

Wie gut ist breitbandige Sprachübertragung wirklich? Ein Lagebericht zur Erweiterung des E-Modells

Sebastian Möller¹, Nicolas Côté^{1,2}, Marcel Wältermann¹ und Alexander Raake¹

¹Quality and Usability Lab, Deutsche Telekom Labs, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Deutschland

²France Télécom R&D, 22300 Lannion, Frankreich; Email: sebastian.moeller@telekom.de

Einleitung

Breitbandige (*wideband*, WB) Sprachübertragung bei 50-7000 Hz ermöglicht gegenüber derjenigen im Standard-Telefonband (*narrow-band*, NB, 300-3400 Hz) eine bessere Übertragungsqualität. Auf einer Gesamt-Qualitätsskala, die insbesondere in der Netzplanung Verwendung findet, wurde der Qualitätsvorsprung mit knapp 30% angegeben [1][2]. Jedoch ist bislang noch völlig unklar, in welchen Situationen dieser Vorsprung tatsächlich besteht. So kann er durch niederbitratige Kodierverfahren oder andere Störungen zunichte gemacht werden, sodass breitbandige Sprachübertragung unter Umständen schlechter als eine ungestörte ISDN-Schmalbandübertragung ist.

In diesem Beitrag wird zunächst versucht, den Qualitätsvorsprung auf einer kontextunabhängigen Skala zu verorten. Aus praktischen Gründen wird hierzu die sog. Transmission-Rating-Skala (*R*-Skala) des E-Modells, eines parametrischen Vorhersagemodells für schmalbandige Telefonübertragung, verwendet [3]. In experimentellen Untersuchungen wurden Transformationen zwischen Beurteilungsskalen und der kontextunabhängigen *R*-Skala etabliert. Auf der so verankerten *R*-Skala können anschließend Qualitätsbeeinträchtigungen durch lineare und nichtlineare Verzerrungen, durch Paketverluste oder durch Hintergrundgeräusche quantifiziert werden. Daraus werden Schritte zur Etablierung eines kompletten breitbandigen E-Modells abgeleitet.

Qualitätsvorsprung durch WB-Übertragung

Die Bestimmung der Sprachübertragungsqualität erfolgt sowohl im NB- als auch im WB-Fall mit Hilfe von Hör- oder Konversationsversuchen mittels Beurteilungsskalen. Die Urteile auf den meist 5-stufigen ACR-Skalen werden zu einem sog. Mean Opinion Score (MOS) je Übertragungskanal oder je Sprachprobe gemittelt. Dabei hängen die mittleren Urteile i. Allg. vom Testkontext ab: So wird eine schmalbandige Verbindung in einem reinen NB-Test normalerweise ein besseres Urteil erhalten als in einem gemischten NB/WB-Test. Der Unterschied liegt in der unterschiedlichen Referenz begründet, die beim Qualitätsbeurteilungsprozess angenommen wird, siehe z.B. [4]. Durch Vergleich mit einer etablierten „Telefon-Referenz“ im NB-Fall und einer (bislang wenig etablierten) qualitativ hochwertigen Referenz im NB/WB-Fall wird das Qualitätsurteil im NB/WB-Test schlechter ausfallen, obwohl das Wahrnehmungsereignis dasselbe ist.

Zur Quantifizierung des Qualitätsvorsprungs muss also der Beurteilungskontext, der den Qualitätsvorsprung bestimmt, fest an die Qualitätsskala gekoppelt werden – und somit von der Beurteilungsskala, welche im Experiment verwendet

wird, getrennt werden. Die Qualitätsskala selbst ist dann kontextunabhängig, d.h. sie erlaubt die Quantifizierung von Qualität im „absoluten“ Maßstab. Die Transformation von der Beurteilungsskala zur Qualitätsskala (die die Benutzung der Beurteilungsskala beschreibt) wird dadurch kontextabhängig.

In der Netzwerkplanung bedient man sich seit langem der bereits erwähnten *R*-Skala zur Beschreibung der kontextunabhängigen Qualität. Die Skala lässt im NB-Fall Werte zwischen $R = 100$ (optimale Übertragungsqualität) und $R = 0$ (schlechtest mögliche Übertragungsqualität) zu. Das E-Modell verortet einen Schätzwert der Gesamtqualität durch Subtraktion von Störungsanteilen (sog. Impairment-Faktoren) vom Maximalwert R_0 dieser Skala; letzterer wird nur durch den Störabstand begrenzt, der wiederum durch Leitungsrauschen oder Hintergrundgeräusche eingeschränkt werden kann:

$$R = R_0 - I_s - I_d - I_{e,eff} \quad (1)$$

Hierbei beschreibt I_s die simultan zum Sprachsignal auftretenden Störungen, I_d die verzögert auftretenden Störungen, und $I_{e,eff}$ die durch Kodierer und Paketverluste zeitvariant auftretenden Störanteile.

