

Erweiterung der Netzwerkplanung auf super-breitbandige Sprachübertragung

Sebastian Möller^{1,2}, Gabriel Mittag¹, Friedemann Köster¹, Vincent Barriac³, Nicolas Côté⁴

¹ *Quality and Usability Lab, TU Berlin; E-Mail: sebastian.moeller@tu-berlin.de, gabriel.mittag@tu-berlin.de, friedemann.koester@googlegmail.com*

² *Deutsches Forschungszentrum für Künstliche Intelligenz, Berlin*

³ *Orange Labs Services, Lannion; E-Mail: vincent.barriac@orange.com*

⁴ *Wavely, Villeneuve d'Ascq; E-Mail: nicolas.cote@wavely.fr*

Einleitung

Zur Planung von Sprachübertragungsnetzen werden üblicherweise Rechenmodelle eingesetzt, welche den Einfluss unterschiedlicher Störungen auf die Gesamt-Sprachqualität abschätzen. Das bekannteste Modell ist das E-Modell, welches aus Parametern, die einzelne Störungsursachen beschreiben (z.B. Abschwächungen, Verzögerungen, Rauschpegel), perzeptive Beeinträchtigungs-Faktoren berechnet, die auf einer Gesamtqualitätsskala als additiv angenommen werden [1][2]. Diese Beeinträchtigungs-Faktoren werden von einem Maximalwert R_{max} abgezogen, um die angenommene Gesamtqualität in einer Konversationssituation zu berechnen.

Bislang ist das E-Modell für schmalbandige (300-3400 Hz) und eingeschränkt auch für breitbandige (50-7000 Hz) Sprachübertragung definiert und standardisiert [3][4]. Im Zuge der Erweiterung auf Super-Breitband (20-14000 Hz) oder Vollband (20-20000 Hz) stellt sich die Frage, welcher Maximalwert R_{max} dort Verwendung finden sollte, und wie sich die wichtigsten Störungen daraus ableiten. Hierzu sollen in diesem Beitrag Hörtestergebnisse unterschiedlicher Provenienz analysiert und mittels zweier unterschiedlicher Verfahren auf die Gesamtqualitätsskala transformiert werden.

Im folgenden Abschnitt werden wir kurz auf die Qualitätsskalen des E-Modells eingehen und beschreiben, wie prinzipiell eine Erweiterung auf Super-Breitband oder Vollband möglich ist. Anschließend werden zwei Datenbanken mit Hörtestergebnissen beschrieben, auf welche die Verfahren angewendet werden. Die daraus erhaltenen Maximalwerte werden analysiert und mit Ergebnissen anderer Forschergruppen verglichen. Auf Basis dieser Ergebnisse hat sich die International Telecommunication Union (ITU-T) inzwischen entschieden, einen Maximalwert zu empfehlen, um daraus Beeinträchtigungs-Faktoren für verschiedene Super-Breitband-Codex abzuleiten. Abschließend werden notwendige Schritte mit dem Ziel eines super-breitbandigen E-Modells diskutiert.

Qualitätsskalen des E-Modells

Das E-Modell verfolgt das Ziel, auf Basis messbarer Parameter eines Sprachübertragungsnetzes eine Abschätzung der für ein Gespräch erzielbaren Konversationsqualität zu berechnen. Die Parameter orientieren sich klassischerweise an den einzelnen Übertragungswegen, welche in Form eines

frequenz-gewichteten Abschwächungswertes (eines sog. *Loudness Ratings*) und einer mittleren Verzögerungszeit vereinfacht beschrieben werden. Zusätzlich werden additive Rauschquellen (für Hintergrundgeräusche, leitungsrauschen und Quantisierungsrauschen) mittels eines gewichteten Pegels dargestellt. Nichtlineares Verhalten, wie bspw. Kodierverzerrungen und durch Paketverluste hervorgerufene Unterbrechungen, werden wiederum durch sog. *Equipment Impairment Factors* oder Paketverluststraten als skalare Werte beschrieben. Eine solche Vereinfachung ist für die Netzwerkplanung notwendig, da viele Details der einzelnen Komponenten zum Planungszeitpunkt noch nicht verfügbar sind.

