Sprachqualitäts-Modellierung: Von Konversations-Phasen über Gespräche bis zur Dienstqualität

Sebastian Möller, Dennis Guse, Friedemann Köster, Benjamin Weiss

Quality and Usability Lab, TU Berlin, E-Mail: sebastian.moeller@tu-berlin.de; dennis.guse@alumni.tu-berlin.de; friedemann.koester@tu-berlin.de; benjamin.weiss@tu-berlin.de

Einleitung

Bei der Messung und Vorhersage von Telefon-Sprachqualität beschränkt man sich zumeist auf die Betrachtung der reinen Hörsituation. Durch Hörversuche können sowohl die Gesamtqualität einer Sprachprobe als auch einzelne perzeptive Dimensionen quantifiziert, und deren Ergebnisse anschließend durch Modelle geschätzt werden. Allerdings ist unklar, wie diese nur für die Hörsituation relevanten Messungen tatsächlich mit der Qualität eines interaktiven Gespräches, oder gar mit der Qualität des gesamten Dienstes, zusammenhängen.

In diesem Beitrag soll eine Kette von Modellen zur Erfassung der verschiedenen Beiträge gebildet werden. Ausgehend von perzeptiven Dimensionen der Hörqualität, wie sie bspw. von Wältermann bestimmt wurden, sowie von Dimensionen der Sprech- und Interaktionsqualität, welche Köster [6] kürzlich analysierte, wird zunächst ein multidimensionaler perzeptiver Raum für die Einzelphasen (Hören, Sprechen, Interagieren) aufgespannt. Für jede dieser Phasen kann ein separater Qualitätsschätzer gebildet werden. Um den Einfluss der einzelnen Phasen auf die Gesamtqualität zu gewichten bedarf es darüber hinaus einer Simulation von Konversations-Verhalten; diese könnte bspw. mittels Agenda-basierter Modelle aus der Mensch-Maschine-Interaktion geschehen. Empirische Ergebnisse von Köster [6] sowie Modelle von Weiss [15] zeigen, dass für die Gesprächsqualität neben dem arithmetischen Mittelwert auch die schlechteste Phasen-Qualität entscheidend ist. Beim Telefonie-Dienst kommt es nun zu wiederholten Gesprächen; das Aggregieren der Gesprächsqualität zur wahrgenommenen Dienstqualität kann dann durch langzeitige Mittelung und Gewichtung erfolgen. Hierzu wurden von Guse et al. [4] bereits Modelle vorgestellt. In der Summe ergibt sich somit eine Kette von Vorhersagemodellen von Einzelphasen zum gesamten Dienst. Im Beitrag werden hierfür Ideen aufgezeigt und Forschungslücken benannt.

Phasen der Sprachkommunikation

Eine sprachliche Kommunikationssituation zwischen zwei Personen lässt sich aus Sicht eines Gesprächspartners prinzipiell in drei Phasen einteilen [11]: Hören, Sprechen, und Interagieren, d.h. der (u.U. mehrfache) Wechsel zwischen Hören und Sprechen. Diese Phasen entsprechen 4 Zuständen: (1) Teilnehmer A spricht und Teilnehmer B nicht; (2) Teilnehmer B spricht und Teilnehmer A nicht; (3) beide Teilnehmer sprechen gleichzeitig (sog. *Double Talk*); oder (4) kein Teilnehmer spricht (sog. *Mutual Silence*). Die Phase „Interagieren“ ergibt sich also aus dem Wechsel zwischen diesen Zuständen, d.h. den Zustandsübergängen zwischen (1) und (2) bzw. umgekehrt, entweder direkt oder über die Zustände (3) oder (4). In der Literatur sind verschiedene Ansätze zur Beschreibung der Zustandsübergänge bekannt, z.B. eine *Speaker Alternation Rate*, eine *Conversational Temperature*, etc. Ein Überblick findet sich bspw. bei Schoenenberg [12].

