Qualitätsdimensionen bei der Sprachübertragung in modernen Telekommunikationsnetzen

Marcel Wältermann[1], Kirstin Scholz[2], Alexander Raake[3], Ulrich Heute[2], Sebastian Möller[3]

[1] *Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 44780 Bochum, Deutschland*

[2] *Lehrstuhl für Netzwerk- und Systemtheorie, Christian-Albrechts-Universität, Kaiserstr. 2, 24143 Kiel, Deutschland*

[3] *Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Deutschland*

Einleitung

Die Qualität übertragener Sprache wird üblicherweise in Hörversuchen ermittelt, innerhalb derer Versuchspersonen (VPn) unter definierten Bedingungen absolute Qualitätsurteile für dargebotene Sprachproben abgeben. Daraus wird für jede Sprachprobe ein integraler *Mean Opinion Score* (MOS) berechnet, der die Qualität als Einzahlwert widerspiegelt. Welche *perzeptiven Dimensionen* des Sprach-Hörereignisses diesen Urteilen zugrunde liegen ist zunächst unbekannt.

Die Betrachtung der zugrunde liegenden perzeptiven Dimensionen bietet allerdings die Möglichkeit, einen generischen Ansatz zur Beurteilung der Ende-zu-Ende Qualität moderner Telekommunikationsnetze zu entwickeln [1]. Für die Vielzahl von Faktoren, die die Sprachqualität hier beeinflussen – neben verschiedenartigen Endgeräten reichen diese von konventionellen Leitungseigenschaften wie Signal-Rauschabstand bis hin zu zeitvarianten Effekten wie Paketverlusten bei VoIP – können so z.B. perzeptiv motivierte Algorithmen für die Schätzung von Qualitätsurteilen entwickelt werden. Hierfür liefert die vorgestellte Arbeit eine Grundlage, indem die bezüglich der Qualität relevanten Dimensionen identifiziert wurden [2]. Hierzu wurden auditive Experimente mit anschließenden multidimensionalen Analysen durchgeführt, die zwei verschiedenen Paradigmen folgen: Multidimensionale Skalierung und Semantisches Differential. Durch eine lineare Abbildung der resultierenden orthogonalen Dimensionen auf Gesamtqualitätsurteile, die in einem weiteren Versuch gewonnen wurden, können Wichtungsfaktoren der Dimensionen bzgl. der Gesamtqualität ermittelt werden.

Sprachproben

Für die auditiven Tests wurden zwei Sätze aus [3] verwendet (ein Sprecher, eine Sprecherin). Diese wurden mit Hilfe eines Leitungssimulators für leitungsbasierte und paketbasierte Verbindungen kontrolliert verarbeitet [2]. Bei der Auswahl der Sprachproben musste ein Kompromiss getroffen werden: Einerseits sollen die Sprachproben ein möglichst großes Spektrum möglicher Störungen umfassen, um den *Wahrnehmungsraum* vollständig zu erfassen. Andererseits ist die Zahl der Proben aufgrund des Aufwands der Experimente begrenzt.

Tab. 1 stellt die verwendeten Sprachproben dar. Diese umfassen verschiedenartige Codecs, flache und handapparat-typische Sendefilter, Bandbreiteneinschränkung (BP), Verwendung von Freisprechern (HFT, *Hands-Free Terminal*), Leitungs- und Hintergrundrauschen, Geräuschreduktions-Verfahren (nach Boll, Ephraim-Mallah), zufällige Paketverluste und Unterbrechungen mit kosinusförmigen Rampen.

Abkürz.	Codec	Sendefilter	zusätzl. Störung
C1	G.711	Handapp.	-
C2	G.726	Handapp.	-
C3	G.729A	Handapp.	-
C4	AMR	Handapp.	-
H	G.711	flach	HFT
T1	G.711	Handapp.	BP 0.5-2kHz
T2	G.711	flach	-
I1	G.729A	Handapp.	10% Paketverluste
I2	G.729A	Handapp.	20% Paketverluste
I3	G.711	Handapp.	10% Unterbr.
HN	G.711	flach	HFT, Htgr.-Rauschen
HR1	G.711	flach	HFT, Boll
HR2	G.711	flach	HFT, Ephraim-Mallah
CN	G.711	Handapp.	Ltgs.-Rauschen