Die Parameter werden nun auf eine Gesamt-Qualitätsskala, die sog. Transmission-Rating-Skala R transformiert. Dabei bedient man sich der Annahme, dass perzeptiv unterschiedliche Arten von Störungen additiv auf der Transmission-Rating-Skala überlagert werden können. So berechnet man je nach Bandbreite der Übertragung zunächst einen vereinfachten Störabstand R_{max} , von dem dann weitere Beeinträchtigungen (nichtlineare Verzerrungen, Echos, etc.) in Form sogenannter Beeinträchtigungs-Faktoren abgezogen werden. Hieraus ergibt sich ein Gesamt-Transmission-Rating R , welches indikativ für die Konversationsqualität der Übertragung steht. Das Gesamt-Transmission-Rating R lässt sich anschließend mittels einer S-förmigen monotonen Beziehung in einen *Mean Opinion Score* (MOS) transformieren, wie man ihn mittels eines Konversationstests als Mittelwert absoluter Kategorien-Bewertungen einer Anzahl von Versuchspersonen erhalten würde. Eine detaillierte Analyse und Beschreibung des Verfahrens findet sich in [5].

Erweiterung auf Super-Breitband

Das E-Modell wurde ursprünglich für schmalbandige Sprachübertragung entwickelt. Für diese Bandbreite wurde der maximale Störabstand auf 100 normiert, und die Transmission-Rating-Skala so limitiert, dass das Intervall $R \in [0; 100]$ auf den Bereich $MOS \in [1; 4,5]$ transformiert wird. Der Wert 100 entspricht also der bestmöglichen, der Wert 0 der schlechtesten möglichen Qualität. Dabei wird implizit davon ausgegangen, dass Versuchspersonen bei einer absoluten Kategorienzuordnung auf einer 5-stufigen Skala kaum alle den besten Wert ankreuzen, sondern sich im Mittel ein Wert von maximal 4,5 ergeben würde. Unter Berücksichtigung

eines geringen Leitungsrauschens sowie des durch logarithmische PCM hervorgerufenen Quantisierungsrauschens ergibt sich für eine klassische ISDN-Verbindung ein Wert von $R = 93,1$, entsprechend einem MOS von ca. 4,4.

Für breitbandige Sprachübertragung (50-7000 Hz) wurde das E-Modell vor einigen Jahren so erweitert, dass der Qualitätsbereich der schmalbandigen Übertragung auf der Transmission-Rating-Skala konstant bleibt. Hierzu wurden Hörtestergebnisse verwendet, bei denen Versuchspersonen innerhalb des gleichen Tests sowohl schmalbandige als auch breitbandige Verbindungen zu beurteilen hatten. In einem solchen „gemischten“ (schmalbandigen und breitbandigen) Test ist zu beobachten, dass Versuchspersonen schmalbandige Verbindungen schlechter bewerten, als wenn sie diese (gleichen) Verbindungen in einem rein schmalbandigen Test zu bewerten haben; für die gleichen Verbindungen ist der MOS im schmalbandigen Test also geringer als im gemischten Test.

Um trotzdem den R -Wert für den schmalbandigen Fall konstant zu halten ist zunächst eine Transformation der Ergebnisse des gemischten Tests von MOS auf R notwendig. Zur Verwendung der entsprechenden Transformation des E-Modells ist es notwendig, dass alle auditiv bestimmten MOS-Werte im Intervall $MOS \in [1; 4,5]$ liegen; ist dies nicht der Fall, d.h. treten höhere Werte als $MOS = 4,5$ auf, so müssen diese zunächst mittels linearer Transformation auf diesen Bereich begrenzt werden. Die aus der Transformation erhaltenen R -Werte werden dann so linear normiert, dass für den schmalbandigen Fall der im Test beobachtete niedrigere R -Wert wieder auf den ursprünglichen R -Wert eines rein schmalbandigen Tests heraufskaliert wird, der mit dem E-Modell vorhergesagt wird. Daraus ergibt sich dann für breitbandige Verbindungen ein höherer maximaler R -Wert. In verschiedenen Tests wurden nach dieser Methode Werte für R_{max} für den breitbandigen Fall bestimmt, und im Rahmen der internationalen Standardisierung bei der ITU-T ein genereller Maximalwert für breitbandige Übertragung von $R_{max} = 129$ festgelegt [4].

