

## Wideband E-Modell

Alexander Raake, Sebastian Möller, Marcel Wältermann

Deutsche Telekom Laboratories, TU Berlin, Email: alexander.raake@telekom.de

### Einleitung

Ausgehend von bisherigen Arbeiten zur Modellierung der Sprachqualität von breitbandiger Telefonsprache (wideband, WB; 50-7000 Hz) stellen wir in diesem Beitrag ein erstes vollständiges WB E-Modell vor. Beim E-Modell handelt es sich um einen parameterbasierten Modellansatz zur Vorhersage der Sprachqualität im Rahmen der Netzwerkplanung. Bislang war ein allgemeiner Einsatz des E-Modells (ITU-T Rec. G.107) auf schmalbandige Telefonnetze beschränkt (narrowband, NB; 300-3400 Hz). In eigenen Vorarbeiten haben wir den Qualitätsvorteil von Breitband über Schmalband auf etwa 30% beziffert, und damit die Qualitätsskala des E-Modells von 0-100 auf 0-129 erweitert [2, 3]. In anschließenden Arbeiten haben wir eine WB-Version des E-Modells für die reine Hörsituation entwickelt, die den Qualitätseinfluss unterschiedlicher Beeinträchtigungsfaktoren umfasst [1]: (1) Lineare Verzerrungen aufgrund unterschiedlicher Übertragungsbandpässe oder Endgeräte, (2) NB- und WB-Sprachkodierer, (3) VoIP Paketverluste, (4) Rauschen auf der Leitung oder im Senderraum. Auf Basis von Konversations-tests erweitern wir das Modell im vorliegenden Beitrag um den Einfluss von Echo und Verzögerungen, also Störungen, die sich nur in einer Konversationsituation auswirken. Das so entstehende Gesamtmodell stellt die erste Version eines weitgehend vollständigen WB-E-Modells dar.

### Hörsituation

Das zunächst von uns vorgestellte "Wideband", d.h. auf Breitband-Sprache erweiterte E-Modell gilt für Störungen wie sie bereits in der Hörsituation wahrgenommen werden [1]. Als Eingangsgrößen des E-Modells dienen die technischen Faktoren die eine Telefonverbindung beeinträchtigen können, die in Form von technischen Parametern ausgedrückt werden. Das E-Modell basiert auf der Annahme, dass diese Parameter auf eine psychologisch-motivierte Qualitätsskala übersetzt und gemäß der Art der Beeinträchtigung zu *Impairmentfaktoren* zusammengefasst werden können, und dass sich diese auf der E-Modell Skala additiv verhalten, siehe Gl. 1, die Grundgleichung für WB-Fall.

$$R = R_{0,WB} - I_{SWB} - I_{dWB} - I_{e,effWB} + A. \quad (1)$$

$R$  ist der Ausgabewert des E-Modells auf der sog. *Transmission Rating Scale*, die für NB von 0-100, für WB von 0-129 reicht. Dabei ist  $R_{0,WB}$  das *Basic Signal-to-Noise Ratio*, d.h. die Basisqualität aufgrund der Pegelverhältnisse aus übertragener Sprache und dem auf der Leitung vorliegenden Gesamttrauschen (Leitungsrauschen und auf Leitungspegel transformierte Geräuschpegel für Sende- und Empfangsseite).  $I_{SWB}$  ist der *Simultaneous*

Set	Ta	T	TELR	BW	
	[ms]	[ms]	[dB]		
Set 1	0	0	100	NB	Group 1
	0	20	15	WB	
	0	100	15	WB	
	0	100	25	NB	
	0	100	35	NB	
	0	100	45	NB	
Set 2	0	0	100	WB	Group 1
	0	20	25	WB	
	0	20	25	WB	
	0	50	15	WB	
	0	50	25	WB	
	0	400	25	WB	
Set 3	0	0	100	WB	Group 2
	0	50	35	WB	
	0	100	15	NB	
	0	100	25	WB	
	0	100	35	WB	
	0	100	45	WB	
Set 4	200	0	100	WB	Group 2
	200	0	100	NB	
	400	0	100	WB	
	400	0	100	NB	
	800	0	100	WB	
	800	0	100	NB	

**Abbildung 1:** Verwendetes Block-Design des Konversations-tests mit Angabe der Versuchsbedingungen pro Block und damit Versuchsgruppe.  $T_a$  ist die Einweg-Verzögerung,  $T$  die Echo-Laufzeit (Einweg),  $TELR$  das Talker Echo Loudness Rating, d.h. die gewichtete Abschwächung des Echsignals und  $BW$  die jeweils verwendete Bandbreite.