Im NB-Fall besteht eine feste Zuordnung zwischen *R*- und MOS-Skala, welche in Abb. 1 skizziert ist (durchgezogene Linie). Diese berücksichtigt die Sättigung an den Skalenendpunkten, die bei auditiven Experimenten unausweichlich auftritt. Sie zeigt außerdem, dass die MOS-Skala typischerweise nur bis zum Maximalwert 4,5 ausgenutzt wird.

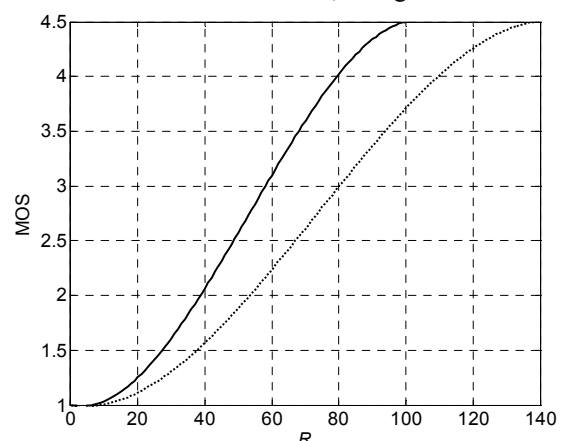


Abbildung 1: Zusammenhang zwischen *R*- und MOS-Skala bei NB (durchgezogen) und NB/WB (gestrichelt)

Zur Ableitung einer entsprechenden Transformationsregel für den WB-Fall wird ein indirektes Verfahren vorgeschlagen: Eine Anzahl von NB-Verbindungen wird zunächst in einem reinen NB-Test beurteilt, und anschließend zusammen mit anderen WB-Verbindungen in einem gemischten

NB/WB-Test. Aus den MOS-Werten des NB-Tests lassen sich mittels der durchgezogenen Transformations-Kennlinie aus Abb. 1 R -Werte ableiten; diese werden als R_{NB} bezeichnet. Durch Anwendung der gleichen Transformationsregel auf die MOS-Werte des NB/WB-Tests lassen sich äquivalente $R_{NB/WB}$ -Werte ableiten, die durch die schlechtere Bewertung in diesem Test geringer ausfallen.

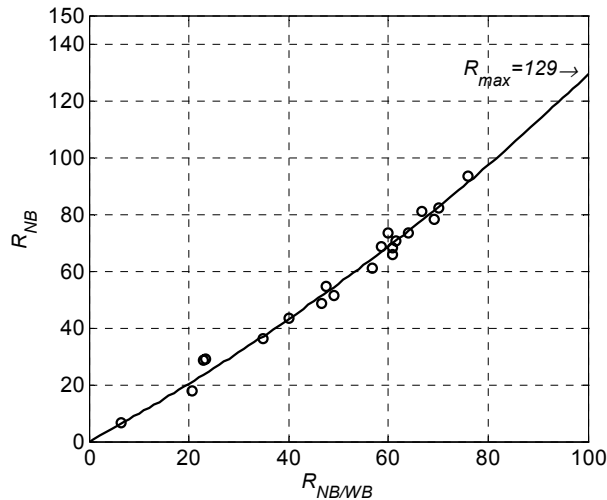


Abbildung 2: Erweiterung der R -Skala durch Extrapolation

Zwischen zusammengehörigen R_{NB} - und $R_{NB/WB}$ -Paaren lässt sich eine Interpolationslinie ziehen; diese ist beispielhaft für einen Test und eine exponentielle Funktion in Abb. 2 gezeigt. Der Schnittpunkt dieser Linie mit der Achse $R_{NB/WB} = 100$ ist der maximale R -Wert im WB-Fall; im Mittel ergibt sich hier der schon angedeutete Wert von 129, was einer 29-prozentigen Qualitätsverbesserung (ausgedrückt auf der R -Skala) entspricht. Für diesen Wert ist in Abb. 1 eine dem WB-Fall entsprechende Transformationsregel zwischen MOS und R aufgezeichnet (gestrichelte Linie).

