

# Endgeräte im E-Modell

Sebastian Möller<sup>1</sup>, Frank Kettler<sup>2</sup>, Hans-Wilhelm Gierlich<sup>2</sup>,  
Nicolas Côté<sup>3</sup>, Alexander Raake<sup>4</sup>, Marcel Wältermann<sup>1</sup>

<sup>1</sup> *Quality and Usability Lab, Deutsche Telekom Labs, TU Berlin, E-Mail: vorname.nachname@telekom.de*

<sup>2</sup> *HEAD Acoustics GmbH, Herzogenrath, E-Mail: {h.w.gierlich, frank.kettler}@head-acoustics.de*

<sup>3</sup> *Université de Bretagne Occidentale, Plouzané, Frankreich, E-Mail: nicolas.cote@univ-brest.fr*

<sup>4</sup> *Assessment of IP-based Applications, Deutsche Telekom Labs, TU Berlin, E-Mail alexander.raake@telekom.de*

## Einleitung

Das E-Modell in ITU-T Empf. G.107 [1] stellt ein Planungswerkzeug zur parametrischen Vorhersage der Sprachqualität schmalbandiger Telefonnetze dar. Dabei wird für die gesamte Übertragungsstrecke Mund-zu-Ohr ein Qualitätswert geschätzt, der neben den Netzwerkeinflüssen ganz erheblich von den verwendeten Endgeräten bestimmt wird. Endgeräte werden bislang aber nur in Form sogenannter *Loudness Ratings* berücksichtigt, welche für Handapparate definiert sind; andere Endgeräte wie Headsets oder Freisprecher mit der ihnen inhärenten Signalverarbeitung werden bislang nicht betrachtet.

In diesem Beitrag sollen erste Schritte erläutert werden, Endgeräte adäquat im E-Modell zu berücksichtigen. Der Fokus liegt zunächst auf der Störgeräuschunterdrückung, deren Einflüsse durch 2-3 neue Parameter deutlich besser vorhergesagt werden können, wie experimentelle Daten belegen. Im Ausblick werden erste Vorschläge zur Modellierung weiterer Effekte in zukünftigen Erweiterungen der ITU-Empfehlung gemacht.

## E-Modell

Das E-Modell liefert eine Vorhersage der Sprachqualität in einer Kommunikationssituation für den gesamten Übertragungspfad Mund-zu-Ohr. Dabei werden einzelne Störungen auf eine Beeinträchtigungs-Skala (R-Skala) umgerechnet, auf der sich die perzeptiven Effekte additiv überlagern sollen. Die so transformierten Beeinträchtigungen (*Impairment Factors*) werden vom maximal erreichbaren Störabstand  $R_0$  (ebenfalls transformiert auf die R-Skala) abgezogen, um einen Qualitätsschätzwert  $R$  für die gesamte Übertragung zu berechnen. Durch Endgeräte bedingte Störungen werden bislang wie folgt berücksichtigt:

- $R_0$  berechnet sich aus dem sende- und empfangsseitigen Hintergrundgeräusch sowie Leitungsrauschen, deren Pegel jeweils auf die R-Skala transformiert werden. Bei der Berechnung des sendeseitigen Anteils  $Nos$  wird der A-gewichtete Pegel durch die *Loudness Ratings* des Handapparates  $SLR$  (sendeseitig) und  $RLR$  (empfangsseitig) sowie durch die gewichtete Differenz zwischen den Sensitivitäten für Direktschall vs. Diffusschall  $Ds$  abgeschwächt:

$$Nos = Ps - SLR - Ds - 100 + 0.004(Ps - SLR - RLR - Ds - 14)^2 \quad (1)$$

Eine Störgeräuschunterdrückung ist dabei nicht vorgesehen.

- Akustische wie auch Netzwerk-Echos werden über ihr *Loudness Rating* ( $TELR$  für Sprecher-Echos,  $WEPL$  für Hörer-Echos) und die zugehörigen Laufzeiten berücksichtigt; spezielle Effekte von Echo-Unterdrückern oder Pegelwaagen werden nicht erfasst.
- Codiervverzerrungen werden über einen speziellen *Impairment Factor*  $Ie,eff$  erfasst, welcher Codec-spezifisch ist und aus auditiven Tests sowie instrumentellen Maßen berechnet werden kann.
- Der Einfluss von Signalverzögerungen wird über den *Impairment Factor*  $Id$  erfasst, welcher aus der Verzögerungszeit  $Ta$  sowie dem Echo (s.o.) berechnet wird.

