

Erweiterung des E-Modells für Super-Wideband-Sprachübertragung

Marcel Wältermann, Izabela Tucker, Alexander Raake, Sebastian Möller

Deutsche Telekom Laboratories, TU Berlin, 10587 Berlin, Deutschland, Email: marcel.waeltermann@telekom.de

Einleitung

Das E-Modell ist ein von der internationalen Standardisierung für Telekommunikation (ITU-T) empfohlenes Werkzeug für die parametrische Vorhersage von Sprachübertragungsqualität [1]. Es liefert Qualitätsschätzwerte für verschiedenste Übertragungsbedingungen, wie sie bei herkömmlicher Schmalbandsprachübertragung („narrowband“, 300-3400 Hz) auftreten können. Ebenso lässt sich die wahrgenommene Qualität anhand wichtiger Einflussfaktoren bei der Übertragung von Breitbandsprache („wideband“, 50-7000 Hz) quantitativ vorhersagen.

In diesem Beitrag wird das E-Modell erweitert, um so genannte Super-Wideband-Sprachübertragung abzubilden, d.h. die Übertragung im Frequenzband von 50 bis 14000 Hz. Dies geschieht durch eine Aufweitung der Qualitätsskala des E-Modells auf der Basis der Ergebnisse von zwei auditiven Experimenten, die im Folgenden vorgestellt werden. Es wird eine lineare Extrapolation abgeleitet, die zu einer universellen Skala führt, auf der Qualitätsvorhersagen für „narrowband“ (NB), „wideband“ (WB) und „super-wideband“ (SWB) einheitlich dargestellt und direkt miteinander verglichen werden können. So lässt sich insbesondere der maximale Qualitätsgewinn bei der SWB-Übertragung quantifizieren.

Das E-Modell

Das E-Modell liefert Sprachqualitäts-Vorhersagen auf der Grundlage eines additiven Zusammenhangs verschiedener Einflussfaktoren, aus denen sich der *Transmission-Rating-Faktor* R als Qualitätsindex berechnen lässt:

$$R = R_0 - I_s - I_d - I_{e,eff} + A, \text{ mit } R \in [0; R_{0,max} + A]. \quad (1)$$

Der R -Wert hängt mit dem *Mean Opinion Score (MOS)* [1;4.5], den man üblicherweise aus Hör- oder Konversationstests gewinnt, über eine S-förmige Transformationsvorschrift zusammen [1]. Der *Impairment-Faktor* I_s repräsentiert signal-simultane Störungen wie korreliertes Rauschen, I_d beschreibt Störungen, die verzögert zum Signal auftreten (z.B. Sprecherecho) und $I_{e,eff}$ quantifiziert Kodier- und Paketverlusteinflüsse. Kann von einer paketverlustfreien Übertragung ausgegangen werden, wird dieser Faktor durch den *Equipment Impairment-Faktor* I_e ersetzt. Der *Advantage-Faktor* A beschreibt den qualitätsrelevanten Vorteil einer bestimmten Technologie (z.B. Mobilfunk). Für die vorliegende Untersuchung wird ohne Einschränkung der Allgemeinheit angenommen, dass $I_s = I_d = A = 0$ und $I_{e,eff} = I_e$.

Die Größe R_0 reflektiert das Signal-zu-Rausch-Verhältnis (z.B. Umgebungs- und Leitungsrauschen) für einen bestimmten Bandbreiten-Kontext. Kann eine ansonsten

störungsfreie Verbindung angenommen werden, so gilt $R_{0,max,NB} = 100$ für einen NB-Kontext und $R_{0,max,WB} = 129$ für einen WB-Kontext [2].

SWB-Erweiterung des E-Modells

Die Erweiterung der bisherigen WB-Version des E-Modells für SWB-Sprachübertragung erfolgt in zwei Schritten (vgl. [2][3]):

1. Durchführung von zwei auditiven Experimenten zur Bestimmung der Sprachqualität im WB-Kontext (Exp. 1) und im SWB-Kontext (Exp. 2). Exp. 2 enthält eine Untermenge der Bedingungen aus Exp. 1.
2. Transformation der resultierenden *MOS*-Werte aus beiden Experimenten auf die R -Skala und Vergleich der R -Werte der gemeinsamen Bedingungen. Der Grad der Kompression der NB/WB-Bedingungen in Exp. 2 erlaubt die Bestimmung von $R_{0,max,SWB}$.

