Audio-Visuelle Qualität: Zum Einfluss des Audiokanals auf die Videoqualitäts- und Gesamtqualitätsbewertung

Falk Schiffner[1] & Sebastian Möller[2]
Quality and Usability Lab – Technische Universität Berlin
[1] *falk.schiffner@tu-berlin.de* [2] *sebastian.moeller@telekom.de*

Einleitung

Audio-Visuelle-Kommunikationsdienste sind weit verbreitet [1, 7] und die Schätzung der wahrgenommenen audio-visuellen Qualität dieser Dienste geschieht oftmals mittels instrumenteller Verfahren [1] (ITU-T G.1070 [6]). Der Ansatz, solche Schätzungen mittels eines wahrnehmungsbasierten Modells zu verbessern, wurde bereits bei Sprach-Telefonie verfolgt [8].

Die vorliegende Studie dient der Untersuchung der Formulierung von Audio- und Videoqualitätsurteilen.

Forschungsfragen

- *Hat der Audiokanal einen Einfluss auf die Qualitätsbewertung des Videos?*
- *Wenn ja, wie groß ist der Einfluss und unterscheidet er sich bei unterschiedlichen Videostörungen?*

Es handelt sich um eine erste Studie zur Ermittlung des qualitätsrelevanten Wahrnehmungsraums für Videotelefonie.

Experiment

Versuchteilnehmer

Für diese Studie wurden 20 Teilnehmer über ein Probandenprotal akquiriert. Die Teilnehmergruppe bestand aus 10 männlichen und 10 weiblichen Personen. Das Durchschnittsalter betrug $30,2$ Jahre (Stdev $7,96$). Alle Teilnehmer wurden vor Beginn einer Sehprüfung (Ishihara-Test, Snellen-Sehtafeln) und einer Hörprüfung (frequenzkonstante Békésy-Audiometrie $(250, 1k, 4k, 8k[Hz])$ [2]) unterzogen. Keiner der Probanden musste von der Studie ausgeschlossen werden. Alle Teilnehmer haben eine Entschädigung erhalten.

Versuchsmaterial und Beeinträchtigungen

In der Tabelle 1 sind die Daten zum Testmaterial aufgelistet. In Abbildung 1 ist beispielhaft ein Screenshot einer Videoprobe zu sehen.

In diesem Experiment wurde ausschließlich das Videomaterial gestört. Das Audiosignal blieb unbeeinträchtigt. Für die Störungen des Videomaterials wurden Beeinträchtigungen aus dem *Reference Impairment System for Video* (RISV) (ITU-T P.930 [5]) gewählt. Zusätzlich wurden Kodierungseffekte, Effekte von Paketverlusten und kombinierte Störungen eingefügt. Ziel war es ein breites Spektrum von möglichen Videostörungen abzudecken.

Tabelle 1: Übersicht – Versuchsmaterial

VIDEO:	
Material	Videotelefonie / Kopf-und-Schulter Szene
Auflösung	$640x480$
Wiederholrate	$25fps$
Bilddiagonale	$18,5cm$
Betrachtungsabstand	ca. $60cm$
AUDIO:	
Material	Sprache (passend zum Video)
Abtastrate	$16kHz$
Bitrate	$256kbps$
Darbietung	Kopfhörer (binaural)
Schalldruckpegel	$73dBSpl$ (mono) / $79dBSpl$ (binaural)

Eine Übersicht der Beeinträchtigungen mit kurzer Beschreibung ist in Tabelle 2 zu sehen. Die Störungen wurden mit Hilfe von *Matlab*, *ffmpeg*, *netem* und *TrafficControl (TC)* eingefügt. Bei der *RISV artifical Bockiness* wurden zwei verschiedene Blockgrößen verwendet. Es wurde $5x5$ und $8x8$ Pixel-Blöcke durch Mittelung der Farbwerte erstellt. Für die Erzeugung des Unschärfe-Effekts (*RISV artifical Blurring*), wurde der in der ITU-T P.930 [5] vorgeschlagene Filter Nr.1 verwendet.
Um den Unschärfe-Effekt noch zu erhöhen wurde ein eigener Unschärfefilter (*„Filter7“*) verwendet. Beide Filter wurden zeilenweise auf jeden Frame angewendet.
Das Bildruckeln (*RISV artifical Jerkiness*) wurde durch das Halten eines Einzelbildes über eine gewissen Anzahl von Frames (6 od. 11) erzeugt.
Das, in das Bild eingebrachte, künstliche Rauschen (*RISV artifical NoiseQ*), wurde das Videomaterial zuerst vom RGB- in der YCbCr-Farbraum überführt. Anschließend wurden nur die Luminanzwerte der einzelnen Pixel beeinträchtigt. Hierfür wurde je nach Störungsstärke, für jeden Frame ein Fehlermuster generiert und die Luminanz auf zufällige Werte gesetzt.
Um Kodierungseffekte einzufügen, wurde eine two-pass Kodierung mit dem $H.264$ Videokodierer mit zwei verschiedenen Bitrate gewählt. Hierbei wird im ersten Schritt ein Profil des Videomaterials anlegt. Im zweiten Schritt folgt dann die Kodierung des Videomaterials anhand des Profils. Dies führt zu einer gleichbleibenden Störungsstärke über die gesamte Dauer des Videos. TrafficControl und Netem wurden zur Erzeugung von Paket-

