

# Netzwerkplanung für die Sprachübertragung im Vollband

Sebastian Möller<sup>1,2</sup>, Gabriel Mittag<sup>1</sup>

<sup>1</sup> *Quality and Usability Lab, TU Berlin, E-Mail: sebastian.moeller@tu-berlin.de; gabriel.mittag@tu-berlin.de*

<sup>2</sup> *Speech and Language Technology, DFKI Berlin*

## Einleitung

Durch die Einführung komplett IP-basierter Übertragung rückt die Sprachkommunikation über den gesamten hörbaren Bereich (0-20000 Hz, Vollband) in Reichweite. Frühere Untersuchungen [1] haben gezeigt, dass sich dadurch eine um ca. 48% höhere Gesamtqualität erzielen lässt gegenüber dem klassischen Telefonieband (300-3400 Hz, Schmalband). Allerdings verursachen Kodierer, Paketverluste und Verzögerungszeiten Beeinträchtigungen der Gesamtqualität, die eine auf Qualität zielende Netzwerkplanung notwendig machen.

In der Studiengruppe 12 der International Telecommunication Union (ITU-T) wurde deshalb kürzlich die erste Vollband-Version eines Netzwerkplanungsmodells, des sog. E-Modells, standardisiert. Aufbauend auf den Ergebnissen, die bei der DAGA 2018 vorgestellt wurden [1], berücksichtigt dieses Modell auch die Effekte von EVS-Sprachkodierern, Beeinträchtigungen durch Paketverluste, sowie die Störung des Konversationsverhaltens durch Verzögerungen.

Im Beitrag wird zunächst die Struktur des E-Modells erläutert. Anschließend wird die Modellierung der Beeinträchtigungen von Kodierverzerrungen, von Paketverlusten, und des Einflusses von Verzögerungen auf die Konversationsqualität beschrieben. Die Modellierungen werden jeweils zu Ergebnissen von Hör- bzw. Konversationstests in Bezug gesetzt. Für den Einfluss von Verzögerungen werden experimentelle Ergebnisse vorgelegt, die zeigen, dass das nun standardisierte E-Modell recht konservative Vorhersagen liefert. Möglichkeiten der Erweiterung des Modells um akustische Schnittstellen sowie um weitere Konversations-Störungen werden diskutiert.

## Struktur des E-Modells

Ziel eines Netzwerkplanungsmodells wie des E-Modells ist es, auf Basis messbarer Parameter eines Sprachübertragungsnetzes eine Abschätzung der für ein Gespräch erzielbaren Konversationsqualität zu berechnen. Die Parameter des E-Modells 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, elektrisches Leitungsrauschen) mittels eines A- oder psophometrisch gewichteten Pegels dargestellt. Nichtlineares Verhalten, wie bspw. Kodierverzerrungen und durch Paketverluste hervorgerufene Unterbrechungen, werden wiederum durch sog. Equipment-Impairment-Faktoren oder Paketverlustraten 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 transformiert. Beim E-Modell bedient man sich dazu der sog. Transmission-Rating-Skala  $R$ , die im Schmalband-Fall von  $R_{max} = 100$  (optimale Qualität) bis 0 (schlechteste Qualität) reicht. Für die Breitband-Version des E-Modells wurde diese Skala so erweitert, dass die maximal erzielbare Qualität einer Breitband-Verbindung (100-7000 Hz) einem Wert  $R_{max} = 129$  entspricht, und eine optimale Schmalband-Verbindung weiterhin einen Wert  $R_{max} = 100$  erreicht. Für super-breitbandige Sprachübertragung wurde in [1] durch Vergleich mit den Ergebnissen von 3 Testdatenbanken gezeigt, dass ein Wert von  $R_{max} = 148$  realistisch ist; wiederum wurde dieser Wert so bestimmt, dass die Werte von  $R_{max} = 129$  für Breitband und  $R_{max} = 100$  für Schmalband erhalten bleiben. Weitere Untersuchungen von NTT [2] zeigten, dass es keinen signifikanten Unterschied zwischen den Bewertungen von (ansonsten ungestörten) super-breitbandigen und vollbandigen Verbindungen gab. Daher wird auch für das Vollband-E-Modell ein maximaler Wert  $R_{max} = 148$  angenommen.