Tabelle 1: Sprachproben

Multidimensionale Analysen

Multidimensionale Skalierung (MDS)

In diesem Experiment wurde die Ähnlichkeit paarweise präsentierter Sprachproben von 14 VPn (6 weibl., 8 männl., 21-30 J.) bewertet [2]. Dazu wurde eine kontinuierliche, bipolare Skala mit den Skalen-Enden „sehr ähnlich“ bzw. „überhaupt nicht ähnlich“ verwendet. Die grundsätzliche Idee der MDS ist die Abbildung der Ähnlichkeiten auf metrische Distanzen [4]: Je höher die Unähnlichkeit zweier Sprachproben bewertet wird, desto größer ist die Distanz zwischen den entsprechenden Punkten, die die Sprachproben in einem n-dim. Raum repräsentieren. Durch die Lage der Punkte kann versucht werden, die n Dimensionen entsprechend charakteristischer Eigenschaften korrespondierender Sprachproben zu interpretieren und zu benennen.

Mit $I = 14$ Sprachproben wurden $I \cdot (I-1) = 182$ Paarvergleiche pro VP und Sprecher durchgeführt, um den Raum zu bestimmen. Die Dimensionalität n des Raumes wurde anschließend derart festgelegt, dass einerseits der Fehler zwischen den Ähnlichkeitsurteilen und den Distanzen möglichst gering wurde, und andererseits die Dimensionen anhand der Konfiguration der Punkte interpretierbar sind.

Semantisches Differential (SD)

Die Tatsache, dass bei der MDS a-priori keinerlei Anhaltspunkte über Eigenschaften der Sprachproben existieren, erschwert die Interpretation der Lösung. Diesen Nachteil gleicht die Technik des Semantischen Differentials aus. Hier beurteilte eine weitere Gruppe von 18 VPn (9 weibl., 9 männl., 21-31 J.) die Ausprägung der die Sprachprobe beschreibenden Attribute. Diese Attribute wurden zunächst in Vorversuchen bestimmt und in Form von *Antonymen* (Gegenwörtern) angegeben (Tab. 2). Sie bildeten die Enden kontinuierlicher, bipolarer Skalen. Pro Sprachprobe war ein Satz von 13 Attributen zu bewerten ($13 \cdot I$ Urteile pro VP und Sprecher).

verzerrt - unverzerrt	*indirekt - direkt*	*dunkel - hell*
dumpf - ungedämpft	*wackelig - fest*	*dünn - voll*
deutlich - undeutlich	*fern - nah*	*klar - unklar*
unterbrochen - kontinuierlich	*knisternd - nicht knisternd*	
zischend - nicht zischend	*unverrauscht - verrauscht*	

Tabelle 2: Antonyme

Eine anschließende Faktorenanalyse fasst korrelierte Attribute zu signifikanten Faktoren F_i zusammen. Eine Benennung der Faktoren wird ausgehend von den dem Faktor zugrunde liegenden Attributen und der Abbildung der Sprachproben im resultierenden Raum bestimmt.

Ergebnisse

Die Faktorenanalyse des SD ergab einen 3-dim. Raum (gemittelt über beide Sprecher). In Tab. 3 sind die Faktorwerte der abgebildeten Sprachproben dargestellt.

	F_1	F_2	F_3		F_1	F_2	F_3
C1	1.01	0.91	-1.21	H	-0.89	0.65	0.14
C2	0.86	0.22	0.51	HN	-0.53	0.27	1.78
C3	0.81	0.73	-0.78	HR1	-1.16	-0.41	0.13
C4	0.60	0.53	-0.52	HR2	-1.52	-0.15	0.40
CN	1.06	0.77	2.18	I1	0.17	-1.29	-0.16
T1	-1.86	0.55	-0.95	I2	0.25	-1.97	-0.56
T2	0.39	0.98	-1.02	I3	0.81	-1.79	0.06

