

# Ein neuer Qualitätsfaktor für die Modellierung linearer Verzerrungen bei der Übertragung von Sprache

Marcel Wältermann und Alexander Raake

Quality und Usability Lab, Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Deutschland

{marcel.waeltermann, alexander.raake}@telekom.de

## Einleitung

Die Qualität übertragener Sprache lässt sich mit dem sog. E-Modell [1] vorhersagen, einem von der ITU-T empfohlenen Werkzeug für die Netzwerkplanung. In der aktuellen Version des E-Modells sind Sprachqualitätsvorhersagen für lineare Verzerrungen, wie sie etwa durch mangelhafte elektroakustische Eigenschaften von Endgeräten entstehen, nicht möglich. Diese können allerdings einen erheblich qualitätsmindernden Faktor darstellen [2][3]. Ebenso wird die Qualitätsbeeinträchtigung von Schmalband- gegenüber Breitband-Sprachübertragung (50-7000 Hz) lediglich implizit durch eine gemeinsame Größe, den sog. *Equipment Impairment Factor* abgedeckt [4].

In diesem Beitrag wird ein neuer Qualitätsfaktor vorgestellt, mit dessen Hilfe lineare Verzerrungen explizit berücksichtigt werden. Dieser sog. *Bandwidth Impairment Factor* wird auf der Grundlage zweier einfacher Parameter aus der „Übertragungsfunktion“ der Telefonstrecke geschätzt. Die Aufspaltung herkömmlicher Equipment Impairment Faktoren in einen linearen und einen Residualanteil erlaubt eine nahtlose Erweiterung des bestehenden Modells und die Vorhersage des Einflusses von Kanalfiltern, Endgeräten und Kodierern.

## Das E-Modell

Das E-Modell berücksichtigt bei der Vorhersage der Sprachqualität verschiedene Klassen von Störungen. Diese Störungsklassen werden durch Qualitätsfaktoren parametrisiert, sog. *Impairment Factors*. Hierzu zählen das Signal-Rausch-Verhältnis ( $R_0$ ), signal-simultane Störungen ( $I_s$ ) und verzögerte Störungen wie etwa Echoeinflüsse ( $I_d$ ). Verzerrungen aufgrund von verschiedenen Kodierern werden unter dem sog. *Equipment Impairment Factor*  $I_e$  zusammengefasst, der unter der Berücksichtigung von Paketverlusten erweitert wird zum *Effective Equipment Impairment Factor*  $I_{e,eff}$ . Diese Faktoren lassen sich auf einer Gesamtqualitätsskala, der sog.  $R$ -Skala, aufsummieren und ergeben die Gesamtqualität der Verbindung:

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

Die  $R$ -Werte können in  $MOS$ -Werte (*Mean Opinion Scores*) überführt werden, die üblicherweise den mittleren Versuchspersonenurteilen aus auditiven Sprachqualitätstests entsprechen ( $MOS \in [1; 4.5]$ ).  $R_{0,max}$  stellt die höchste zu erreichende Qualität dar. In der konventionellen Schmalband-Telefonie (narrowband (NB), 300-3400 Hz) gilt  $R_{0,NB,max} = 100$  und entspricht einem

$MOS$ -Wert von 4.5. Ungestörte Breitbandsprache (wideband (WB), 50-7000 Hz) hat gegenüber einem Standard-ISDN Kanal (G.711-Kodierung,  $R_{G.711} = 93.2$ ) einen Qualitätsvorteil von ca. 36% [4], d.h. es gilt  $R_{0,WB,max} = 129$ . In derselben Studie wurden  $I_e$ -Werte für Breitband-Sprachkodierer abgeleitet. Durch den Zusammenhang

$$I_{e,WB} = (129 - 93.2) + I_{e,NB} \quad (2)$$

lässt sich die Qualität von NB-Kodierern auch auf der erweiterten  $R$ -Skala darstellen und mit WB-Kodierern vergleichen.

Die Additivitätseigenschaft in Gl. (1) ist allerdings nicht länger gültig, wenn zwei NB-Kodierer im WB-Kontext in Reihe geschaltet werden. Da der Bandbreiten-Nachteil der NB-Kodierer implizit durch den Faktor  $I_{e,WB}$  abgebildet wird, würde bspw. im Fall einer zweifachen G.711-Kodierung ein unrealistisch hohes Impairment von  $2 \cdot 36$  entstehen. Jedoch kann der Bandbreiten-Nachteil durch einen separaten Faktor abgefangen und so das Problem umgangen werden (s. nächster Abschnitt).