Dasselbe Verfahren der Erweiterung auf der R -Skala soll nun wiederum angewendet werden, um auch für den super-breitbandigen Fall einen neuen Maximalwert zu bestimmen. Dazu müssen die MOS-Werte eines „gemischten“ Tests (in diesem Fall Super-Breitband gemischt mit Schmalband und/oder Breitband) zunächst mit der für den entsprechenden Schmalband- oder Breitband-Fall etablierten Berechnungsvorschrift (S-förmige Relation des schmalbandigen [3] oder breitbandigen E-Modells [4]) auf die R -Skala transformiert werden; ggf. ist hierzu eine vorherige Begrenzung auf den Bereich $MOS \in [1; 4,5]$ notwendig. Die so erhaltenen R -Werte sind dann „komprimiert“ auf den Bereich, der für Schmalband oder Breitband vorgesehen ist, und sie müssen so heraufskaliert werden, dass die R -Werte für die schmalbandigen bzw. breitbandigen Verbindungen wieder auf denjenigen, vordefinierten Werten des (schmalbandigen oder breitbandigen) E-Modells landen. Bei gleicher Skalierung ergibt sich dann der R -Wert für die beste, anderweitig ungestörte Super-Breitband-Verbindung als maximal im Super-Breitbandfall erreichbarer Wert R_{max} .

Es ist aber auch ein alternativer Weg denkbar, der zusätzlich beschränkt werden soll, um die Nichtlinearität des Zusammenhangs zwischen MOS und R zu untersuchen. Dazu wird ein direkter linearer Zusammenhang zwischen den MOS-Werten und den vordefinierten R -Werten der schmal- und/oder breitbandigen Verbindungen in einem gemischten Test etabliert, in dem neben diesen schmal- und/oder breitbandigen Verbindungen auch super-breitbandige Verbindungen bewertet wurden. Dieser lineare Zusammenhang wird nun genutzt, um den zum super-breitbandigen MOS-Wert zugehörigen R -Wert abzulesen, und somit R_{max} für den super-breitbandigen Fall zu bestimmen.

Verwendete Datenbanken

Zur Berechnung von R_{max} im Super-Breitbandfall standen zwei Datenbanken zur Verfügung, bei denen sowohl super-breitbandige als auch schmal- und/oder breitbandige Verbindungen im selben Test auf einer 5-stufigen ACR-Skala bewertet wurden.

Die erste Datenbank wurde von Orange im Jahr 2015 erstellt, um die Qualität von EVS- und Opus-Kodierern zu vergleichen [6]. Französische Sprachdaten wurden 24 französischen Muttersprachlern in einem Hörversuch diotisch über Sennheiser280pro-Kopfhörer präsentiert, deren Übertragungsbandbreite auf das Super-Breitband eingegrenzt wurde. Von den 54 unterschiedlichen Kanalbedingungen waren 61% super-breitbandig und 39% breitbandig. Die einzigen im Sprachmaterial vorkommenden Störungen waren die verwendeten Kodierer bei unterschiedlichen Bitraten, sowie randomisiert eingefügte Paketverluste unterschiedlicher Verlustraten. Für die breitbandigen Kodierer existieren bereits die zugehörigen Parameter-Werte für das breitbandige E-Modell [7], sodass die zugehörigen Transmission-Rating-Werte R bestimmt werden können; dabei handelt es sich um den AMR-WB-Kodierer bei Bitraten zwischen 12,65 und 23,85 kbit/s.

Die zweite Datenbank wurde von der Fa. Rohde & Schwarz zur Entwicklung des instrumentellen Schätzers POLQA (jetzt ITU-T Rec. P.863) erstellt und uns für die vorliegende Umrechnung freundlicherweise zur Verfügung gestellt [8]. Sie umfasst 52 unterschiedliche Kanalbedingungen zwischen Schmalband und Vollband (23% Schmalband, 37% Breitband, 19% Super-Breitband, 21% Vollband), welche nur durch Kodierer und Paketverluste gestört waren. Da sich der MOS-Wert der ansonsten ungestörten Vollband-Bedingungen kaum von dem der Super-Breitband-Verbindung unterschied wird hier davon ausgegangen, dass durch das Vorhandensein von Vollband-Verbindungen keine Verzerrung der Testergebnisse auftrat. Die Stimuli wurden diotisch präsentiert und von 24 Probanden auf der 5-stufigen ACR-Skala bewertet.