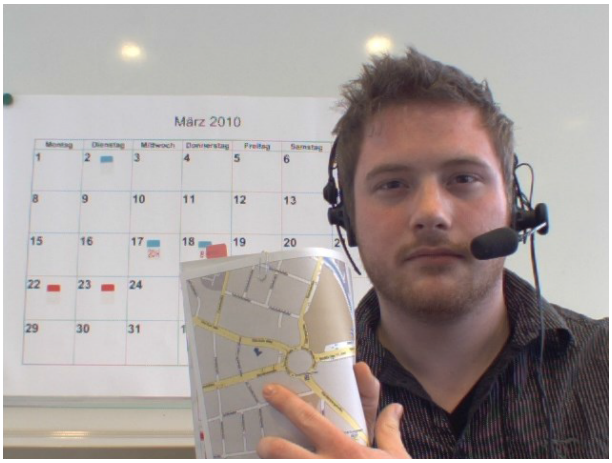

Abbildung 1: Screenshot – Bsp. Videomaterial, Kopf-und-Schulter Szene.

verlusten verwendet.
Bei den kombinierten Ströungen wurden *artifical Blurring* und *artifical Jerkiness* mit *artifical Blockiness* kombiniert. Jedoch handelt es sich hier um nicht gleichmäßig verteilte Blöcke im Bild, sondern die Blockgröße bestand aus $20x20$ Pixel-Blöcken. Innerhalb eines Frames waren nur 30% der Blöcke gestört. Zusätzlich waren nur 35% der Frames einer Videoprobe zufällig gestört. Dies führte zu einer gewollten sehr unregelmäßigen Anmutung der Störung.

Versuchsprozedur

Der Versuch war in zwei 20-minütige Teile (Durchgang A und Durchgang B) unterteilt. Im Durchgang A wurde das Videomaterial ohne Audiosignal dargeboten. Im Durchgang B wurde das Video mit passendem Audiosignal (Sprache) dargeboten. Die Reihenfolge der Durchgänge A und B wurde bei der Hälfte der Teilnehmer gewechselt. Die Darbietung der Stimuli innerhalb der Durchgänge wurde randomisiert. Zu Beginn eines jeden Durchgangs gab es ein kurzes Training, damit sich die Versuchsperson mit der Versuchsaufgabe und der Bewertungsoberfläche vertraut machen konnten. Es wurde mittels einer kontinuierlichen 11 Punkte-Skala (vgl. ITU-T P.910 [3]) bewertet. In Abbildung 3 ist examplarisch die Bewertungsskala zu sehen. Im Durchgang B wurde die Reihenfolge der Bewertungsskalen varriert. Zuerst wurde die Gesamtqualität abgefragt. Anschließend wurde die Video- und Audioqualität abgefragt. Die Reihenfolge der Skala für die Video- und Audioqualität wurden von Versuchsperson zu Versuchsperson gewechselt. Dadurch sollte ein Reihenfolgeneffekt ausgeschlossen werden. Die Teilnehmer konnten zwischen den beiden Durchgängen eine fünfminütige Pause einlegen.

Ergebnisse

In der Abbildung 2 sind die Bewertungen aus beiden Durchgängen zu sehen. Die grünen Balken stellen die Qualitätsbewertung des Audiokanals dar. Es ist zu erkennen, dass das ungestörte Sprachsignal einen konstanten Qualitätswert (*„gut“*) erzielt (vgl. Abb. 3). Hieraus wird

Tabelle 2: Kurzbeschreibung der Beeinträchtigungen für das Videomaterial.