Die oben genannten Phasen sind jedoch nicht wahrnehmungsbezogen. Der Gesprächsteilnehmer kann zwar während aller Phasen prinzipiell wahrnehmen und daher die Qualität beurteilen, allerdings unterschiedlich gut, da die eigene Sprechaktivität kognitive Ressourcen bindet. Urteile in der Sprech- und Interaktionsphase werden daher i. Allg. nicht sehr analytisch sein können. Eine wahrnehmungsbezogene Betrachtungsweise ist aber für die diagnostische Bestimmung der Telefon-Sprachqualität wichtig, da unterschiedliche (physikalische bzw. algorithmische) Störungen des Übertragungskanals sich teilweise nur in einzelnen Phasen auswirken (bspw. ein Echo nur beim Sprechen, oder eine Verzögerung nur beim Interagieren), und in diesen Phasen die Bewertung ebenjener Störung vollzogen werden muss – u.U. auch bei eingeschränkten kognitiven Ressourcen. Kenntnisse der perzeptiven Eigenschaften, die in einzelnen Phasen vorherrschen, sowie Kenntnisse über ihre Wertigkeit bei der Bildung eines Gesamt-Qualitätsurteils für ein Gespräch, sind daher wichtig für die realistische Abschätzung der Konversationsqualität.

Zur Untersuchung wahrnehmungsbezogener Größen ist es zunächst notwendig, die Dimensionen des Wahrnehmungsraumes zu bestimmen, d.h. den Raum zu vermessen. Hierzu werden üblicherweise Methoden der multidimensionalen Analyse verwendet (Ähnlichkeitsbewertung und MDS oder Semantisches Differenzial und Hauptkomponentenanalyse), allerdings wurden diese bislang unseres Wissens nach vor allem in passiven Situationen verwendet. Für die Sprech- und die Interaktionsphase entwickelten Köster und Möller [7] daher ein neues Verfahren, welches zunächst das Vorhandensein der drei Phasen postuliert und getrennt für jede Phase einen multidimensionalen Raum aufspannt. Es ergaben sich 4 Dimensionen für die Hörsituation (basierend auf Wältermann: *Coloration*, *Discontinuity*, *Noisiness* und *Non-optimum Loudness*; letztere ist zu den anderen dreien allerdings nicht orthogonal), 2 Dimensionen für die Sprechsituation (*Impact on Speaking*, *Degradation of One's Own Voice*), und eine Dimension in der Interaktionsphase (*Interactivity*). Beim gemeinsamen Testen aller drei Phasen in einer realen Konversation fallen dann einige dieser

perzeptiven Dimensionen wieder zusammen, wahrscheinlich da nicht ausreichend Ressourcen zur gleichzeitigen Trennung aller 7 Dimensionen zur Verfügung stehen.

Bislang wurden nur zeitlich stabile Störungen betrachtet, welche über die Dauer der jeweiligen Phase oder des gesamten Gespräches konstant bleiben. Insbesondere bei mobilen und IP-vermittelten Gesprächen variieren die Störungen (und damit auch die Qualität) über der Zeit. Bittet man einen Gesprächsteilnehmer zum Abschluss eines Gespräches dann um ein Qualitätsurteil, so wird sich dieses wiederum aus den einzelnen Wahrnehmungen während des Gespräches – und den dabei erfahrenen Phasen – zusammensetzen. Hierfür wurden in der Vergangenheit sog. Call-Quality-Modelle entwickelt, welche zumeist eine zeitliche Mittelung, verbunden mit einer stärkeren Gewichtung der nahe am Beurteilungszeitpunkt liegenden Abschnitte (sog. *Recency Effect*) sowie der stärker beeinträchtigten Abschnitte (sog. *Peak Rule*), vorsehen, vgl. Weiss et al. [15]. Diese Modelle wurden auf Basis subjektiver Qualitätsurteile gebildet, welche in einer simulierten Gesprächssituation erfasst wurden; die Probanden hatten dabei zumeist eine Höraufgabe, welche nur durch die Abfrage einzelner inhaltlicher Information durch eine Sprechaufgabe unterbrochen wurde. Eine wirkliche selbstbestimmte Interaktivität stellte sich bei diesen Versuchen nicht ein.