Die je nach Übertragungsbandbreite maximal erzielbare Qualität wird nun durch verschiedene Beeinträchtigungen gemindert. Zur Berechnung der Stärke der jeweiligen Beeinträchtigung 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 das Gesamt-Transmission-Rating  $R$ , indem man vom Maximalwert  $R_0$  die weiteren Beeinträchtigungen (nichtlineare Verzerrungen, Verzögerungen, Echos, etc.) in Form sogenannter Beeinträchtigungsfaktoren abzieht. Im Fall des Schmalband-E-Modells [3] wird der Maximalwert  $R_0$  zunächst aus den Parametern der Übertragungsstrecke, die die Signalabschwächung und die Rauschleistungen beschreiben, berechnet, sodass sich bei Abwesenheit von rauschen und optimaler Lautstärke ein Wert  $R_0 = R_{max} = 100$  ergibt. Im Falle des Breitband-E-Modells [4] und des Vollband-E-Modells [5] wird hier derzeit ein fester Wert  $R_0, WB = 129$  bzw.  $R_0, FB = 148$  angenommen. Die weiteren Beeinträchtigungsfaktoren werden nach den in [3], [4] und [5] angegebenen Formeln berechnet.

Das Gesamt-Transmission-Rating  $R$  ist zunächst ein Planungswert, der nicht direkt mit subjektiven Urteilen vergleichbar ist.  $R$  lässt sich allerdings mittels einer S-förmigen (fast-)monotonen Beziehung in einen geschätzten *Mean Opinion Score*,  $MOS$ , transformieren, wie man ihn mittels eines Hör- oder Konversationstests als Mittelwert absoluter Kategorien-Bewertungen einer Anzahl von Versuchspersonen erhalten würde. Hierzu wird  $R$  zunächst durch den Maximalwert  $R_{max}$  geteilt und anschließend in einen geschätzten  $MOS_{CQE}$  transformiert:

$$\text{Für } R_x < 0: \quad MOS_{CQE} = 1$$

$$\text{Für } 0 < R_x < 100: \quad MOS_{CQE} = 1 + 0,035 R_x + R_x (R_x - 60) \cdot$$

$$(100 - Rx) \cdot 7 \cdot 10^{-6} \quad (1)$$

$$\text{Für } Rx > 100: \quad MOS_{CQE} = 4,5$$

mit

$$Rx = R/R_{max} \quad (2)$$

## Modellierung der Beeinträchtigung von Kodierverzerrungen und Paketverlusten

Kodierverzerrungen werden im E-Modell als Einzahlwerte, sog. Equipment-Impairment-Faktoren  $I_e$  modelliert, die direkt die Beeinträchtigung durch den jeweiligen Kodierer auf der Transmission-Rating-Skala repräsentieren. Diese Einzahlwerte lassen sich entweder aus den Ergebnissen von Hörversuchen extrahieren, bei denen der entsprechende Kodierer zusammen mit weiteren, bereits bzgl. ihres  $I_e$ -Wertes bekannten Kodierern, in Form eines MOS bewertet wurden. Hierbei werden die MOS-Werte zunächst mittels einer Inversen zu Formel (1) in  $R$ -Werte transformiert, und anschließend auf der  $R$ -Skala linear interpoliert. Alternativ können auch anstelle der Hörtest-Ergebnisse Vorhersagen signalbasierter Modelle wie der ITU-T-Standards P.862 [6] oder P.863 [7] verwendet werden. Details der dazu notwendigen Verfahren sind für den Schmalband- bzw. Breitband-Fall in [8], [9], [10] und [11] beschrieben.