Beeinträchtigung durch WB-Störungen

Mit Hilfe der neuen Transformation R -MOS lassen sich nun WB-Störungen auf der erweiterten R -Skala quantifizieren. Unter der Annahme, dass das Additivitätsprinzip des E-Modells im WB-Fall gültig ist, ergeben sich Impairment-Faktoren aus der Differenz zwischen optimalem $R_o = 129$ und dem aus dem auditiven beobachteten MOS-Wert transformierten R -Wert der gestörten Verbindung. Bislang wurden für folgende Störungen Impairment-Faktoren etabliert:

Sprachkodierer: In [1] wurden mit Hilfe der dargestellten Methode I_e -Werte für eine Reihe von WB-Kodierern bestimmt. Sie bewegen sich zwischen 1 (bei G.722.2@23,05 kbit/s) und 39 (bei G.722.2@6,6 kbit/s). D.h. dass bei niederbitratigen Kodierern der Qualitätsvorsprung durch die Kodierverzerrung wieder zunichte gemacht werden kann. Für asynchrone Tandems zeigt sich, dass die Reihenfolge der Kodierer einen Einfluss hat; die Additivität aus Gl. (1) ist also nicht vollständig gegeben.

Paketverluste: Durch Abwandlung der in [2] gefundenen Formel zur Abhängigkeit von $I_{e,eff}$ von der Paketverlustrate Ppl (in %) und einem kodierer-spezifischen Robustheitsfaktor Bpl lassen sich $I_{e,eff}$ -Werte berechnen, die sowohl die nicht-linearen Kodierverzerrungen als auch die zeitvarianten Störungen umfassen [1]. Die Ableitung war dabei etwas weni-

ger zuverlässig, und es wurden bislang nur zufällige (nicht gehäufte) Fehlermuster erfasst.

Bandbreite: In [5] wird eine Methode beschrieben, um den Einfluss linearer Störungen vom Gesamt- $I_{e,eff}$ zu trennen. Dazu wird ein bandbreiten-abhängiger Faktor I_{bw} eingeführt, der aus der äquivalenten Rechteck-Bandbreite (in Bark) und der Mittenfrequenz (in Hz) berechnet wird. Es verbleibt ein Term I_{res} , der sowohl die nichtlinearen Kodierverzerrungen als auch mögliche andere Störeinflüsse umfasst.

Hintergrundgeräusche: In [6] werden Untersuchungen zur Abhängigkeit des R_o vom Störabstand vorgestellt. Es zeigt sich, dass R_o auch bei hohen Störabständen noch zum Maximalwert 129 ansteigt; dies steht im Gegensatz zum NB-Fall, bei dem oberhalb eines Störabstandes von ca. 40 dB ein konstanter R -Wert beobachtet wird. Die Abflachung der Kurve im NB-Fall führt dazu, dass sich der Qualitätsvorsprung von ca. 36 Punkten bei hohem Störabstand auf weniger als 10 Punkte bei mittlerem und niedrigem Störabstand verringert. Bei der Kombination von Kodierern und Hintergrundgeräuschen zeigt sich, dass Kodierstörungen durch Rauschen maskiert werden.

Nicht-optimale Lautheit: Unter Verwendung verschiedener Lautheits-Modelle werden in [7] $I_{e,loud}$ zur Beschreibung auf der erweiterten R -Skala abgeleitet.

Ausblick

Die bislang gefundenen Impairment-Faktoren bilden einen ersten Grundstock zu einer pragmatischen Erweiterung des E-Modells. Dabei kann zunächst von einem R_o -Wert für WB-Verbindungen ausgegangen werden, wie er z.B. in [6] bestimmt wird. Von diesem können dann lineare Anteile in Form eines I_{bw} und nichtlineare Anteile in Form eines I_{res} abgezogen werden; letztere lassen sich aus den zuvor bestimmten I_e breitbandiger Kodierer berechnen. Der Einfluss von Paketverlusten kann in diesem Faktor mit berücksichtigt werden; $I_{e,loud}$ sollte ebenfalls einbezogen werden. Untersuchungen zu Konversations-Störungen wie Echos oder nicht-optimales Rückhören fehlen bislang.

Literatur

- [1] Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M.: Impairment Factor Framework for Wideband Speech Codecs. IEEE Trans. Audio, Speech, and Language Processing 14(6), 1969-1976, 2006
- [2] Raake, A.: Speech Quality of VoIP – Assessment and Prediction. Wiley, Chichester, West Sussex, 2006
- [3] ITU-T Rec. G.107: The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, Geneva, 2005
- [4] Jekosch, U.: Voice and Speech Quality Perception – Assessment and Evaluation. Springer, Berlin, 2005
- [5] Wältermann, M., Raake, A.: Ein neuer Qualitätsfaktor für die Modellierung linearer Verzerrungen bei der Übertragung von Sprache. DAGA 2008, Dresden
- [6] Ramirez, J.-P., Raake, A., Wältermann, M., Möller, S.: Predicting Speech Quality under Noise for a Wideband E-Model. DAGA 2008, Dresden
- [7] Côté, N., Gautier-Turbin, V., Möller, S., Raake, A.: Assessment of Different Loudness Models for Perceived Speech Quality. DAGA 2008, Dresden