## Der Bandwidth Impairment Factor

Neben der im vorangehenden Abschnitt aufgeführten potentiellen Verletzung der Additivitätseigenschaft in der derzeitigen Version des E-Modells ist es zudem nicht möglich, reine Störungen aufgrund von Einschränkungen in der Bandbreite durch Kodierer, Endgeräte oder Kanalfilter explizit zu erfassen. Um diesem Nachteil entgegen zu wirken, wurde in [3] ein dedizierter Impairment Faktor eingeführt, der sog. *Bandwidth Impairment Factor*  $I_{bw}$ . Dieser wurde aus auditiven Experimenten abgeleitet, in denen hauptsächlich linear gefilterte Sprache in einem WB-Kontext qualitativ bewertet wurde.  $I_{bw}$  berechnet sich wie folgt:

$$I_{bw} = 0.035 \cdot |s| - 0.0067 \cdot s - 7.4 \cdot \frac{z_{bw}}{\text{Bark}} + 129.2, \quad (3)$$

mit  $s = \frac{f_c}{\text{Hz}} - 9.9 \cdot \left( \frac{z_{bw}}{\text{Bark}} + 101.8 \right)$ .

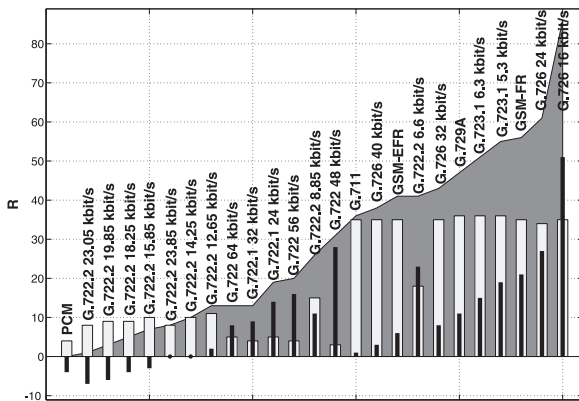
Hierbei beschreibt  $z_{bw}$  die Übertragungsbandbreite in Bark [5] und  $f_c$  die Mittenfrequenz in Hz, für die ein bandbreitenabhängiges Optimum existiert. In der Praxis kann  $z_{bw}$  durch die sog. *Equivalent Rectangular Bandwidth (ERB)* [3] angenähert werden, die wiederum aus dem Amplitudenspektrum des Übertragungskanals nach dem Verfahren aus [6] geschätzt werden kann. Die Mittenfrequenz  $f_c$  ergibt sich dann aus dem geometrischen Mittel der Eckfrequenzen der auf die Hz-Skala transformierten *ERB*.

Mit Hilfe dieses neuen Faktors können beliebige Frequenzgänge der Übertragungsstrecke auf der erweiterten  $R$ -Skala als Skalar quantifiziert werden. Da Unterschiede der Übertragungsbandbreite bisher in  $I_{e,WB}$  zusammengefasst sind, wird in diesem Beitrag eine Zerlegung dieses Faktors in einen linearen Anteil, dargestellt durch  $I_{bw}$ , und einen Residualanteil  $I_{res}$  vorgeschlagen:

$$I_{e,WB} = I_{bw} + I_{res} . \quad (4)$$

Im Folgenden soll gezeigt werden, dass auch der Einfluss von Kodierern, für die eine einwandfreie Schätzung des Amplitudenspektrums oft nicht gelingt, sinnvoll durch  $I_{bw}$  angegeben werden kann. Hierzu wird angenommen, dass für die betrachteten Codecs  $I_{e,WB}$  bekannt ist.

Abb. 1 zeigt für eine Reihe von NB- und WB-Kodierern verschiedener Bitraten die  $I_{e,WB}$ -Werte (aus ITU-T Rec. G.113) in aufsteigender Reihenfolge (dunkelgrau), die Schätzergebnisse für  $I_{bw}$  (hellgraue Balken), und die sich aus Gl. (4) ergebenden Residual-Faktoren  $I_{res}$  (schwarze Balken). Anhand der Abbildung lässt sich zunächst eine



**Abbildung 1:**  $I_{e,WB}$  (dunkelgrau),  $I_{bw}$  (hellgrau),  $I_{res}$  (schwarz).

Gruppierung zwischen störungsfreien NB- ( $I_{bw} \approx 36$  für G.711) und WB-Kodierern ( $I_{bw} \rightarrow 0$  für PCM) erkennen. Der perzeptive Unterschied zwischen NB und WB ist demnach durch Gl. (3) nachweislich bestätigt und erfüllt Gl. (2). Folglich können zusätzliche (nicht-lineare) Störungen durch spezifische NB-Kodierer unter dem Residualanteil  $I_{res}$  zusammengefasst werden, der perzeptive Effekte verschiedener Natur widerspiegeln kann (abhängig z.B. vom Kodierprinzip und der Bitrate). Die Schätzung für PCM (50-7000 Hz) ist hier nicht exakt, da Gl. (3) unter der Annahme abgeleitet wurde, dass ein Bandpass von 200-7000 Hz die beste Qualität darstellt [3].