Im vorliegenden Paper sollen Ideen präsentiert werden, wie sich die Gesamtqualität einer realistischen Konversationssituation auf Basis der Qualitäten der einzelnen Phasen vorhersagen lassen könnte, und wie sich die Qualität der Konversation dann über mehrere zeitlich aufeinander folgende Nutzungen zu einer Gesamtqualität eines Dienstes integrieren lässt. Dabei stehen Rechenmodelle im Vordergrund, mit deren Hilfe sich Qualität aus instrumentell messbaren Größen (zumeist Signalen oder Parametern) vorhersagen lässt. In den folgenden Abschnitten werden zunächst Vorhersagemodelle für die 3 Phasen und die dabei relevanten perzeptiven Dimensionen betrachtet. Daran schließt sich eine kurze Diskussion der Integration der drei Phasen zur Konversationsqualität, sowie der zeitlichen Integration über mehrere Konversationen hin zu einer Dienstqualität an. Das Paper schließt mit einer Einordnung und einem Ausblick auf zukünftige Arbeiten.

Vorhersagemodelle für die Hörsituation

Bereits seit langer Zeit existieren Modelle zur Schätzung der Gesamtqualität in der Hörphase. Hierbei wird zumeist ein perzeptiv motivierter Vergleich zwischen (ungestörtem) Eingangs- und (gestörtem) Ausgangssignal einer Übertragungsstrecke durchgeführt, und das Ergebnis dieses Vergleiches – also ein perzeptiv motivierter Abstand – dann auf ein mittleres Qualitätsurteil auf einer 5-stufigen Skala (*Mean Opinion Score*, MOS) transformiert. Dabei müssen beide Signale zunächst zeitlich aufeinandergelegt werden (was bei zeitlich variierenden Verzögerungen ein Problem darstellen kann), und perzeptiv weniger relevante Unterschiede (bspw. leichte Lautheitsunterschiede) müssen vor dem Vergleich ausgeglichen werden. Das hierzu derzeit in ITU-T Rec. P.863 standardisierte Modelle (POLQA) erfasst eine Vielzahl praktisch relevanter Störungen des Übertragungskanals sowie der Endgeräte, sowohl für den Schmalband- (ohne Endgeräte) als auch für den Super-Breitband-Fall. Die Korrelationen zwischen auditiv bestimmten und geschätzten MOS-Werten betragen dabei üblicherweise über 0,90, teilweise über 0,93 (bei schmalbandigen Daten), bestimmt auf unabhängigen Testdatenbanken. Auch ohne ein ungestörtes Referenzsignal (d.h. sogenannt nicht-intrusiv) erlaubt das Modell aus ITU-T Rec. P.563 eine Vorhersage für den Schmalbandfall. Zur Anpassung des Modells wurden hier allerdings Schätzungen des intrusiven Vorgänger-Modells nach ITU-T Rec. P.862 (PESQ) als Referenz genommen. Eine nicht-intrusive Variante für super-breitbandige Sprache ist derzeit in Planung (ITU-T SG12 Study Item P.SPELQ). Auf Basis von zuvor bestimmten oder angenommenen Planungs-parametern erlaubt das E-Modell (ITU-T Rec. G.107) ebenfalls die Vorhersage der Hörqualität für Handapparat-basierte Gespräche, sofern die entsprechenden Parameter der Sprechphase und der Interaktion (Rückhören, Echo, Verzögerungen) auf Optimalwerte gesetzt werden. Die dabei erzielten Korrelationen sind üblicherweise geringer als bei signalbasierten Modellen, da mit Planungswerten gerechnet wird, und keine Informationen über die tatsächlich am Ohr des Zuhörers ankommenden Sprachsignale vorliegen.