Je nach verwendetem Kodierer und damit implementierter Paketverlust-Behandlung wirken sich Paketverluste unterschiedlich auf die wahrgenommene Qualität aus. Zur Quantifizierung der damit einhergehenden Beeinträchtigung werden im E-Modell zwei Parameter verwendet: Zum einen die Paketverlustrate  $Ppl$  (in %), und zum anderen der Parameter  $Bpl$ , der die Robustheit des Kodierers und der implementierten Fehlerbehandlung gegenüber Paketverlusten beschreibt. Aus diesen Parametern sowie dem Basiswert  $I_e$  wird die Gesamt-Beeinträchtigung durch Kodierer und Paketverluste auf der  $R$ -Skala im Schmalband- und Breitband-E-Modell wie folgt quantifiziert:

$$I_{e,eff} = I_e + (95 - I_e) \cdot \frac{Ppl}{Ppl + Bpl} \quad (3)$$

Das beschriebene Verfahren soll nun auf den Vollband-Fall erweitert werden. Hierzu stehen zunächst die beiden Datenbanken von Orange (O2) und Rohde & Schwarz (R&S) zur Verfügung, die bereits in [1] zur Berechnung des maximalen Wertes  $R_{max}$  verwendet wurden. Daneben konnten Ergebnisse von 3 weiteren Datenbanken mit einbezogen werden:

- Orange (O1): Dieser Hörversuch beinhaltet französische Sprachproben, die mit breitbandigen sowie super-breitbandigen Kodierern sowie teilweise Paketverlusten gestört waren. Die insgesamt 20 Testkonditionen wurden von 24 Probanden diotisch abgehört und auf der 5-stufigen MOS-Skala bewertet.
- Qualcomm (QC): Dieser Hörversuch verwendete Sprachproben in amerikanischem Englisch, welche durch EVS-Kodierer bei verschiedenen Bitraten gestört waren. Die 49 Testkonditionen wurden von 32 Probanden bzgl. MOS bewertet, und die Ergebnisse sind in [12] beschrieben.
- NTT: Dieser Hörversuch beinhaltet 37 japanische Testkonditionen mit Kodierern und signalkorreliertem

Rauschen, welche von 24 Probanden bzgl. MOS bewertet wurden. Die Ergebnisse finden sich in [13].

Aus den Ergebnissen der 5 Hörversuchen sollen nun Werte für  $I_e$  für verschiedene EVS-Kodierer abgeleitet werden. Hierzu sind drei Schritte notwendig:

1. Da die  $MOS_{CQE}$  in Gl. (1) auf den Bereich  $MOS_{CQE} \in [1;4,5]$  begrenzt sind, die subjektiven MOS aber im Bereich  $MOS \in [1;5]$  abdecken, muss zunächst eine Begrenzung auf den kleineren Bereich erfolgen:
$$MOS_{norm,i} = \frac{MOS_i - 1}{MOS_{max} - 1} 3,5 + 1 \quad (4)$$
2. Die so normierten MOS-Werte können nun mittels einer Invertierung von Gl. (1) und  $Rx = R/1,48$  auf die  $R$ -Skala transformiert werden.
3. Aus den so berechneten  $R$ -Werten lassen sich nun  $I_e$ -Werte als Differenz zwischen den  $R$ -Werten der kodierten und der unkodierten, ansonsten aber ungestörten super-breitbandigen oder vollbandigen Testkondition berechnen.