Der Bandbreiteneinfluss von WB-Kodierern ist i.a. gering, zeigt jedoch eine deutlichere Abhängigkeit vom verwendeten Kodierverfahren und der Bitrate als im Schmalbandfall. Z.B. ist der Bandbreiteneinfluss von G.722.2 Kodierern größer als für die anderen Kodierer. Die negative Residualkomponente reflektiert den Qualitätsvorteil aufgrund der Nichtlinearität des Kodierers, der die Bandbreiteneinschränkung teilweise kompensiert. In einem zweiten Schritt wurden 24 Kombinationen (Tandems) der Kodierer untersucht. Wie bereits erwähnt

schlägt der Versuch, die zugehörigen Werte  $I_{e,WB}$  zweier in Reihe geschalteter Kodierer auf der  $R$ -Skala zu summieren, in vielen Fällen fehl (s. z.B. [4]). Eine Fehlerquelle scheint die implizite Addition der linearen Störungsanteile zu sein. In der Tat ergibt die Analyse der Faktoren  $I_{bw,1}$  und  $I_{bw,2}$  zweier in Reihe geschalteten Kodierer 1 und 2, dass der Bandbreiteneinfluss in erster Näherung durch den Kodierer mit kleinerer Bandbreite bestimmt ist, also  $I_{bw,Tandem} \approx \max\{I_{bw,1}, I_{bw,2}\}$ .

Diese Beziehung ist intuitiv nachvollziehbar, denn die Frequenzgangverzerrung eines Kodiertandems entspricht sicherlich nicht der Summer der Bandbreiten-Impairments der Einzelkodierer. D.h., die Summe  $I_{e,WB,1} + I_{e,WB,2}$  erzeugt einen minimalen Fehler von  $\min\{I_{bw,1}, I_{bw,2}\}$  (unter der Annahme, dass keine Interaktionen zwischen den Komponenten  $I_{bw,1}$  und  $I_{bw,2}$  vorhanden sind). Allerdings führt eine Subtraktion des Fehlerterms von  $I_{e,WB,1} + I_{e,WB,2}$  lediglich zu einer Fehlerreduktion in den meisten Fällen, was darauf schließen lässt, dass  $I_{res}$  weitere perzeptive Dimensionen enthält, für die eine Additivität nicht unbedingt korrekt sein muss.

## Zusammenfassung

In diesem Beitrag wurde ein neuer E-Modell Impairment Faktor, der *Bandwidth Impairment Factor*  $I_{bw}$  vorgestellt, mit dessen Hilfe lineare Verzerrungen einer Übertragungsstrecke auf plausible Weise erfasst werden können. Für die Integration in das E-Modell werden konventionelle Equipment Impairment Faktoren zerlegt in einen linearen und einen Residualanteil, wobei der lineare Anteil von  $I_{bw}$  abgebildet wird. Die vorgeschlagene Erweiterung des E-Modells bildet somit die Basis, um Ende-zu-Ende Sprachqualität inklusive der Frequenzverzerrungen von Kodierern, Kanalfiltern und Endgeräten vorhersagen zu können.

Diese Studie wurde unterstützt von der Deutschen Forschungsgemeinschaft (DFG), MO1038/5-2.

## Literatur

- [1] ITU-T Rec. G.107: The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, CH-Geneva, 2005.
- [2] Moore, B.C.J.; Tan, C.T.: Perceived Naturalness of Spectrally Distorted Speech and Music. In: *J. Acoust. Soc. Am.*, 114(1), 408-419, 2003.
- [3] Raake, A.: *Speech Quality of VoIP – Assessment and Prediction*. Wiley, UK-Chichester, West Sussex, 2006.
- [4] 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.
- [5] Zwicker, E.; Fastl, H.: *Psychoacoustics – Facts and Models*. Springer, D-Berlin, 1999.
- [6] Scholz, K.; Wältermann, M.; Huo, L.; Raake, A.; Heute, U.; Möller, S.: Estimation of the Quality Dimension “Directness/Frequency Content” for the Instrumental Assessment of Speech Quality. In: *Proc. ICSLP 2006*, USA-Pittsburgh PA, 2006, 1523-1526.