Neben der Gesamtqualität lassen sich auch einzelne perzeptive Dimensionen der Hörphase schätzen. Côté entwickelte hierfür das signalbasierte Modell DIAL [2], welches mit Referenzsignal arbeitet und insbesondere für *Coloration* sehr gute Korrelationen (> 0,95) liefert. Die Korrelationen für *Noisiness* und *Discontinuity* lagen allerdings etwas niedriger. Auch Scholz [13] und Huo [5] entwickelten hierfür dimensionsbasierte Modelle. Köster et al. entwickelten darüber hinaus nicht-intrusive Modelle ohne Referenzsignal. Die dabei erzielten Korrelationen zwischen auditiven und geschätzten Qualitätsurteilen lagen bei 0,91 für *Noisiness* [8] und bei 0,93 für C*oloration* [10]. Allerdings fehlt noch ein nicht-intrusives Modell für *Discontinuity*.

Vorhersagemodelle für die Sprechphase

Bislang sind nur wenige Vorhersagemodelle für die Sprechsituation bekannt. Das bekannteste signalbasierte Modell ist das von Appel und Beerends [1] entwickelte PESQM, welches an PESQ/POLQA angelehnt ist, allerdings als gestörtes Signal das zurückkommende Signal betrachtet, welches das Sprechersignal überlagert. Auch das parametrische E-Modell berücksichtigt Störungen in der Sprechsituation, die durch nicht-optimales Rückhören oder Sprecherechos begründet sind, und berechnet hieraus einen Beeinträchtigungsfaktor, der in einen MOS-Wert umgerechnet werden kann.

Köster [6] entwickelte darüber hinaus einfache referenzbasierte Maße für die beiden perzeptiven Dimensionen der Sprechsituation. Die Korrelationen sind allerdings aufgrund der einfachen Eingangsgrößen und der dürftigen Datenlage bislang noch nicht zufriedenstellend (< 0,65). Es ist zu erwarten, dass auf einer besseren Datenbasis

aussagekräftige Schätzer auch für diese perzeptiven Dimensionen gebildet werden können.

Vorhersagemodelle für die Interaktionsphase

Das Abwechseln von Sprechen und Hören, wie o.a. als „Interaktionsphase“ bezeichnet, unterliegt sehr komplexen kommunikativen Phänomenen wie der Verhandlung geeigneter Übergabepunkte, welche z.B. durch prosodische Merkmale angedeutet werden können, sowie dem sozialen Verhältnis der Gesprächsteilnehmer. Der genaue Verlauf der Interaktion hängt darüber hinaus stark von Zweck des Gespräches ab, welcher für die Evaluation regelmäßig durch die Definition von Test-Szenarien simuliert wird. Daher hängen auch die Beurteilungsergebnisse hinsichtlich der Qualität der Interaktion stark von den verwendeten Szenarien sowie den Spezifika der Gesprächsteilnehmer ab.

Zur Abschätzung des Einflusses von Verzögerungen auf das Interaktionsverhalten definierte die ITU-T früher feste Grenzwerte (ITU-T Rec. G.114), welche zulässige Einweg-Verzögerungszeiten festlegten. Diese Grenzwerte werden zunehmend abgelöst von parametrischen Modellen des Einflusses auf die Gesamtqualität, wie sie bspw. vom E-Modell getätigt werden. Das Modell geht dabei im Normalfall von einer „freien“ Konversation ohne besonderen Zeitdruck aus, und berücksichtigt (als Planungsmodell) noch einen Sicherheitspuffer. Bei Annahme einer hoch interaktiven Situation wurde eine Verschärfung der Beeinträchtigung mit ins Modell aufgenommen, welche von Raake vorgeschlagen wurde (ITU-T Rec. G.107).