Die beschriebene Prozedur wurde auf alle 5 Testdatenbanken angewendet, und die damit erhaltenen Ergebnisse sind in Tab. 1 dargestellt. Da bei der oben beschriebenen Prozedur die lineare Interpolation der Verfahren aus [8], [9], [10] und [11] zunächst mangels bereits definierter  $I_e$ -Werte unterbleiben muss zeigt sich einige Varianz in den erhaltenen  $I_e$ -Werten. Diese Werte wurden daher in Studiengruppe 12 der ITU-T diskutiert, und es wurde vereinbart, dass ein gewichteter Mittelwert zur Definition endgültiger  $I_e$ -Werte verwendet werden soll, wobei die beiden Orange-Datenbanken eine Gewichtung von  $\frac{1}{2}$  gegenüber den anderen Datenbanken bekommen sollten, da dort nur weniger unterschiedliche Störungen abgedeckt waren [14]. Der so berechnete gewichtete Mittelwert ist in der Spalte „Ø“ in Tab. 1 angegeben.

Tabelle 1: Werte für  $I_e$  und  $Bpl$  für EVS-Kodierer bei unterschiedlichen Bitraten, sowie Schätzungen von P.863.

Bitrate	O1	O2	RS	QC	NTT	Ø	P.863	Bpl
9.6	22.4	17.5	28.7	19.4		22.7	34.3	13
13.2	16.4	11.1	21.2	15.4	18.0	17.1	24.8	11.7
16.4	13.9	6.0	12.3	7.6	13.3	10.8	16.3	10.3
24.4	5.1			7.6	11.4	7.2	8.7	11.4
32.0		7.6		10.5	7.4	8.7	17.2	9.3
48.0				11.7	8.7	10.2	2.2	9.6

Um die so erhaltenen  $I_e$ -Werte auf Konsistenz zu überprüfen wurden 32 Sprachproben der Datenbank nach ITU-T Rec. P.501 [15] mit EVS-Kodierern kodiert und dekodiert und die sauberen Quelldaten sowie die gestörten dekodierten Daten als Eingabe zum signalbasierten Modell nach ITU-T Rec. P.863 [16] verwendet. Die vom signalbasierten Modell geschätzten MOS-Werte wurden ebenfalls nach der o.a. Prozedur in  $I_e$ -Werte transformiert, und die Mittelwerte über alle Sprachproben sind in der Spalte „P.863“ in Tab. 1 angegeben. Die mit diesem Verfahren geschätzten  $I_e$ -Werte sind allgemein etwas höher als diejenigen, die aus den

Hörversuchen bestimmt wurden; nur bei der höchsten Bitrate bleibt unklar, wieso der Hörversuch hier einen höheren Wert lieferte als bei den beiden darunter liegenden Bitraten.

Für den EVS-Kodierer bei 16,4 kbit/s wurden die P.501-Daten auch durch randomisiert verteilte Paketverluste gestört. Die gestörten und die ungestörten Sprachdaten wurden wiederum als Eingangsdaten für das signalbasierte Modell nach ITU-T Rec. P.863 verwendet, und die damit geschätzten MOS-Werte in  $I_e$  transformiert. Aus den Schätzungen wurde mittels des mit „O“ gekennzeichneten Wertes aus Tab. 1 ein Wert für  $B_{pl}$  für diesen Kodierer geschätzt, wobei als Kriterium eine Minimierung des mittleren quadratischen Fehlers gewählt wurde und die Konstante in Gl. (3) variabel gehalten wurde. Die so erhaltenen Werte sind in Abb. 1 für verschiedene Paketverlustraten eingezeichnet. Ebenfalls eingezeichnet ist die Schätzung des schmalbandigen E-Modells nach Gl. (3) mit einem konstanten Term von  $c = 95$ . Die insbesondere zu höheren Verlustraten beobachteten Differenzen zwischen Modellvorhersagen und aus den P.863-Schätzungen bestimmten Werten lassen sich durch Veränderung der Konstanten minimieren; optimale Übereinstimmung ergibt sich für eine Konstante von  $c = 132$ , d.h. Gl. (3) wird ersetzt durch:

$$I_{e,eff} = I_e + (132 - I_e) \cdot \frac{P_{pl}}{P_{pl} + B_{pl}} \quad (5)$$

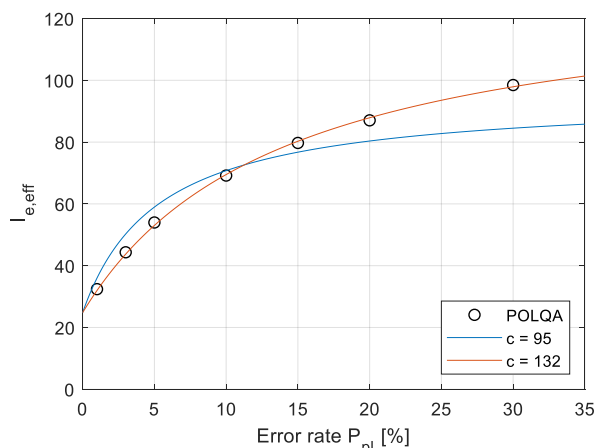


Abbildung 1: Vergleich der aus P.863 geschätzten  $I_{e,eff}$  und den Vorhersagen des E-Modells nach Gl. (2) bzw. Gl. (4).

## Modellierung der Beeinträchtigung von Verzögerungen

Zur Modellierung der Beeinträchtigungen durch reine Verzögerungen (ohne Echo) wurden an der TU Berlin und der University of Western Sydney zwei Konversationstests durchgeführt:

- TUB 1: Dieser Konversationstest wurde mit einem superbreitbandigen System durchgeführt, bei dem sich die Ende-zu-Ende-Verzögerung in 8 Schritten zwischen 0 und 2500 ms einstellen ließ und die Probanden Stereo-Headsets trugen, um akustische Echos zu vermeiden [17]. 24 Probanden führten strukturierte Konversationen durch, die mittels sog. Short-Conversation-Test-Szenarien [18] vorgegeben wurden (bspw. Bestellung einer Pizza oder Vereinbarung eines Arzttermins). Sie bewerteten anschließend die Gesamtqualität der Konversation auf der 5-stufigen MOS-Skala.

- TUB 2: Der zweite Konversationstest wurde von Uhrig und Michael an der University of Western Sydney in Australien durchgeführt [19]. Es wurden wiederum SCT-Szenarien verwendet, diesmal auf Englisch. Das verwendete Kommunikationssystem arbeitete im Vollband und erlaubte die Variation der Ende-zu-Ende-Verzögerung im Bereich 0 bis 1600 ms. Die 20 Probanden verwendeten monaurale Headsets und bewerteten die Konversationsqualität nach jeder Konversation.

Das Standard-E-Modell verwendet einen Beeinträchtigungsfaktor  $I_{dd}$  zur Modellierung der Beeinträchtigung durch Verzögerungen  $T_a$ :

$$\text{Für } T_a \leq 100 \text{ ms: } I_{dd} = 0 \quad (6a)$$

Für  $T_a > 100$  ms:

$$I_{dd} = 25 \left\{ (1 + X^6)^{\frac{1}{6}} - 3 \left( 1 + \left[ \frac{X}{3} \right]^6 \right)^{\frac{1}{6}} + 2 \right\} \quad (6b)$$

mit: 
$$X = \frac{\log \frac{T_a}{100}}{\log(2)} \quad (7)$$

Abb. 2 zeigt einen Vergleich zwischen den Modellvorhersagen und den MOS-Werten aus den beiden Konversationstests. Hierzu wurden die MOS-Ergebnisse aus beiden Tests zunächst soweit linear verschoben, dass sich für die verzögerungsfreie Verbindung ( $T_a = 0$  ms) ein MOS-Wert von 4,5 ergibt, wie ihn das E-Modell in Gl. (2) implizit annimmt. Es zeigt sich eine recht gute Übereinstimmung zwischen Modellvorhersagen und Konversationstest-Ergebnissen für den Bereich  $T_a = 0 \dots 1700$  ms; für höhere Verzögerungen, wie sie in der Datenbank TUB1 auftauchen, sind die Vorhersagen zu optimistisch.