Ergebnisse

Die Ergebnisse der Skalierung sind für die direkte Umrechnung aus MOS (linke Bildhälften) sowie die Umrechnung mit vorheriger Transformation auf die R -Skala (rechte Bildhälften) in Abbildung 1 für die Orange-Datenbank und in Abbildung 2 für die Rohde&Schwarz-Datenbank dargestellt. Gekennzeichnet sind jeweils die Datenpunkte für

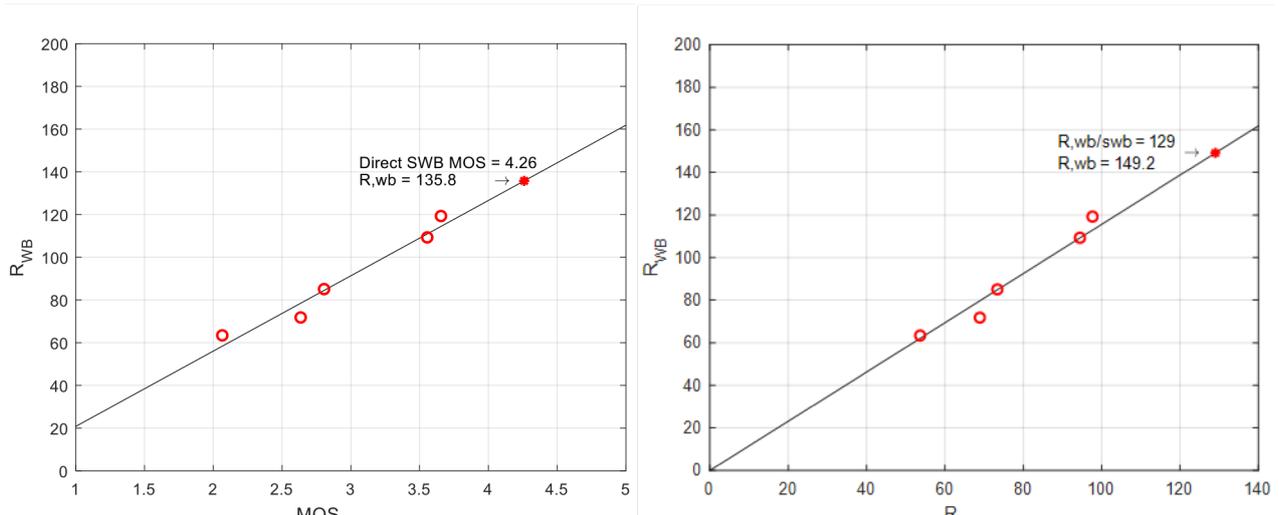


Abbildung 1: Vergleich der unskalierten Testergebnisse (x-Achse Index WB/SWB) und der hochskalierten Ergebnisse (y-Achse, Index WB) der Orange-Datenbank. Links: Direkter Vergleich MOS-R; rechts: Vergleich nach vorheriger Transformation auf die R-Skala. Mit Pfeil gekennzeichnet ist jeweils der Datenpunkt für die ansonsten ungestörte Super-Breitband-Verbindung ergibt.