Köster [6] unternahm darüber hinaus eine Schätzung der (einzigen) perzeptiven Dimension „Interaktivität“ dieser Phase. Die Vorhersagekraft ist aufgrund der Simplizität des Ansatzes und der geringen Daten bislang noch sehr moderat.

Aggregation zur Gesamtqualität einer Dienst-Episode

Es stellt sich die Frage, wie sich die Qualitätsurteile der Hör-, der Sprech- und der Interaktionsphase nun zu einer Gesamtqualität für eine Konversation (bspw. einen Anruf) aggregieren lassen. Hierzu bestehen unterschiedliche Möglichkeiten.

Zunächst können die drei Phasen linear kombiniert werden um die Gesamtqualität einer Konversation zu bestimmen, vgl. Köster [6]. Dazu wurde in der Vergangenheit zumeist die Annahme getroffen, dass eine für die Hörsituation bestimmte (subjektiv gemessene oder instrumentell vorhergesagte) Qualität auch für eine Konversationssituation aussagekräftig ist, sofern keine speziellen Beeinträchtigungen durch Rückhören, Echos oder Verzögerung vorliegen. Diese Annahme ist prinzipiell akzeptabel. Allerdings zeigen Vergleiche zwischen Hör- und Konversationstests auch, dass erstere stärker analytisch sind und die zur Verfügung stehende Skala (bspw. die 5-stufige MOS-Skala) von den Versuchspersonen besser ausgenutzt wird, während sich die Urteile in einem Konversationstest zumeist auf den mittleren Skalenbereich fokussieren. Diesem Umstand könnte aber durch eine einfache Transformationsregel (bspw. Kompression der Urteile beim Übergang zwischen Hör- und Konversationstest) Rechnung getragen werden. Die Voraussetzung ist hier aber natürlich, dass keine speziellen Konversations-Beeinträchtigungen vorliegen.

Im parametrischen E-Modell wurde stattdessen explizit versucht, Hör-, Sprech- und Interaktions-Beeinträchtigungen auf einer gemeinsamen Skala zu integrieren. Die dabei verwendete sog. Transmission-Rating-Skala stammt vom Vorgängermodell (von der Fa. Bellcore), und die Anteile der verschiedenen Störungsarten zueinander wurden durch Vermischung unterschiedlicher Testergebnisse eingestellt.

Etwas strukturierter erfolgt die Integration im sog. Call-Quality-Modell. Dabei geht man aber nicht von den erwähnten drei Phasen aus, sondern unterteilt die Konversation in Gesprächsabschnitte von ca. 8s, aus denen sich Gespräche von typischerweise 1-2 Minuten zusammensetzen lassen. Die Abschnitte sind durch die Hörsituation bestimmt, in den (etwa gleich langen) Pausen dazwischen wird eine Versuchsperson aufgefordert, eine Frage mündlich zu beantworten, also zu sprechen; die Qualitätsurteile dieser Abschnitte reflektieren also nur das Hören. Aus den auditiven Bewertungen der Hörsituation lassen sich Gesamtqualitätswerte für die simulierte Konversationssituation schätzen, wobei üblicherweise schlechte Momente und Momente, die zeitlich nahe zum Beurteilungszeitpunkt liegen, stärker gewichtet werden (sog. Recency-Effekt und/oder Peak-End-Rule). Eine solche zeitliche Aggregation erzielt üblicherweise bessere Korrelationen als ein reiner arithmetischer Mittelwert (Beispielwerte aus [9] und [15]: Schmalband: R=0,94, RMSE=0,20; gemischt Schmalband/Breitband: R=0,92, RMSE=0,31). Der gleiche Aggregationsansatz funktioniert auch, wenn man die auditiven Bewertungen der Hörsituation durch instrumentelle Schätzer (z.B. POLQA, E-Modell) ersetzt. Hierbei wurden für den Schmalbandfall gute Übereinstimmungen (Bsp. PESQ: R=0,87, RMSE=0,26), für den gemischten Schmalband-/Breitband-Fall immer noch ordentliche Übereinstimmungen (Bsp. POLQA: R=0,81, RMSE=0,50) berichtet [9].