Tabelle 3: Faktorwerte

Am positiven Ende von F_1 sind Sprachproben zu finden, die sich im verwendeten Codec unterscheiden. Das negative Ende wird von HFT- sowie der bandbeschränkten Sprachprobe T1 gebildet. Dieser Faktor ist hoch korreliert ($\varrho \geq 0.9$) mit den Begriffen *fern-nah*, *indirekt-direkt* (Oberbegriff: **Direktheit**) und *dünn-dick*, *dumpf-ungedämpft*, *dunkel-hell* (Zusammenhang mit dem **Frequenzgehalt**, d.h. mit den im Signal enthaltenen Frequenzen). F_2 ist hoch korreliert mit den Begriffen *unterbrochen-kontinuierlich* und *wackelig-fest* einerseits, und mit Paketverlusten und Unterbrechungen behafteten Sprachproben (I1-I3) am negativen Ende andererseits. Dieser Faktor lässt sich offenbar mit **Kontinuität** benennen. F_3 kann aufgrund hoher Korrelation mit *unverrauscht-verrauscht* und *nicht zischend-zischend* bzw. aufgrund der verrauschten Sprachproben am positiven Ende (HN, CN) als **Rauschhaftigkeit** interpretiert werden.
Die MDS ergab für beide Sprecher einen 4-dim. Raum [2], wobei **Direktheit** und **Frequenzgehalt** hier als Einzeldimensionen auftauchen. Beide korrelieren wiederum hoch mit Faktor 1 des SD, so dass der 3-dim. Raum als stabil angesehen werden kann.

Modellierung der Gesamtqualität

In einem weiteren Hörversuch wurden Gesamtqualitätsurteile zu den Sprachproben gesammelt [2]. Durch multiple lineare Regression lassen sich Wichtungsfaktoren b_i finden, mit deren Hilfe die MOS-Werte durch die Koordinaten der entsprechenden Sprachproben im 3-dim. Raum modelliert werden können. Die Modellierung folgt dem Zusammenhang $MOS = \sum_{i=1}^{3} b_i F_i$, die Varianzabdeckung beträgt $R^2 \approx 90\%$.
Der so erhaltene Koeffizient $b_2 = 0.70$ deutet eine hohe „Wichtigkeit“ des Faktors **Kontinuität** bezüglich der Gesamtqualität an (je höher die Kontinuität der Sprachproben, desto besser die Qualität). Der Koeffizient für **Rauschhaftigkeit** beträgt $b_3 = -0.47$ und für **Direktheit/Frequenzgehalt** $b_1 = 0.46$.

Zusammenfassung und Ausblick

Mit Hilfe zweier unabhängiger multidimensionaler Analysen konnte eine stabile Abbildung des Wahrnehmungsraumes gewonnen werden, der die für die Sprachqualität in modernen Telekommunikationsnetzen relevanten perzeptiven Dimensionen widerspiegelt. Diese sind, in der Reihenfolge der „Wichtigkeit“ bezüglich der Gesamtqualität: **Kontinuität**, **Rauschhaftigkeit** und **Direktheit/Frequenzgehalt**. Die Dimensionen bilden die Grundlage zur Entwicklung eines neuen Verfahrens für die Schätzung von Sprachqualität [1][5].

Diese Arbeit wurde unterstützt von der Deutschen Forschungsgemeinschaft (MO 1038 und HE 4465).

Literatur

[1] Heute, U.; Möller, S.; Raake, A.; Scholz, K.; Wältermann, M.: Integral and Diagnostic Speech-Quality Measurement: State of the Art, Problems, and New Approaches. In: Proc. Forum Acusticum 2005, Budapest 2005, 1695-1700

[2] Wältermann, M.: Bestimmung relevanter Qualitätsdimensionen bei der Sprachübertragung in modernen Telekommunikationsnetzen. Dipl.-Arbeit, Ruhr-Univ. Bochum, 2005

[3] ITU-T Rec. P.501 Amendment I, International Telecommunication Union, CH-Genf, 2004

[4] Kruskal, J.; Wish, M.: Multidimensional Scaling, Vol. 07-011 of Quantitative Applications in the Social Sciences (E.M. Uslaner, ed.), Sage, US-Newbury Park CA, 1978

[5] Scholz, K.; Wältermann, M.; Huo, L.; Raake, A.; Möller, S.; Heute, U.: Messtechnische Erfassung der Qualitätsdimension 'Direktheit/Frequenzgehalt' zur instrumentellen Analyse und Beurteilung von Sprachqualität. DAGA 2006, Braunschweig, 2006