Auditive Experimente

Exp. 1 (WB-Kontext) enthielt u.a. folgende NB und WB Bedingungen: Eine „direkte“ WB-PCM-Bedingung, die WB-Kodierer G.722.2, G.722 und G.729.1, die NB-Kodierer G.711, G.729A und G.726, verschiedene Kodierer-Tandems, verschiedene Paketverlustraten und sendeseitiges Hintergrundrauschen. In Exp. 2 (SWB-Kontext) waren neben einer „direkten“ SWB-PCM-Bedingung und einigen reinen Bandpässen vier SWB-Kodierer bei verschiedenen Bitraten enthalten (AMR-WB+, Speex, G.722.1 Annex C und ein anonymer Kodierer-Prototyp). Zusätzlich wurden NB- und WB-Kodierer und Bedingungen mit Paketverlusten und Hintergrundrauschen aus Exp. 1 hinzugefügt.

Beide Hörversuche wurden nach den Vorgaben in ITU-T Rec. P.800 durchgeführt. Die verschiedenen Hörergruppen bestanden aus 100 (Exp. 1) und 27 (Exp. 2) Teilnehmern. Das Hörvermögen der Teilnehmer in Exp. 2 wurde mittels Tonaudiometrie getestet, was zum Ausschluss von 6 Hörern führte. Die Hörbeispiele wurden diotisch präsentiert.

Erweiterung der R -Skala

Gemäß den Anforderungen in [1] wurden die aus den Experimenten erhaltenen *MOS*-Werte [1;5] linear in den Bereich [1;4.5] transformiert (vgl. ITU-T Rec. P.833). Die so normierten *MOS*-Werte wurden zunächst zu Werten $R_{WB,100} \in [0; 100]$ (Exp. 1) und $R_{SWB,100} \in [0; 100]$ (Exp. 2) auf die R -Skala überführt [1]. Der Skalenbereich [0; 100] ist allerdings nur gültig für Bedingungen, die in einem reinen NB-Kontext betrachtet werden. Da *MOS*-Werte von der im Versuch enthaltenden Bedingung be-

ster Qualität abhängen, zeigen sich im Vergleich zu einem reinen NB-Experiment Komprimierungseffekte der NB-Bedingungen auf der *MOS*-Skala in Exp. 1, das in einem WB-Kontext durchgeführt wurde. Ebenso sind sowohl die NB- als auch die WB-Bedingungen in Exp. 2 komprimiert aufgrund der enthaltenden SWB Bedingungen.

Die Kompression in Exp. 1 kann durch die in [2] gefundene Expansionsvorschrift kompensiert werden

$$R = R_{WB,129} = 1.29 \cdot R_{WB,100} . \quad (2)$$

Durch die Erweiterung auf den Skalenbereich [0;129] lassen sich die *R*-Werte für die NB- und WB-Bedingungen valide abbilden. In Analogie zu Gl. (2) lassen sich die Ergebnisse aus Exp. 2 ebenfalls auf den Bereich [0;129] überführen:

$$R_{SWB,129} = 1.29 \cdot R_{SWB,100} . \quad (3)$$

Die nun noch verbleibende Kompression der SWB-Versuchsbedingungen lässt sich in Abb. 1 erkennen. Dort sind die (komprimierten) *R*-Werte derjenigen Bedingungen abgebildet, die beiden Experimenten gemeinsam sind (x-Achse: $R_{SWB,129}$ aus Exp. 2; y-Achse: $R = R_{WB,129}$ aus Exp. 1). Um eine vollständige Entkomprimierung

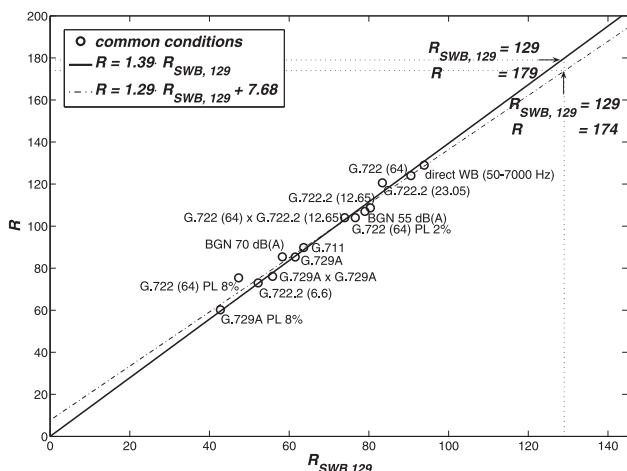


Abbildung 1: *R*-Skalenerweiterung für SWB. BGN .. Hintergrundrauschen, PL .. Paketverlust. Bitraten in Klammern.