Eine bessere und sinnvollere Vorhersage wäre möglich, wenn sich neben der Beurteilung auch das Sprech- und Hörverhalten der Gesprächsteilnehmer vorhersagen und ggf. simulieren ließe. Hierzu wäre eine Simulation von Konversationsverhalten unter Berücksichtigung der auftretenden Störungen sinnvoll (bspw. häufigeres Ins-Wort-Fallen bei Verzögerungen, oder Adoption eines Walkie-Talkie-Verhaltens bei starken Verzögerungen; Verlangsamung des Sprechens bei starken Echos). Ein erster Vorschlag für eine solche Simulation wurde an der TU Berlin gemacht, basierend auf Agenda-Modellen, wie sie in der Simulation von Mensch-Maschine-Interaktion Verwendung finden, allerdings wurde er bislang noch nicht validiert.

Sobald eine solche Simulation zur Verfügung steht, lassen sich Schätzer für die Hörphase, die Sprechphase und die Interaktionsphase gemäß ihrer Auftretenshäufigkeit in einem Dialog gewichten und zu einer Gesamtqualitätsschätzung

integrieren. Dabei sind wahrscheinlich wieder Effekte wie die Peak-End-Rule oder der Recency-Effekt zu berücksichtigen. Die entsprechenden Modelle zur Schätzung der Qualität während der einzelnen Interaktionsphasen könnten wiederum signalbasiert oder parametrisch arbeiten.

Aggregation zur Gesamtqualität des Dienstes

Alle oben angegebenen Modelle beschäftigen sich mit der Qualität eines einzelnen Gespräches oder einzelner Phasen hiervon. Telekommunikationsdienste werden aber üblicherweise mehrfach konsekutiv genutzt. Daraus ergibt sich für den Nutzer eine Gesamtqualität eines Dienstes, über mehrere Nutzungs-Episoden hinweg. Ältere Untersuchungen von Bellcore belegen, dass die dabei bewertete Qualität nicht identisch mit der Qualität eines jeden Gespräches sein muss [3]. Wahrscheinlich liegen dem beobachteten Unterschied psychologische Beurteilungs-Verzerrungen zugrunde.

Guse und Möller widmeten sich diesem Effekt und versuchten, multi-episodische Qualitätsbeurteilungen (also „Dienstqualität“) aus den Bewertungen einzelner Nutzungs-Episoden vorherzusagen. Dabei wurde ein zeitlich abklingendes „Vergessen“ schlechter Bewertungen sowie ein zeitliches „Mittelungsfenster“ beobachtet und mittels gewichteter Fensterung modelliert. Die Korrelationen der Modellvorhersagen mit den Dienstqualität-Bewertungen sind jedoch bislang sehr begrenzt, was neben den einfachen Modellierungsansätzen auch durch die schlechte Datenlage begründet sein dürfte. Auch die Reputation des Dienste-Anbieters kann eine Rolle spielen. Darüber hinaus ist bislang offen, wie sich die Bewertungen unterschiedlicher Dienste eines Anbieters gegenseitig beeinflussen.

Diskussion und offene Fragen

Zur Optimierung und Überwachung der Qualität interaktiver Telekommunikationsdienste sind diagnostische Modelle notwendig. In unserem Beitrag wählen wir hierfür einen Ansatz, der perzeptive Dimensionen einzelner Interaktionsphasen schätzt und anschließend zu einem Qualitätsurteil für ein oder mehrere Gespräche integriert. Der Überblick zeigt, dass insbesondere für die Sprech- und die Interaktionsphasen noch gute Vorhersagemodelle fehlen.