Im Folgenden zeigen wir, wie durch geringfügige Modifikationen die wichtigsten Effekte von Störgeräuschreduktionen erfasst werden können.

## Störgeräuschreduktion im E-Modell

Störgeräuschreduktionen produzieren vor allem zwei Arten von Artefakten: Verbleibendes Störgeräusch sowie Veränderungen des bereinigten Sprachsignals. Diese können wie folgt erfasst werden:

- Verbleibendes Störgeräusch tritt während der Sprachaktivität wie auch in den Pausen auf. Die in ITU-T Empf. G.160 [2] definierten Parameter  $SNRI$  (*SNR Improvement during speech*, in dB) und  $TNLR$  (*Total Noise Level Reduction*, in dB) zeigen die jeweilige Verbesserung an. Wir nehmen vereinfacht eine jeweils hälftige Gewichtung von Sprachaktivität und Pausen an und erweitern Formel (1) wie folgt:

$$Nos = Ps - SLR - Ds - 0.5(SNRI - TNLR) - 100 + 0.004(Ps - SLR - RLR - Ds - 14)^2 \quad (2)$$

- Die Veränderungen des Sprachsignals werden in auditiven Tests nach ITU-T Empf. P.835 [3] in Form eines S-MOS-Wertes bewertet. Hierzu gibt es auch ein instrumentelles Maß, welches in ETSI EG 202 396-3 [4] beschrieben ist. Durch Vergleich der MOS-Werte für den ungestörten und den um das Geräusch bereinigten Fall und Transformation auf die R-Skala kann ein *Impairment Factor*  $Inr$  berechnet werden, welcher die Beeinträchtigung des Sprachsignals auf der R-Skala widerspiegelt; dieser Wert wird ebenfalls von  $R_0$  abgezogen. Die genaue Prozedur ist in [5] beschrieben.

## Experimentelle Überprüfung

Zur Überprüfung des Ansatzes wurde ein auditiver Test nach ITU-T Empf. P.835 durchgeführt, und die dabei

verwendeten Stimuli wurden zur Bestimmung der Eingangsparameter des E-Modells verwendet. Die Stimuli wurden mit 7 verschiedenen mobilen Handapparaten in störgeräuschbehafteter Umgebung mit einem Kunstkopf HEAD acoustics Typ HMS II.3 gemäß ITU-T Empf. P.58 und ITU-T Empf. P.310 und einer Pinna nach Typ 3.4 gemäß ITU-T Empf. P.57 aufgezeichnet. 8 englische Sätze von 4 Sprechern (2m, 2w) wurden mit einem aktiven Sprachpegel von -1.7 dBPa am Mundreferenzpunkt reproduziert; dieser Pegel ist 3 dB höher als normal, um den Lombard-Effekt anzunähern. Der Aufbau zur Reproduktion von Hintergrundgeräuschen folgte ETSI EG 202 396-1, wobei 5 verschiedene Geräuschtypen (Auto, Straße, Cafeteria, Kreuzung, Büro) zum Einsatz kamen.

Die verrauschten Sprachstimuli wurden von 24 Hörern in ruhiger Umgebung gemäß ITU-T Empf. P.835 bewertet. Hierzu wurden sie zunächst IRS-gefiltert und dann mit einem aktiven Sprachpegel von -21 dBPa (entspricht 73 dB) diotisch über Kopfhörer präsentiert. Hieraus ergaben sich der o.a. S-MOS sowie die Bewertung der Gesamtqualität (G-MOS), welche als Zielgröße für die Validierung dient.

Die Eingangsparameter des E-Modells wurden auf Standardwerte gemäß [1] gesetzt, mit Ausnahme des Hintergrundgeräuschpegels (gemessen), der *Loudness Ratings SLR* (gemessen) und *RLR* (10 dB – *SLR*), *I<sub>e,eff</sub>* (20 für 2G-Handsets, 3 für 3G-Handsets), und des Wertes für *D<sub>s</sub>* (Default-Wert *D<sub>s</sub>* = 3 und zwei gemessene Varianten mit rosa Rauschen bzw. Cafeteria-Geräusch). Die Parameterwerte für *SNRI* und *TNLR* wurden ebenfalls mit allen Handapparaten und Hintergrundgeräuschen gemessen. Zur Bestimmung von *Inr* wurden der S-MOS-Wert sowie eine TOSQA-Schätzung für den ungestörten Fall herangezogen, da die unverrauschten Sprachstimuli leider nicht im auditiven Test bewertet worden waren.