Name	Beschreibung
Reference	Ungestörtes Videomaterial
RISV artifical Blockiness 5x5/8x8	homogene Störung, alle Frames (zwei Blockgrößen)
RISV artifical Blurring ITU(F1)/ Filter7	homogene Störung, alle Frames (zwei Unschärfefilter)
RISV artifical Jerkiness 6/11 Frames	Bildruckeln (6 bzw. 11 Frames gehalten)
RISV artifical NoiseQ 3% / 15%	Salz & Peffer Rauschen (x% der Pixel pro Frame)
H264 Bitrate 28 / 56*kbps*	H.264 - Codec Bitrate (2-pass Kodierung)
Packet Loss $0,5\%$ / $1,5\%$	H.264 - Codec (Bitrate: Hoch), TC, Netem
Artifical Impair. Combi I (Blurr+Block)	artifical Blurr.(F1) + artifical Blocki (hetrogene)
Artifical Impair. Combi II (Jerki+Block)	artifical Jerki.(6) + artifical Blocki (hetrogene)

der Schluss gezogen, dass keine Videostörung die Bewertung des Audiokanals beeinflusste. Bei der Betrachtung der Videoqualitätsurteile, zeigt sich, dass mit steigender Störungsintensität das Qualitätsurteil sinkt (vgl. Abb. 2 bspw. *Blockiness* oder auch *Noise*).
Bei der Gegenüberstellung der Videoqualitätsurteile aus dem Durchgang A (Abb. 2, rote Balken = *Video_only*) mit den Videoqualitätsurteilen aus dem Durchgang B (Abb. 2, dunkelblaue Balken = *Video_with Audio*) ist kein signifikanter Unterschied zwischen den Urteilen auszumachen. Es wurde eine ANOVA durchgeführt ($F(1,28) = 0,048, p = 0,82$). Somit kann der Schluss gezogen werden, dass die Urteile zur Videoqualität unabhängig von der Anwesenheit des Audiokanals sind. Es zeigt sich auch, dass die Stärke der Videobeein-

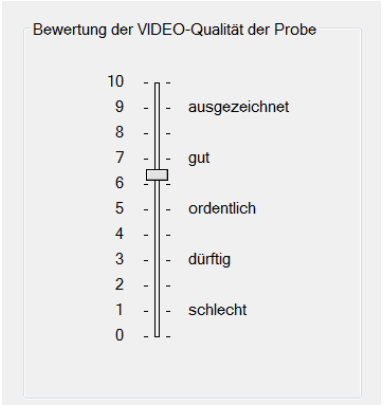

Abbildung 3: Verwendete kontinuierliche 11 Punkte-Skala – Bsp.: Bewertung der Videoqualität.