Darüber hinaus ist die Simulation von Gesprächsverhalten essentiell. Nur damit lässt sich eine korrekte Integration der unterschiedlichen wahrgenommenen Größen gemäß ihrem (angenommenen) Beitrag zum Gesprächsverlauf erreichen. Für eine solche Simulation ist insbesondere Wissen über kommunikative Effekte notwendig, bspw. *Turn-Taking*, *Grounding*, etc. Eine solche Simulation kann zunächst auf semantischer Ebene geschehen, sollte danach allerdings auch auf der Signalebene erfolgen, um den Einfluss moderner Signalverarbeitung korrekt zu berücksichtigen. Hierzu können moderne Sprachsyntheseverfahren nützlich sein, mit deren Hilfe realistische Signale generiert werden können.

Unabhängig davon ist auch die genaue Untersuchung des zeitlichen Verlaufes von Qualitätsurteilen notwendig. Nur so lassen sich tatsächlich für einen Dienst aussagekräftige Vorhersagen erzielen. Idealerweise lässt sich dies mit (allerdings zeit- und kostenintensiven) Feldstudien über mehrere Tage oder Wochen untersuchen.

Literatur

[1] Appel, R., Beerends, J.: On the quality of hearing one’s own voice. Journal of the Audio Engineering Society, vol. 50, no. 4, S. 237–248, 2002.

[2] Côté, N.: Integral and Diagnostic Intrusive Prediction of Speech Quality. Berlin: Springer, 2011.

[3] Duncanson, J.P.: The average telephone call is better than the average telephone call. In: The Public Opinion Quarterly 33.1, S. 112–116, 1969.

[4] Guse, D., Weiss, B., Möller, S.: Modelling multi-episodic quality perception for different telecommunication services: First insights. In: Sixth Int. Workshop on Quality of Multimedia Experience (QoMEx), Singapore, IEEE, S. 105–110, 2014.

[5] Huo, L.: Attribute-based Speech Quality Assessment - Narrowband and Wideband. Kiel: Shaker Verlag, 2015.

[6] Köster, F.: Multidimensional Analysis of Conversational Telephone Speech, Dissertation (eingereicht), Technische Universität Berlin, 2016.

[7] Köster, F., Möller, S.: Introducing a new test-method for diagnostic speech quality assessment in a conversational situation, in Fortschritte der Akustik DAGA 2016: Plenarvortr. u. Fachbeitr. d. 42. Dtsch. Jahrestg. f. Akust.. Berlin: DEGA, 2016.

[8] Köster, F., Mittag, G., Polzehl, T., Möller, S.: Non-intrusive estimation of noisiness as a perceptual quality dimension of transmitted speech, in Proc. 5th International Workshop on Perceptual Quality of Systems, Berlin, Germany: PQS 2016, pp. 74 – 78.

[9] Lewcio, B. Management of Speech and Video Telephony Quality in Heterogeneous Wireless Networks: Cham: Springer, 2014.

[10] Mittag, G., Köster, F., Möller S.: Non-intrusive estimation of the perceptual dimension coloration, in Fortschritte der Akustik, DAGA 2016: Plenarvortr. u. Fachbeitr. d. 42. Dtsch. Jahrestg. f. Akust., 2016.

[11] Richards, D.S.: Telecommunication by Speech. London: Butterworth, 1973.

[12] Schoenenberg, K.: The Quality of Mediated-Conversations under Transmission Delay, Dissertation, Technische Universität Berlin, 2015.

[13] Scholz, K.: Instrumentelle Qualitätsbeurteilung von Telefonbandsprache beruhend auf Qualitätsattributen. Kiel: Shaker Verlag, 2008.

[14] Wältermann, M.: Dimension-based Quality Modeling of Transmitted Speech. Berlin: Springer, 2012.

[15] Weiss, B., Möller, S., Raake, A., Berger, J., Ullmann, R.: Modeling call quality for time-varying transmission characteristics using simulated conversational structures., In: Acta Acustica united with Acustica 95.6, S. 1140–1151, 2009.