*Impairment Factor*, der sprach-simultane Störungen wie signalkorreliertes Rauschen und zu laute Sprachpegel umfasst. Der *Delayed Impairment Factor*  $I_{dWB}$  betrifft alle gegenüber dem Sprachsignal "verzögerten" Störungen, wie Echo und Delay. Der *Effective Equipment Impairment Factor*  $I_{e,effWB}$  beinhaltet die Beeinträchtigung durch Kodierung sowie durch VoIP Paketverluste. Der Advantage Factor  $A$  erlaubt es, die z.B. beim Mobilfunk durch die Möglichkeit der mobilen Nutzung höhere Toleranzschwelle bei Störungen in den Modellvorhersagen abzudecken. In [1] wurden Modellansätze gezeigt für  $R_{0,WB}$  und  $I_{e,effWB}$ , aufgrund der unzureichenden Datenlage noch nicht für  $I_{SWB}$ . Im vorliegenden Beitrag wird  $I_{dWB}$  betrachtet.

### Konversationsituation

Im Rahmen dieser Studie wurde ein als Block-Designs konzipierter Konversations-test durchgeführt. Die insgesamt vier Blöcke wurden dabei für jede Versuchsgruppe zu einem Set aus zwei Blöcken (Sets 1–4) kombiniert, gemäß Abb. 1. Die Konversationen wurden mittels der in [4] beschriebenen *Short Conversation Test* Szenarien durchgeführt. Als Endgeräte wurden Sennheiser HMD 410 Headsets eingesetzt. Jede der vier Gruppen bestand aus sechs Versuchspaaren à zwei Teilnehmern, mit insgesamt  $4 \cdot 6 \cdot 2 = 48$  Versuchspersonen. Jedes der 4 Sets wurde demnach von 24 Versuchspersonen (VPen) bewertet (s. Abb. 1), die beiden NB und WB Referenzbedingungen (ohne Echo und Delay) durch alle 48 VPen.

Die beiden Gesprächspartner eines Durchlaufs befanden sich in zwei akustisch und visuell getrennten Räumen gemäß ITU-T Rec. P.800 (1996). Die randomisierte Rei-

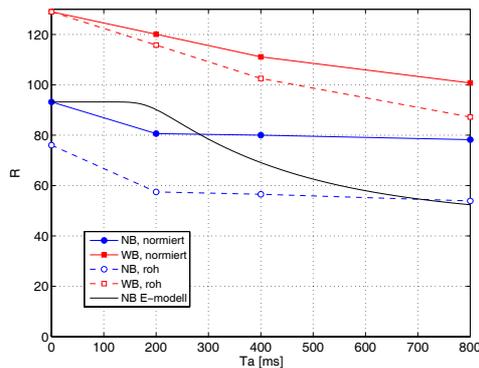


Abbildung 2: Qualität in Abhängigkeit von der Verzögerung  $T_a$  für den echofreien Fall.

henfolge der Gespräche war für jedes VPen-Paar eine andere. Nach jedem Gespräch gaben die VPen eine Bewertung der Gesamtqualität auf der 5-Punkte *Absolute Category Rating*-Skala ab, der sog. MOS-Skala (*Mean Opinion Score*, ITU-T Rec. P.800, 1996). Die auf der ACR/MOS-Skala gewonnenen mittleren Beurteilungen wurden zunächst wie in [1] beschrieben auf die WB-erweiterte E-Modell *R*-Skala transformiert. In einem zweiten Schritt werden die Daten dann üblicherweise anhand von Referenzverbindungen mit bekanntem *R*-Wert normiert. Da sich ein nichtlinearer Bias in den Ergebnissen zeigte, der echo-freie Verbindungen anders betraf als echobehaftete, werden im Folgenden jeweils die Werte mit Normierung auf die für die beiden Referenzverbindungen erwarteten Werte (1) und einmal ohne Normierung betrachtet (2), wobei das Hauptaugenmerk auf den Vergleich NB zu WB gelegt wird.