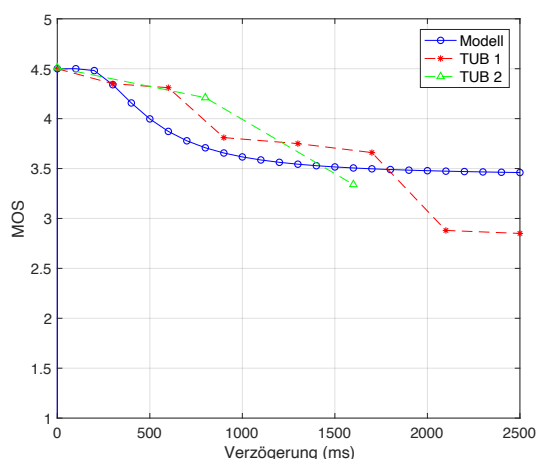


Abbildung 2: Vergleich der Modellvorhersagen nach Gl. (6a) – (7) mit den Ergebnissen der Konversationstests.

Trotz der guten Übereinstimmung der Modellvorhersagen mit den subjektiven Testergebnissen entschloss sich die Studiengruppe 12 der ITU-T, die Vorhersagen des Modells pessimistischer anzulegen: Hierzu wurde ein Vorfaktor 1,48 in die Vollband-Variante eingefügt, die die größere Spannbreite der  $R$ -Werte bei Vollband ( $[0; 148]$ ) gegenüber Schmalband ( $[0; 100]$ ) ausgleichen soll. Damit ändern sich Gleichungen (6a) und (6b) wie folgt:

Für  $T_a \leq 100$  ms:

$$I_{dd} = 0 \quad (8a)$$

Für  $Ta > 100$  ms:

$$I_{dd} = 1,48 \cdot 25 \left\{ (1 + X^6)^{\frac{1}{6}} - 3 \left( 1 + \left[ \frac{X}{3} \right]^6 \right)^{\frac{1}{6}} + 2 \right\} \quad (8b)$$

## Diskussion

Das hier vorgestellte E-Modell erlaubt eine Abschätzung der Einflüsse von Sprachkodierern, Paketverlusten und Verzögerungen auf die Konversationsqualität bei vollbandiger Sprachübertragung. Derzeit gibt es aber noch signifikante Einschränkungen gegenüber der schmalbandigen Version des E-Modells:

- Der Wert für  $R_0$  ist im Vollband-Modell fest bei 148; in ihm werden nicht, wie im Schmalband-Fall, die Einflüsse von Abschwächungen und Rauschen (Hintergrund- und Leistungsrauschen) abgebildet. Zur Modellierung dieser Effekte sind weitere Hör- oder Konversationstests mit unterschiedlichen Abschwächungen und Geräuschen notwendig. Da die Endgeräte sich bei vollbandiger Kommunikation voraussichtlich signifikant von Standard-Handapparaten unterscheiden werden und da diese Endgeräte voraussichtlich auch Mechanismen zur Geräuschunterdrückung beinhalten werden, ist auf eine adäquate und gemeinsame Betrachtung der Endgerät-Hintergrundgeräusch-Raum-Konstellation zu achten.
- Die Vorhersagen für Verzögerungen beschränken sich auf ansonsten saubere, echofreie Verbindungen. Jedoch ist insbesondere bei Freisprechern oder bei schlechten akustischen Ankopplungen von Reflexionen auszugehen, die zu Sprecher- oder Hörerechos führen können. Diese werden in der Vollband-Version des E-Modells bislang nicht behandelt. Auch die Effekte von Echounterdrückern werden bislang nicht berücksichtigt.
- Die Vorhersage der Effekte von Kodierverzerrungen beschränken sich bislang auf den EVS-Kodierer bei 6 unterschiedlichen Bitraten, und für Paketverluste auf eine Bitrate dieses Kodierers. Auch werden bislang nur randomisiert auftretende – und nicht in Bursts auftretende – Paketverluste modelliert.