**Tabelle 1:** Korrelationen (*r*) und Vorhersagefehler (*RMSE*)

E-Modell-Version	Ergebnis	
	<i>r</i>	<i>RMSE</i>
a) original	0,62	0,95
b) <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s</sub></i> = 3	0,70	0,72
c) <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s,pink</sub></i>	0,78	0,88
d) <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s,cafe</sub></i>	0,71	1,65
e) <i>Inr</i>	0,72	1,12
f) <i>Inr</i> , <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s</sub></i> = 3	0,82	0,85
g) <i>Inr</i> , <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s,pink</sub></i>	0,86	1,01
h) <i>Inr</i> , <i>SNRI</i> und <i>TNLR</i> , <i>D<sub>s,cafe</sub></i>	0,73	1,71

Die Korrelationen zu den auditiven Testergebnissen sowie die Vorhersagefehler sind in Tabelle 1 zusammengestellt. Durch Einschluss der Parameter *SNRI* und *TNLR* lässt sich die Korrelation von 0,62 auf 0,78 steigern. Hierbei ist es allerdings von Bedeutung, wie der Wert für *D<sub>s</sub>* bestimmt wird; eine Messung mit rosa Rauschen ergibt die zuverlässigsten Schätzwerte. Bei weiterer Berücksichtigung von *Inr* steigt die Korrelation sogar auf 0,86, allerdings auf Kosten eines leicht erhöhten Vorhersagefehlers.

## Ausblick

Die Berücksichtigung von Geräuschreduktionsverfahren stellt nur einen ersten Schritt zur Modellierung der Effekte von Endgeräten im E-Modell dar. Daneben sind folgende Erweiterungen wünschenswert, welche bislang allerdings noch nicht evaluiert werden konnten:

- Klangverfärbungen durch die akustischen Übertragungswege: Diese könnten durch einen Bandbreiten-Impairment-Faktor berücksichtigt werden, wie er in [6] beschrieben ist. Insbesondere beim Freisprechen werden Klangverfärbungen maßgeblich für die erzielbare Sprachqualität sein.
- Echo-Unterdrücker: Die zusätzliche Echounterdrückung kann mittels *TELR* im E-Modell bereits erfasst werden. Störungen der bereinigten Sprache lassen sich wie bei *Inr* durch einen zusätzlichen *Impairment Factor I<sub>ec</sub>* berücksichtigen, wobei anstelle des ETSI-Modells [4] adäquatere Modelle zum Einsatz kommen sollten, welche allerdings zunächst validiert werden müssen.
- Beeinträchtigung der Konversationstauglichkeit: Diese ist insbes. bei Pegelwaagen oder vergleichbaren Regelungen stark eingeschränkt. Für eine hohe Konversationstauglichkeit muss die Sprachqualität in allen Gesprächssituationen (Hören, Sprechen, Gegensprechen, Konversation) einen Mindestwert aufweisen; genauere Untersuchungen dieser Zusammenhänge stehen aber noch aus.

Zur Validierung der vorgeschlagenen Ansätze sind umfangreiche Datensätze notwendig, zu deren Sammlung die DEGA Unterstützung bieten könnte. Die Ergebnisse könnten dann in einer Überarbeitung des E-Modells münden, welche für 2012 geplant ist.

## Literatur

- [1] ITU-T Rec. G.107, “The E-Model, a Computational Model for Use in Transmission Planning”, International Telecommunication Union, Genf, 2009
- [2] ITU-T Rec. G.160, “Voice Enhancement Devices”, International Telecommunication Union, Genf, 2008
- [3] ITU-T Rec. P.835, “Subjective Test Methodology for Evaluating Speech Communication Systems That Include Noise Suppression Algorithm”, International Telecommunication Union, Genf, 2003
- [4] ETSI EG 202 396-3, “Speech Quality Performance in the Presence of Background Noise – Part 3: Background Noise Transmission – Objective Test Methods”, European Telecommunications Standards Institute, Sophia Antipolis, 2007
- [5] Möller, S., Kettler, F., Gierlich, H.-W., Côté, N., Raake, A., Wältermann, M., “Extending the E-Model to Better Capture Terminal Effects”, in: Proc. 3rd Int. Workshop on Perceptual Quality of Systems (PQS 2010), Bautzen, 6-8 Sept. 2010
- [6] Wältermann, M., Raake, A., “Towards a New E-Model Impairment Factor for Linear Distortion of Narrowband and Wideband Speech Transmission”, in: Proc. ICASSP 2008, Las Vegas NV, 4817-4820, 2008