In Abb. 2 sind die Versuchsergebnisse für echofreie Verbindungen und Verzögerungen  $T_a \in [0, 200, 400, 800]$  ms den Vorhersagen des NB E-Modells gegenübergestellt. Offenbar sagt das E-Modell einen stärkeren Qualitätsabfall vorher als im Konversationstest gemessen. Dies steht im Einklang mit diversen anderen vergleichbaren Studien, wo ein bedeutender Qualitätseinfluss nur für besonders interaktive Konversationen gemessen werden konnte (siehe [2] für Beispiele). Allerdings ist der Abfall der Qualität mit steigendem  $T_a$  für WB ausgeprägter als für NB, was eine höhere Empfindlichkeit von Nutzern für Delay impliziert. Der Abfall der NB E-Modell Kurve geht jedoch über das für den schlechtesten Testfall beobachtete Verhalten hinaus (nicht-normierte WB-Ergebnisse). Bislang wurden die pessimistischen Vorhersagen des Qualitätseinflusses von Verzögerung  $T_a$  seitens der ITU-T akzeptiert, um Netze auch für interaktive Konversationen mit ausreichend hoher Qualität zu planen. Bei Entscheidungen für oder wider einzelne Netzelemente wie Kodierer muss der Einfluss entsprechender Verzögerungswerte jedoch gegen den Gewinn an "Hörqualität" abgewogen werden. In diesem Kontext werden die pessimistischen Vorhersagen des E-Modells derzeit kritisch diskutiert. Die hier beschriebenen Ergebnisse zeigen jedoch, dass zusätzliche Studien unerlässlich sind, um sicher zu stellen, dass interaktive Konversationen über breitbandige Verbindungen nicht zu einer eher noch kritischeren Qualitätsbeurteilung führen als derzeit vom E-Modell vorhergesagt. Auf Basis der hier gezeig-

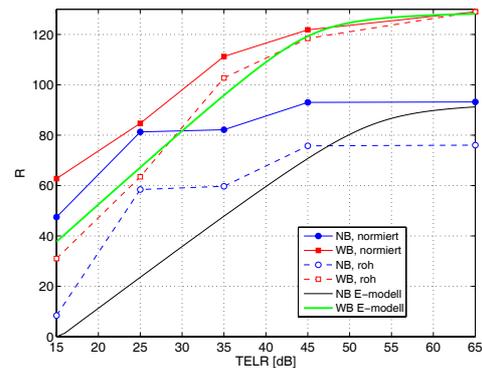


Abbildung 3: Qualität in Abhängigkeit von der Abschwächung des Sprecherechos, für Einweg-Echolaufzeit  $T = 100$  ms und Vorhersagen mittels NB und hier vorgeschlagenen WB E-Modells.

ten Daten kann für den echo-freie Fall zunächst gesetzt werden:  $Id, WB = Id, NB$ .

Im Falle von Sprecherechos zeigt sich für WB ein stärkerer Abfall der Qualität mit abnehmender Echoabschwächung (*TELR*, *Talker Echo Loudness Rating*) als für NB (Abb. 3). Die zusätzliche lineare Verschiebung zwischen den Kurven lässt sich durch einen entsprechenden *TELR*-Offset von  $\Delta TELR = 18$  dB in der betroffenen E-Modell Gleichung nachbilden. Abb. 3 zeigt die resultierenden Modellvorhersagen und die gute Übereinstimmung mit den Testergebnissen.

In weiterführenden Studien müssen die Ergebnisse validiert werden, bevor das Modell standardisiert werden kann. Zudem muss untersucht werden, ob Nutzer von WB-Verbindungen grundsätzlich empfindlicher auf Verzögerungen reagieren, wie die vorliegende Studie impliziert, oder Verzögerung in unserem Test bei NB-Verbindungen nur weniger auffiel, da die VPen ihr Qualitätsurteil auf den deutlicheren Unterschied zwischen NB und WB stützten.

## Danksagung

Die Autoren danken Hr. Thomas Heinrich für die Durchführung der Konversationstests.

## Literatur

- [1] Raake, A., Möller, S., Wältermann, M., Côté, N., Ramirez, J.-P.: Parameter-based Prediction of Speech Quality in Listening Context – towards a WB E-model. In: Proc. IEEE QoMEX (2010).
- [2] Raake, A.: Speech Quality of VoIP – Assessment and Prediction, John Wiley & Sons, GB-Chichester (2006).
- [3] Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M.: Impairment Factor Framework for Wideband Speech Codecs. IEEE Trans. Audio, Speech, Lang. Proc. (2006) 14(6), 1969-1976.
- [4] Möller, S.: Assessment and Prediction of Speech Quality in Telecommunications. Kluwer Academic Publishers, USA-Boston(2000).