will be produced in 2005-2007

→ *Your contributions are very welcome!*

Please contact the Rapporteurs:

sebastian.moeller@rub.de
alexander.raake@limsi.fr

References

ITU-T Contribution COM 12-14 (2001). *Influence of the transmission channel on speech recognizer and spoken dialogue system performance*. Federal Republic of Germany (Author: S. Möller), ITU-T SG12 Meeting, 19-23 Feb. 2001, CH-Geneva.

ITU-T Delayed Contribution D.XXX (2005). *Proposal for parameters describing the performance of speech technology devices*. Federal Republic of Germany (Author: S. Möller), ITU-T SG12 Meeting, 18-27 January 2005, CH-Geneva.

ITU-T Rec. G.107 (2003). *The E-model, a computational model for use in transmission planning*. International Telecommunication Union, CH-Geneva.

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

ITU-T Rec. P.85 (1994). *A method for subjective performance assessment of the quality of speech voice output devices*. International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.851 (2003). *Subjective quality evaluation of telephone services based on spoken dialogue systems*. International Telecommunication Union, CH-Geneva.

Möller, S. (2005). *Quality of telephone-based spoken dialogue systems*. Springer, US-New York NY.

Möller, S. (2002). *A new taxonomy for the quality of telephone services based on spoken dialogue systems*. In: Proc. 3rd SIGdial Worksh. on Discourse and Dialogue, US-Philadelphia, 142-153.

Möller, S., Krebber, J., Raake, A. (2004). *Performance of speech recognition and synthesis in packet-based networks*. In: Proc. 8th Int. Conf. on Spoken Language Processing (Interspeech 2004 - ICSLP), KR-Jeju Island, Vol. 2, 1541-1544.

Smeele, P., Möller, S., Krebber, J. (2004). *Evaluation of the speech output of a smart-home system in a car environment*. In: Proc. 8th Int. Conf. on Spoken Language Processing (Interspeech 2004 - ICSLP), KR-Jeju Island, Vol. 3, 2221-2224.

Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A. (1997). *PARADISE: A framework for evaluating spoken dialogue agents*. In: Proc. 35th Meet. Assoc. Comp. Ling., ES-Madrid, 271-280.