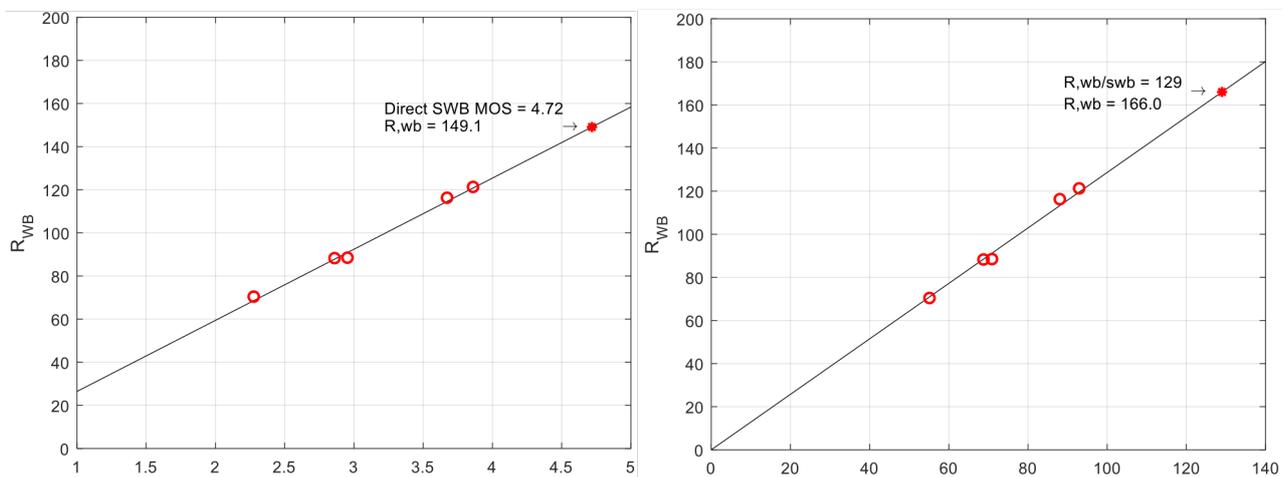


Abbildung 2: Vergleich der unskalierten Testergebnisse (x-Achse, Index WB/SWB) und der hochskalierten Ergebnisse (y-Achse, Index WB) der Rohde&Schwarz-Datenbank. Links: Direkter Vergleich MOS-R; rechts: Vergleich nach vorheriger Transformation auf die R-Skala. Mit Pfeil gekennzeichnet ist jeweils der Datenpunkt für die ansonsten ungestörte Super-Breitband-Verbindung ergibt.

die ansonsten ungestörte Super-Breitband-Verbindung, aus denen sich der maximale Wert R_{max} ergibt.

Zunächst fällt auf, dass sich alle Datenpunkte recht gut um die Interpolationskurven gruppieren, d.h. dass die Abweichungen recht gering sind. Demgegenüber sind die MOS-Werte für die ansonsten ungestörte Super-Breitband-Verbindung in beiden Tests stark unterschiedlich (4,26 bei Orange und 4,72 bei Rohde&Schwarz); dies unterstreicht die Notwendigkeit einer Normierung und Transformation der Ergebnisse.

Die zugehörigen maximalen R -Werte variieren zwischen $R_{max} = 135,8$ (Orange-Datenbank, Berechnung direkt aus MOS) und $R_{max} = 166$ (Rohde&Schwarz-Datenbank, Berechnung nach Transformation), und mitteln sich über die beiden Datenbanken und Berechnungsmethoden zu $R_{max} = 150$. Insgesamt ergeben sich bei der Berechnung nach vorheriger Transformation stets geringere Werte R_{max} als bei Berechnung direkt aus MOS.

Diskussion und Vergleich

Trotz guter Gruppierung um die Interpolationskurve zeigen sich Unterschiede in den ermittelten R_{max} -Werten, je nach verwendeter Datenbank und Transformationsmethode. Zur Ermittlung stabiler Werte sollen die hier berechneten Werte daher mit denen einer japanischen Forschergruppe bei NTT verglichen werden [9]. Diese verwendete acht bei NTT aufgezeichnete Datenbanken (vier davon breitbandig, vier gemischt breitbandig-super-breitbandig) und stellte Interpolationskurven nach vorheriger Transformation auf die R -Skala auf. Die zur Interpolation verwendeten Kanalstörungen bestanden allerdings bei der Hälfte der Datenbanken aus signalkorreliertem Rauschen, zur anderen Hälfte aus Kodierern und signalkorreliertem Rauschen; auch berechnungsverfahren unterschied sich teilweise vom hier vorgestellten. Trotz völlig unterschiedlicher Datenbanken (in japanischer Sprache) und unterschiedlichen Versuchspersonen bewegen sich die ermittelten R_{max} -Werte zwischen 138 und 155, sofern die Interpolationsgerade wie im

hier beschriebenen Verfahren durch den Nullpunkt gelegt wird. Im Mittel ergab sich über alle acht Datenbanken ein Wert von $R_{max} = 146$. Dieser Wert liegt nur 4 Punkte unter dem hier ermittelten Mittelwert, was eine erstaunlich gute Übereinstimmung darstellt.