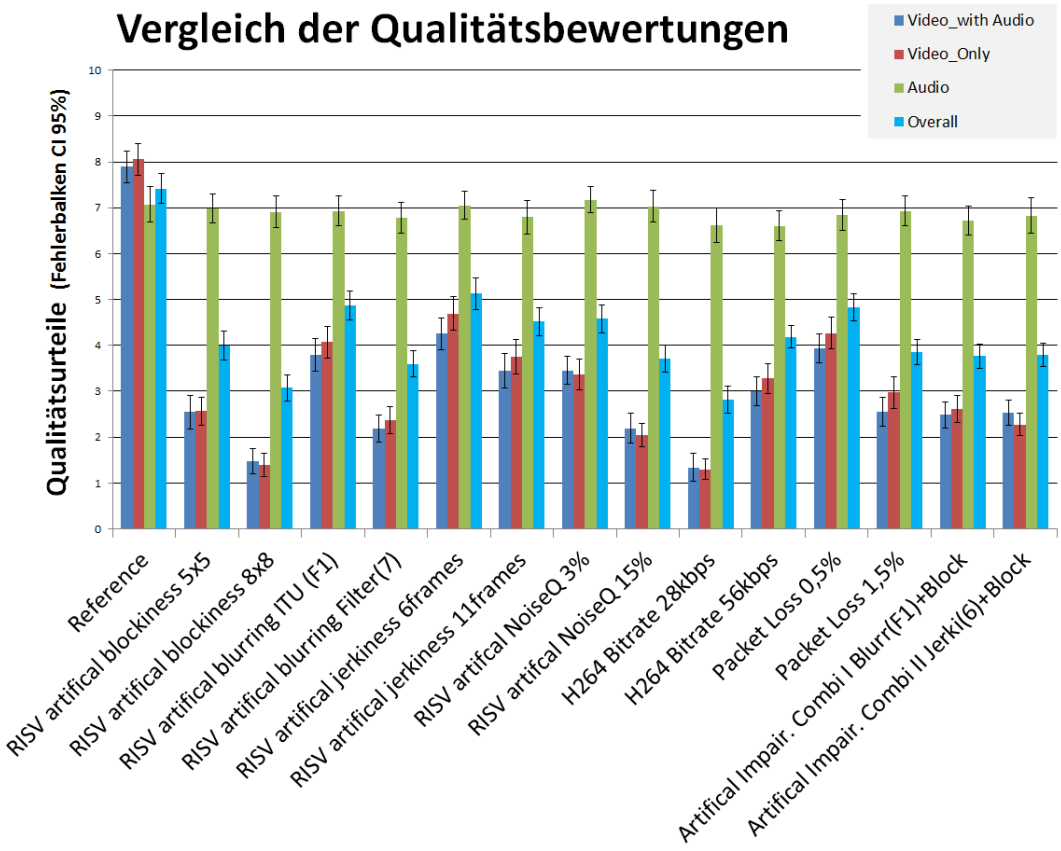

Abbildung 2: Die subjektiven Qualitätsbewertungen beider Durchgänge (0 = schlecht möglichste Qualität - 10 = best mögliche Qualität), Fehlerbalken: 95% Konfidenzintervall (CI95%).

trächtigungen zu groß gewählt war, da alle Videobeeinträchtigung einen Qualitätswert unterhalb von *„ordentlich“* (vgl. Abb. 3) erzielt haben. In folgenden Studien sollte die Spanne der Beeinträchtigungen, so gewählt werden, dass die gesamte Breite der Bewertungsskala genutzt wird.

Bei der Betrachtung der Urteile zur *Gesamtqualität* ist zu beobachten, dass bei allen beeinträchtigten Videoproben die Gesamtqualität höher ist, als die jeweilige Videoqualität. Der Grund dafür ist im Einfluss der *„guten“* Sprachsignale zu sehen, welche, bei der Integration der einzelen Urteile zur Gesamtqualität, die schlechten Videoqualitätsurteile nach oben ziehen.

Die in Abschnitt beschriebene variierde Reihenfolge der Bewertungsskalen in Durchgang B wurde untersucht. Aus dem t-Test ($t(28) = 0,22, p = 0,83$) konnte kein Einfluss der Bewertungsskalenreihenfolge beobachtet werden. Somit ist es unerheblich, ob Bewertungsreihenfolge *Gesamt-, Video- und Audioqualität* oder *Gesamt-, Audio- und Videoqualität* war.

Schlussfolgerung

Anhand des Experiment konnte gezeigt werden, dass der Audiokanal keinen Einfluss auf die Qualitätsbewertung des Videos hat. Somit können Versuche zur Videoqualität und der Beschreibung von Störungen im Video ohne Audio durchgeführt werden. Dieses Ergebniss bestätigt auch, dass die Vorgehensweise bei der Schätzung der audio-viusellen Qualität, richtig ist. Hierbei wird die Video- sowie die Audioqualität zunächst getrennt geschätzt und anschließend über eine Gewichtung zusammengeführt, um so die Gesamtqualität zu bestimmen (vgl. ITU-T G.1070 [6]).

Ausblick

In Zukunft soll ein wahrnehmungsbasierter Videoqualitätsschätzer entwickelt werden. Dieser Schätzer soll mit einem Sprachqualitätsschätzer kombiniert werden, um die audio-visuelle Qualität von Videotelefonie zu ermitteln.

Literatur

[1] B. Belmudez & S. Möller, „Audiovisual quality integration for interactive communications“, EURPSIP Journal on Audio, Speech, and Music Processing 2013

[2] E. Lehnhardt & R. Laszig, „Praxis der Audiometrie“, 9. Auflage, Thieme Verlag 2009

[3] ITU-T Rec. P.910, “Subjective video /quality assessment methods for multimedia applications”, Int. Telecomm. Union, Geneva, 04/2008

[4] ITU-T Rec. P.830, “Subj. performance assessment of telephone-based and wideband digital codecs”, Int. Telecomm. Union, Geneva, 02/1996

[5] ITU-T Rec. P.930, “Principles of a reference impairment system for video”, Int. Telecomm. Union, Geneva, 04/1996

[6] ITU-T Rec. G.1070, “Methods for Subjective Determination of Transmission Quality”, Int. Telecomm. Union, Geneva, 07/2012

[7] Internet, Statista, „Anteil der Nutzer von Internet- oder Videotelefonie in Deutschland in den Jahren 2008 bis 2015 ”, letzter Besuch 02/2016

[8] M. Wältermann, A. Raake, S. Möller „Modeling of Integral Quality Based on Perceptual Dimensions - A Framework for a New Instrumental Speech-Quality Measure“, ITG Fachtagung Sprachkommunikation, D-Aachen, 2008. VDE Verlag GmbH