und damit valide *R*-Werte der NB-, WB- und SWB-Bedingungen im SWB-Kontext zu erhalten, wurden die *R*-Werte von Exp. 2, $R_{SWB,129}$, auf diejenigen von Exp. 1 abgebildet:

$$R = a \cdot R_{SWB,129} + b . \quad (4)$$

Die Koeffizienten *a* und *b* wurden bestimmt durch Kurvenanpassung. Mit $a = 1.29$ und $b = 7.68$ kann ein *RMSE* von 3.02 erreicht werden. Mit der Randbedingung $b = 0$ ergibt sich $a = 1.39$ und *RMSE* = 3.47. Aufgrund des schlecht rechtfertigbaren Achsenabschnitts wird dieser Zusammenhang im Folgenden betrachtet.

Mit Gl. (4) lässt sich der neue SWB-Bereich der *R*-Skala ermitteln, dessen obere Grenze durch $R_{0,max,SWB}$ bestimmt ist: $R_{0,max,SWB} = 1.39 \cdot R_{0,max,WB} = 1.39 \cdot 129 = 179$. Dies entspricht einem Qualitätsvorsprung von 39% bezüglich WB und 79% bezüglich NB.

SWB Impairment-Faktoren

SWB Equipment Impairment Faktoren $I_{e,SWB}$ der Kodierer aus Exp. 2 können nun anhand von $I_{e,SWB} = R_{0,max,SWB} - R(\text{Bedingung}) = 179 - R(\text{Bedingung})$ bestimmt werden. Aufgrund der Abwärtskompatibilität können die Faktoren $I_{e,SWB}$ für NB und WB Kodierer im Allgemeinen ermittelt werden durch:

$$I_{e,SWB} = \begin{cases} R_{0,max,SWB} - R_{0,max,WB} + I_{e,WB} & (\text{WB}) \\ R_{0,max,SWB} - R_{0,max,NB} + I_{e,NB} & (\text{NB}). \end{cases} \quad (5)$$

In Abb. 2 sind $I_{e,SWB}$ -Werte für einige Bedingungen aus Exp. 2 abgebildet. Die direkte SWB-Bedingung ent-

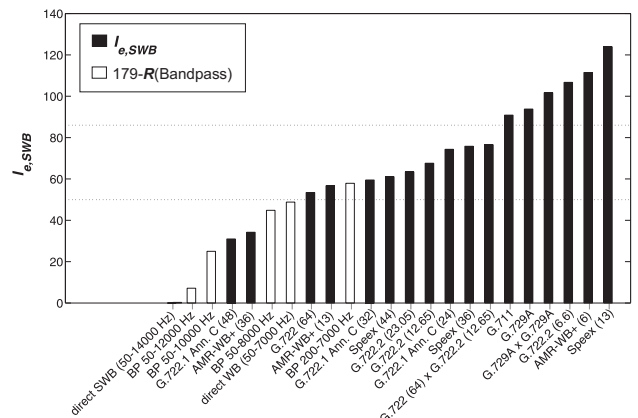


Abbildung 2: Impairment-Faktoren und Bandpass-Einflüsse für einige Bedingungen aus Exp. 2.

spricht einem Impairment-Faktor von Null. Im Allgemeinen steigt die Qualitätseinbuße erwartungsgemäß mit sinkender Bitrate und Audio-Bandbreite. Bei sehr niedrigen Bitraten können SWB-Kodierer die Qualität aufgrund starker nicht-linearer Effekte bzw. interner Bandbreiten-einschränkung u.U. stärker beeinflussen als WB- oder NB-Kodierer.

Zusammenfassung und Ausblick

Die *R*-Skala des E-Modells wurde auf den Wert von $R_{0,max,SWB} = 179$ für SWB-Sprachübertragung erweitert. Dies entspricht einem Qualitätsgewinn von 39% im Vergleich zu WB. Die Abwärtskompatibilität der *R*-Skala erlaubt weiterhin die Darstellung von WB und NB Bedingungen. Weitere Experimente anderer Labore sind notwendig, um die hier gefundenen Ergebnisse zu validieren und die Grundlage für eine Standardisierung zu schaffen.

Literatur

- [1] ITU-T Rec. G.107: The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, Genf, 2009.
- [2] Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M.: Impairment Factor Framework for Wideband Speech Codecs. In: IEEE Trans. Audio, Speech, and Language Proc., 14(6), 1969-1976, 2006.
- [3] Raake, A.: Speech Quality of VoIP – Assessment and Prediction. Wiley, UK-Chichester, West Sussex, 2006.