Trotz der genannten Einschränkungen wurde das vollbandige E-Modell für hinreichend hilfreich erachtet, sodass es im vergangenen Jahr als internationaler Standard ITU-T Rec. G.107.2 [5] verabschiedet wurde.

## Literatur

- [1] Möller, S., Mittag, G., Köster, F., Barriac, V., Côté, N.: Erweiterung der Netzwerkplanung auf superbreitbandige Sprachübertragung, in: Fortschritte der Akustik – DAGA 2018: Plenarvortr. u. Fachbeitr. d. 44. Dtsch. Jahrestg. f. Akust., DE-München, Dtsch. Ges. Akust., Berlin, 2018.
- [2] ITU-T Contr. SG12-C260: Fullband extension of R-value, Source: Nippon Telegraph and Telephone Corporation (NTT), Int. Telecomm. Union, Genf, 2018.
- [3] ITU-T Recommendation G.107: The E-model: a computational model for use in transmission planning, Int. Telecomm. Union, Genf, 2015.

- [4] ITU-T Recommendation G.107.1: Wideband E-model, Int. Telecomm. Union, Genf, 2019.
- [5] ITU-T Recommendation G.107.2: Fullband E-model, Int. Telecomm. Union, Genf, 2019.
- [6] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Int. Telecomm. Union, Genf, 2001.
- [7] ITU-T Recommendation P.863: Perceptual objective listening quality prediction, Int. Telecomm. Union, Genf, 2018.
- [8] ITU-T Recommendation P.833: Methodology for derivation of equipment impairment factors from subjective listening-only tests, Int. Telecomm. Union, Genf, 2001.
- [9] ITU-T Recommendation P.833.1: Methodology for the derivation of equipment impairment factors from subjective listening-only tests for wideband speech codecs, Int. Telecomm. Union, Genf, 2009.
- [10] ITU-T Recommendation P.834: Methodology for derivation of equipment impairment factors from instrumental models, Int. Telecomm. Union, Genf, 2015.
- [11] ITU-T Recommendation P.834.1: Methodology for derivation of equipment impairment factors from instrumental models for wideband speech codecs, Int. Telecomm. Union, Genf, 2015.
- [12] ITU-T Recommendation P.Imp863: Implementer's Guides for ITU-T P.863, Int. Telecomm. Union, Genf, 2016.
- [13] ITU-T Contr. SG12-C158: Equipment impairment factors for EVS codecs, Source: Nippon Telegraph and Telephone Corporation (NTT), Int. Telecomm. Union, Genf, 2018.
- [14] ITU-T Contr. SG12-C306: Merged results of EVS-SWB equipment impairment factor estimations, Source: Deutsche Telekom AG, Orange, Int. Telecomm. Union, Genf, 2018.
- [15] ITU-T Recommendation ITU-T P.501: Test signals for use in telephony, Int. Telecomm. Union, Genf, 2017.
- [16] ITU-T Recommendation P.863: Perceptual objective listening quality prediction, Int. Telecomm. Union, Genf, 2018.
- [17] Köster, F.: Multidimensional analysis of conversational telephone speech, Springer Nature Singapore Pte Ltd., Singapore, 2018.
- [18] ITU-T Recommendation ITU-T P.805: Subjective evaluation of conversational quality, Int. Telecomm. Union, Genf, 2007.
- [19] Uhrig, S., Michael, T., Möller, S., Keller, P., Voigt-Antons, J.-N.: Effects of delay on perceived quality, behavior and oscillatory brain activity in dyadic telephone conversations, in: Proc. 10th Int. Conf. on Quality of Multimedia Experience (QoMEX 2018), 29 - 31 May 2018, Sardinia.