Zusammenfassung und offene Fragen

Im Beitrag wurden Hörtestergebnisse unterschiedlicher Provenienz analysiert und mittels zweier unterschiedlicher Verfahren auf die Gesamtqualitätsskala transformiert. Es zeigte sich, dass je nach Testdatenbank ein Qualitätsgewinn von 36 bis 66% gegenüber dem Schmalbandfall erzielbar ist, und immerhin noch ein Qualitätsgewinn von 5 bis 29% gegenüber dem Breitbandfall. Unter Abwägung der Eigenheiten der einzelnen Tests und Umrechnungsverfahren hat sich daher die Study Group 12 der International Telecommunication Union (ITU-T SG12) in ihrer Sitzung im September 2017 entschieden, einen Maximalwert von $R_{max} = 148$ zu empfehlen.

Mit Hilfe dieses Maximalwertes lassen sich nun weitere Komponenten eines super-breitbandigen E-Modells ermitteln. Als wichtigste Aufgabe gilt hierbei, Beeinträchtigungs-Faktoren für Kodierer abzuleiten, welche im Super-Breitband arbeiten. Hierzu ist es notwendig, die R -Werte der Kodier-Bedingungen vom hier ermittelten Wert R_{max} abzuziehen, da Beeinträchtigungs-Faktoren für Kodierer auf der R -Skala additiv sind. Solche Beeinträchtigungs-Faktoren lassen sich sowohl für Kanalbedingungen mit als auch ohne Paketverlusten bestimmen. Sobald Werte für unterschiedliche Paketverlusten vorliegen kann eine erste Formel für den Einfluss von Paketverlusten (Beeinträchtigungs-Faktor bzw. R -Wert als Funktion der Verlustrate in %) etabliert werden. Beim schmalbandigen und breitbandigen E-Modell verwendet die entsprechende Formel einen Parameter, der die Robustheit des Kodierers gegenüber Paketverlusten kennzeichnet. Es ist zu erwarten, dass auch im Super-Breitbandfall ein solcher Parameter zu bestimmen ist.

Mittelfristiges Ziel der Arbeiten ist die Erstellung eines kompletten super-breitbandigen E-Modells. Dazu sind neben den Beeinträchtigungen durch Kodierer und Paketverluste auch Störungen auf dem analogen Teil der Übertragungsstrecke, wie Hintergrundrauschen am Send- und Empfangsort, Frequenzverzerrungen und Leitungsruschen der analogen Signalpfade, sowie zu hohe oder niedrige Signalpegel zu berücksichtigen. Zur Bestimmung der Konversationsqualität müssen außerdem noch die Effekte von Hörer- und Sprecherechos, sowie die Beeinträchtigung der Konversation durch einfache Verzögerungen, modelliert werden. Hierzu sind Konversationsversuche mit Probanden unerlässlich.

Literatur

- [1] Johannesson, N. O.: The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks, IEEE Communications Magazine, Jan. 1997, 70–79.
- [2] ETSI Technical Report ETR 250: Transmission and Multiplexing (TM); Speech Communication Quality fromMouth to Ear for 3,1 kHz Handset Telephony across

Networks, European Telecommunications Standards Institute, Sophia Antipolis, 1996.

- [3] ITU-T Rec. G.107: The E-model: a computational model for use in transmission planning, International Telecommunication Union, Genf, 2015.
- [4] ITU-T Rec. G.107.1: Wideband E-model, International Telecommunication Union, Genf, 2015.
- [5] Möller, S.: Assessment and Prediction of Speech Quality in Telecommunications, Kluwer Academic Publ., Boston MA, 2000.
- [6] ETSI TR 26.952: Codec for Enhanced Voice Services (EVS) Performance characterization V14.0.0, European Telecommunications Standards Institute, Sophia Antipolis, 2017.
- [7] ITU-T Rec. G.113 Amendment 1: Revised Appendix IV – Provisional planning values for the wideband equipment impairment factor and the wideband packet loss robustness factor, International Telecommunication Union, Genf, 2009.
- [8] ITU-T Contribution SG12-C22: A subjective ACR LOT testing fullband speech coding and prediction by P.863, International Telecommunication Union, Study Group 12 (Source: Rohde & Schwarz), Genf, 2017.
- [9] ITU-T Contribution SG12-C79: Vorlage für Beiträge zur DAGA, International Telecommunication Union, Study Group 12 (Source: Nippon Telegraph and Telephone Corporation (NTT), Japan), Genf, 2017.